5 SAMPLING

Everyone takes surveys. Whoever makes a statement about human behavior has engaged in a survey of some sort.

—*Andrew Greeley, sociologist*

LEARNING GOALS

This chapter explains random probability sampling and describes other methods for obtaining samples. You will also learn about longitudinal and cross-sectional research designs. By the end of the chapter, you should be able to distinguish several types of probability and nonprobability sampling, describe various kinds of longitudinal research designs, and explain the idea of sampling error.

Ask your friends whether they have ever been surveyed about what television shows they watch. Every week the ratings come out with a listing of the most popular shows. Yet so few people seem to know anyone who was ever asked by a pollster what they watched on television that week. It is probably not surprising, given that only a few thousand people or so were actually surveyed. How can such a small sample of people accurately represent the entire nation?

Each election year, public opinion polls are routinely published with the latest voter preferences, and immediately the losing side claims the polls are not accurate. After all, only 1,000 people are surveyed in the typical political poll, and no one we know was ever asked. How then do we explain why the final presidential election results have almost always been within a few percentage points (the margin of error) of the last public opinion poll taken before the election when such a small sample of voters was measured? (see Box 5.1). Typically, polls report the *margin of error*, and this number applies to each candidate's percentage, not to the spread between the candidates. So if

BOX 5.1
POLLING SAMPLES AND ACTUAL RESULTS

Look at the final Gallup Poll results in Table 5.1 for the winning candidate in presidential elections since the 1960s. These are the polls closest to Election Day and include results of likely voters with an allocation of undecided votes. The actual results are almost always within the usual (2 or 3 percent) margin of error, even in the 2016 election. These percentages represent the popular vote, not Electoral College results.

GALLUP POLL			
YEAR	**FINAL POLL**	**ACTUAL RESULTS**	**DIFFERENCE**
1960: Kennedy	50.5%	50.1%	0.4%
1964: Johnson	64%	61.3%	2.7%
1968: Nixon	43%	43.5%	−0.5%
1972: Nixon	62%	61.8%	0.2%
1976: Carter	48%	50.1%	−2.1%
1980: Reagan	47%	50.8%	−3.8%
1984: Reagan	59%	59.2%	−0.2%
1988: G.H.W. Bush	56 %	53.0%	3.0%
1992: Clinton	49%	43.3%	5.7%
1996: Clinton	52%	49.2%	2.8%
2000: G. W. Bush	48%	47.9%	0.1%
2004: G. W. Bush	49%	50.7%	−1.7%
2008: Obama	55%	53.0%	2.0%
2012: Obama	49%	51.1%	−2.1%
2016: Trump*	43.6%	46.2%	−2.6%

* from Real Clear Politics Poll; Gallup no longer conducts presidential election polls.

Source: Based on Peters (2017).

candidate A polls 45 percent and candidate B gets 49 percent and the margin of error is 4 percent, then the true results for candidate A are anywhere between 41 and 49 percent, and between 45 and 53 percent for candidate B. Thus, while candidate B might appear ahead in the polls at 49 percent, candidate A could actually be at 49 percent while candidate B might have support from only 45 percent of those asked.

Polling people for their opinions is a common practice in contemporary society. But not all polls are done accurately, and their reputations can suffer greatly, as was seen during the 2016 U.S. presidential election, despite findings that these "National polls were generally correct and accurate by historical standards" (AAPOR 2017: 2). For example, your friends on Facebook may ask you to vote on some topic, or the local news channel may ask viewers to text their responses to a question about some controversial issue. You would think these findings are valid, given the large number of people who answer. However, even though the quantity of responses to the surveys is substantially greater than the number reporting their television viewing habits to a ratings monitoring company, these polls are not scientifically accurate or generalizable. All these types of polls reflect the opinions of only those Internet users, cell phone texters, or television news viewers who have chosen to participate and may not even represent the opinions of other users or viewers, let alone the public as a whole. Why this is so and how we can conduct a survey with a representative sample with minimal funding and time are the topics of this chapter.

SOME BASIC SAMPLING CONCEPTS

As mentioned in Chapter 1, the goals of research may be to describe, explain, explore, or predict characteristics of a population. A *population* is the total collection of units or elements you want to analyze. Whether the units you are talking about are a country's citizens, schools, editorials in newspapers, Twitter postings, or local businesses, when the population is small enough, you can easily survey every element of the population. Of course, this means you have to be able to define clearly what is part of the population. For example, if you want to study the population of all students on a campus, you need to first determine who is a student: Do you include full-time and part-time students, those on leave of absence but still officially registered as students, or those studying abroad this year? And once you determine the actual population, where do you find a complete and accurate list of them? We know that a population survey of all Americans (that is, the U.S. Census) misses quite a few people.

The *unit of analysis* is the element about which you are observing and collecting data, such as a school, an editorial, a person responding to a political poll, or a Twitter post. Sometimes, however, the element surveyed is merely a mechanism to collect information about some other unit of analysis you need for your research questions. For example, you might survey school principals (the elements selected from the population of all principals) in order to gather information about student-teacher ratios in their schools. The unit of analysis is now the school and its student-teacher ratio, but the element selected randomly was the principal.

Let's focus on a group in which the unit of analysis is also the element selected from the population. If you want to know how everyone at your company feels about a new health benefits policy, and the number of employees is small enough, go survey each person. Then you would have information about the variables in this population of workers; these findings are usually called the *population parameters*. Or consider when a professor wants to find out what every student in a class thinks about the course and the teacher. As is done in most classes, a course evaluation is completed by surveying every unit of the class population, assuming no one is absent that day. The resulting data are the population parameters.

However, the population we are often interested in is usually larger than a class and, with limited time and money, we are unlikely to survey each and every one of the elements in a population. Therefore, in order to figure out what the population parameters are, we need to make inferences from a subset of the population; that is, we need to generate some *statistics* from a *sample* of people chosen to represent the entire population. Technically, the word "statistic" refers to the information we have about the variables in a sample. Population parameters are what we estimate or infer from *sample statistics* when we collect data from a sample because we were unable to do a full population study. In most cases, there will be a difference between the information (the statistics) we gather about the sample and what the true parameters of the population are. This difference is called the *sampling error*. As with political polls, for example, the "margin of error" is often stated in press releases as "plus or minus four percentage points."

In most cases, we need to generate a sample, but we still have to identify and count every element in the population in order to do accurate sampling. Sometimes employees, for example, are on leaves of absence. Should they be included in the list of every element of the population? Is this a study of both full-time and part-time workers? Of middle management and executives? In short, the researcher must first define the study population from which the sample will be selected and to which the resulting statistics will be generalized.

The best rule of thumb is that you can talk about only the population that is represented in the sample. If part-time students or fraternity/sorority members or off-campus students are not part of the population list (sometimes called the *sampling frame*) from which a sample is generated, then the resulting statistics cannot be generalized to all students at the university. The findings must be qualified as applicable only to full-time, on-campus, non-fraternity/sorority students. As discussed later, if the respondents are obtained using what is called a nonprobability sampling method, then the information gathered can be used to describe, explain, or predict information about only those people who are in your sample and completed the survey. You cannot say the data are meant to represent the opinions, behavior, or demographics of all students, or all Facebook users, or any population of elements greater than those sampled.

PROBABILITY SAMPLING

When we go for a medical examination and get a blood test, it's not necessary to take all our blood—a sample will do, thank you very much. Because blood is homogeneous throughout our veins, any vein will do, and as little as a drop or as much as a vial will suffice. Similarly, if the population we wanted to study were totally homogeneous, then any one element would do to represent the whole. But it is a very unlikely event that all elements are exactly the same on all characteristics in reality. Heterogeneity reigns, and our population usually varies quite a bit in attitudes, behavior, and demographics. Therefore, we are obliged to take a sample of elements that represent what is in the population. For example, if the population we are interested in is composed of students with different majors, then our sample should reflect the range of majors in almost the same proportion that they exist on the campus. If 10 percent of the students major in sociology, then 10 percent of the sample should also be sociology majors.

In order to achieve this, we must use a sampling process that is likely to result in a representative sample. This will happen if each person has an equal chance of being chosen for the sample study. The proportion of people in the final sample who are of different ethnicities, genders, sexual orientations, ages, and so on should look pretty close to their distribution in the population. If the population is 60 percent Hispanic, then the sample chosen should also be around 60 percent Hispanic. Unless you wanted to under- or oversample a subgroup (and this can be done by weighting or through disproportionate sampling techniques, both discussed in more advanced textbooks), probability sampling will ensure approximate representation of the population (see Box 5.2).

BOX 5.2
PROBABILITY THEORY, SAMPLING, AND M&MS

M&Ms, my favorite candy, reports the following percentage of colors for their milk chocolate (plain) candies: 13 percent each of brown and red, 20 percent orange, 24 percent blue, 14 percent yellow, and 16 percent green. (These proportions change from time to time, so by the time you do your own candy sampling "research," the numbers may vary, but the concept of sampling distributions remains the same in this example.)

Imagine you have been invited by the company to dip a huge scoop into some large barrel of these candies at their factory. Let's say you picked out exactly 100 M&Ms. How many colors will you be tossing into your mouth? On the basis of their figures, you should have 13 brown, 14 yellow, 13 red, 20 orange, 16 green, and 24 blue ones in your scoop. And if you got only 50, you would have 6 or 7 brown, 7 yellow, 6 or 7 red, 10 orange, 8 green, and 12 blue ones.

So, purely in the interests of research, I went to the store and bought a package of M&Ms. The 1.69-ounce bag had 57 candies in it. Therefore, on the basis of probability theory, it should contain 7.4 (13 percent of 57) brown ones, 8 yellow, 11.4 orange, and so on. Since it's unlikely that there will be 0.4 of a candy, let's round off the numbers and expect around 7 or 8 brown, 8 yellow, 7 or 8 red, 11 or 12 orange, 9 green, and 14 blue. Remember, one bag is a sample of all possible M&Ms (the entire population of candies is not in my little package!), and here's what I find: 14 orange, 9 red, 16 green, 7 brown, 9 yellow, and 2 blue. My colors are pretty close to the overall population distribution, with the exception of the blue and green ones. Perhaps there is some sampling error with a few of the colors, probably due to the small sample size.

BOX 5.2 CONTINUED

To make sure, I take another sample and buy a second package. (Well, somebody has to do this kind of research!) Remember, the proportion would be exactly what the company says it is, if I took all possible samples, but I'm only doing two. Guess what? In this one, out of 56 pieces, I have 6 red ones, 11 orange, 16 green, 4 blue, 13 yellow, and 6 brown ones. As random sample sizes increase, the sample statistics increasingly come closer to the actual population parameters, so let's combine the two packages to achieve a larger sample and see how close my sample comes to the expected distribution of colors (see Table 5.2).

Table 5.2 M&M Sample Color Distribution

	PACK #1 (N = 57)	PACK #2 (N = 56)	TOTAL (N = 113)	EXPECTED (N = 113)
Brown	7 (12.3%)	6 (10.7%)	13 (11.5%)	14.7 (13%)
Yellow	9 (15.8%)	13 (23.2%)	22 (19.5%)	15.8 (14%)
Red	9 (15.8%)	6 (10.7%)	15 (13.3%)	14.7 (13%)
Orange	14 (24.6%)	11 (19.6%)	25 (22.1%)	22.6 (20%)
Green	16 (28.1%)	16 (28.6%)	32 (28.3%)	18.1 (16%)
Blue	2 (3.5%)	4 (7.1%)	6 (5.3%)	27.1 (24%)

I could keep buying bags of candy and counting them by color until I had all the M&Ms ever made, but for a variety of dietary and financial reasons, I don't. By increasing the number of candies by the quantity in each additional bag, the percentage of each color gets closer to the actual number produced by the company. The question becomes: Can I generalize to the entire population of M&Ms manufactured, on the basis of any one sample of a single 1.69-ounce package? Theoretically, if I were to look at all possible samples, taking 57 candies at a time, the average percentage of all brown ones would be exactly 13 percent. Some packets would probably be 15 percent brown, others might be 11 percent, some would have extremely few brown ones, and a few would have extremely many brown candies. But when these sample percentages were averaged together, brown ones would be 13 percent of the total because, at that point, all possible samples of 57 would equal the total number of M&Ms made. If my sample sizes were larger—that is, if I bought those jumbo bags of candy—they would even more closely follow the percentages of the true population figures.

And if I plotted these sample percentages on a graph, they would form a normal curve. But since I wouldn't be able to get all these samples, it is a theoretical distribution. This is what is called the *central limit theorem*, which is discussed more in Chapter 6. It helps me decide whether my particular sample is representative of the true population information or whether I have a sample so extremely different that the probability of having gotten it by chance is less than 5 percent, the traditional cutoff point used in the social sciences to declare that this was not a typical sample or finding.

Our goal, then, is to obtain a sample that is representative of the population. If I asked someone to put together a bag of M&Ms for me, that person might select the colors he or she favored or might try to balance the mix, erroneously assuming that the company makes equal quantities of each color. It would, in either case, be a biased sample and unrepresentative of the true population. Because there is no reason to believe that factory workers are choosing the colors for each bag of M&Ms, we assume that the colors are selected randomly by the packaging machines, and therefore each bag (sample) should be close to the population figures, especially in the larger bags of candy.

BOX 5.2 CONTINUED

My total sample percentages were virtually the same percentage expected for red candies (13 percent), within two percentage points for the brown and orange ones, and pretty close to what was expected for the yellow M&Ms. But there were much larger deviations from the expected for the blue and green candies. Did I just happen to get unusual samples of blue and green ones by chance alone because I looked at so few packages of them? Larger samples would eventually produce numbers closer to the expected. The difference between what I observed and what I expected is called the margin of error or the *sampling error*. Like public opinion polls, my sampling distribution of percentages will not be exactly like the population's actual percentage breakdown of colors, but it will be within a few percentage points plus or minus the expected, especially when the sample is larger. I guess it's time to buy a bigger bag—for research purposes only, of course!

When we can specify the (nonzero) probability of each element in the population being selected for membership in the sample, we have what is called a *probability sample*. The most common types include

- Simple random sampling
- Stratified random sampling
- Systematic random sampling
- Cluster or multistage sampling.

Simple Random Sampling

Don't be fooled by the word "random," which seems to crop up regularly in everyday slang. It is a mistaken use of the word when someone tells you she stood on a street corner and "randomly" gave out questionnaires or walked up to people "randomly" in the mall or asked some "random" person to participate in a study. The researcher may have aimlessly wandered around finding people, but that doesn't mean every person chosen was randomly selected. Did every shopper have an equal chance of being designated a respondent? Not likely, because some people were working, others shopped earlier in the day, and still others were hanging out at the other end of the mall.

Therefore, in order to achieve a true *simple random sample*, you must be able to provide a complete list of all possible units in the population from which to choose a sample. Whether you are randomly selecting ads in a magazine, students on a campus, or clients of a social service agency, somehow you must first get a complete and accurate set of elements according to the criteria you decide. For example, if you want to generate a sample of employees where you work, then you must first decide what an employee is (full-time, has worked at least six months, is not on maternity leave, etc.)

and be able to get a complete list of all who fit the criteria. This becomes the sampling frame from which a sample can be chosen.

All units in the sampling frame may be identified with a number, either computer generated or done by hand, in order to use random sampling techniques. Each ad in a magazine could be assigned a number, every student already has an ID or mailbox number, each employee has a worker or social security number. Using either a table of random numbers (found in many statistics and methods books) or computer-generated ones (such as www.random.org/integers), the units of analysis are chosen. Or their names can be written on pieces of paper, placed in a box, mixed well, and then, like a lottery, picked out at random until the desired number of respondents is selected. Some computer programs can also generate random samples just using the names, so there is no need to first assign numbers to each element in the sampling frame.

Telephone surveys employ probability sampling through *random digit dialing* techniques in which machines generate phone numbers within various area codes and then dial the numbers. Even those with unlisted numbers can be selected using this method. However, people without phones or who use only mobile phones are not part of the sampling frame, and this possibility could bias the resulting sample by underrepresenting the poorest people. See Box 5.3 for a description of this technique as used by a national polling company.

BOX 5.3

CELL PHONES, RANDOM DIGIT DIALING, AND SAMPLING

Here is a description of the sampling methods used by the Pew Research Center for the People and the Press, one of the leading national survey organizations:

> The typical Pew Research Center for the People and the Press national survey selects a random digit sample of both landline and cell phone numbers in all 50 U.S. states and the District of Columbia. As the proportion of Americans who rely solely or mostly on cell phones for their telephone service continues to grow, sampling both landline and cell phone numbers helps to ensure that our surveys represent all adults who have access to either (only about 2 percent of households do not have access to any phone). We sample landline and cell phone numbers to yield a combined sample with approximately 60 percent of the interviews conducted by landline and 40 percent by cell phone. This ratio is based on an analysis that attempts to balance cost and fieldwork considerations as well as to improve the overall demographic composition of the sample (in terms of age, race/ethnicity, and education). This ratio also ensures a minimum number of cell-only respondents in each survey.

BOX 5.3 CONTINUED

The design of the landline sample ensures representation of both listed and unlisted numbers (including those not yet listed) by using random digit dialing. This method uses random generation of the last two digits of telephone numbers selected on the basis of the area code, telephone exchange, and bank number. A bank is defined as 100 contiguous telephone numbers, for example 800–555–1200 to 800–555–1299. The telephone exchanges are selected to be proportionally stratified by county and by telephone exchange within the county. That is, the number of telephone numbers randomly sampled from within a given county is proportional to that county's share of telephone numbers in the U.S. Only banks of telephone numbers containing three or more listed residential numbers are selected.

The cell phone sample is drawn through systematic sampling from dedicated wireless banks of 100 contiguous numbers and shared service banks with no directory-listed landline numbers (to ensure that the cell phone sample does not include banks that are also included in the landline sample). The sample is designed to be representative both geographically and by large and small wireless carriers. . . .

When interviewers reach someone on a landline phone, they randomly ask half the sample if they could speak with "the youngest male, 18 years of age or older, who is now at home" and the other half of the sample to speak with "the youngest female, 18 years of age or older, who is now at home." If there is no person of the requested gender at home, interviewers ask to speak with the youngest adult of the opposite gender. This method of selecting respondents within each household improves participation among young people who are often more difficult to interview than older people because of their lifestyles.

Unlike a landline phone, a cell phone is assumed in Pew Research polls to be a personal device. Interviewers ask if the person who answers the cell phone is 18 years of age or older to determine if they are eligible to complete the survey . . . . This means that, for those in the cell sample, no effort is made to give other household members a chance to be interviewed. Although some people share cell phones, it is still uncertain whether the benefits of sampling among the users of a shared cell phone outweigh the disadvantages.

One of the main challenges of surveying cell phone users is drawing a representative sample of this group. Drawing samples for all telephone surveys is now more complicated because of the introduction of cell phone numbers and number portability (i.e., where people can keep their numbers when they move or change service providers and can port a landline number to a cell phone). Telephone numbers are assigned different prefixes, which can be used to identify whether the number is for a landline or cell phone, but there are also mixed or shared prefixes that include both landline and cell numbers. In addition, people who forward their calls (e.g., from their landline number at home or work to their cell) may appear as a landline number even when they are actually talking on their cell phone.

Source: Used by permission from The Pew Research Center for the People and the Press, www.pewresearch.org/methodology/u-s-survey-research/sampling/.

Stratified Random Sampling

Let's say that you want a sample with comparison groups of equal sizes, even if they are not that way in the population. For example, you know that there are more female employees, but you desire to have half of your sample female and the other half male. What do you do now? Based on probability theory, your final random sample should come out looking like the population with more women in it, yet somehow you have to ensure a 50-50 split. This is where *stratified random sampling* comes in. With this method, you disproportionately stratify (categorize) your sample along the lines you want to analyze by establishing quotas for certain kinds of respondents. Take religion: You might want 20 percent Roman Catholic, 20 percent Protestant, 20 percent Jewish, 20 percent no religion, and 20 percent other religions. So you divide the sampling frame into these categories and within each category (stratum), you take a simple random sample using the methods described previously until you get the proportion of respondents you desire for each category.

Stratified sampling can also be a way to guarantee an exact proportionate representation of the population, even if simple random sampling might yield almost similar results. If you know that the population is 60 percent female and 40 percent male, you can stratify on gender and achieve a 60:40 ratio in your final sample, rather than leave it to chance. You could also stratify on a combination of traits, such as both sex and Instagram users, in which you want each 25 percent of your sample to be men who post on Instagram, men who don't, women who use Instagram, and women who don't use it. Whatever strata you choose, you must first be able to identify people who fit into these categories and then perform random sampling within the strata. That's the key part for a probability sample: Without randomizing respondents, you would have a nonprobability quota sample (discussed in the section on Nonprobability Sampling) and lose the ability to generalize the results to a larger population.

If you don't know how many people in different categories responded until they completed the demographic section of the questionnaire, you can use statistical weighting techniques during the data analysis phase to adjust for under- and over-represented groups. This is discussed in advanced statistics books and available in statistical computer programs like SPSS.

Systematic Random Sampling

Another way of generating a random sample, especially if there is a large number of population members, is to use *systematic random sampling*. This involves taking every *n*th element in the sampling frame until the total is reached. Imagine you need a sample of 100 and you have 5,000 people or any other elements such as magazine ads, schools, or television characters to select from. If you divide 5,000 by 100, you get a sampling interval of 50. So you take every fiftieth person, ad, school, or whatever from the population list.

Let's use mailbox numbers on a college campus as an example of reaching each person in the population. Imagine these begin at number 101 and go up to 1,200, and you want a sample size of 100. You must begin randomly somewhere between and including mailbox number 101 and number 1,200. Since you have 1,100 mailbox numbers and need 100, your sampling interval is 11. Begin by getting a random number from 101 to 1,200 (either using a printed table of random numbers or a computer-generated one). Let's say you randomly select number 500 as your starting point. Mailbox 500 gets a questionnaire and then box number 511, 522, 533, and so on, until you get 100, increasing in sampling intervals of 11. If you want to distribute 200 questionnaires instead, then divide 1,100 by 200 to get a sampling interval of 5.5; you would now choose every sixth mailbox (because there is no such thing as a half mailbox) after starting randomly, such as 500, 506, 512, 518, and so on, until you select 200 mailbox numbers. There is no need with systematic random sampling to generate 200 random numbers, only a random starting one.

The sample resulting from this systematic method is typically quite similar to a simple random sample, unless for some reason there is a pattern to the order of the population elements. For example, consider the case where stratified random sampling selects every tenth house on a city's streets. At some point you notice that every tenth house happens to be on a corner lot containing larger houses, probably with higher-income residents. The final sample could end up disproportionately composed of upper-class respondents. In this situation, a systematic random sample would not generate a representative group of people.

Multistage or Cluster Sampling

When even larger populations are used from which to select a sample, researchers use what is called a *multistage* or *cluster sampling* method. Large national polling agencies typically employ this method to generate samples. This method involves randomly selecting units beginning with larger clusters and moving to smaller ones at each stage. If, for example, you are interested in surveying Americans' attitudes toward a presidential candidate, you might begin by randomly choosing a certain number of states, either stratified by region or simply taken from the entire list of 50. Then, the next stage is to randomly select counties from within the selected states. The third stage involves randomly choosing cities within the counties now selected. The next stage is randomly generating streets, and the final stage is randomly choosing houses on those streets. At each stage, simple, stratified, or systematic random sampling can be done. This is how pollsters can get away with 1,000 people representing the entire population of the country and why it is unlikely you would know anyone selected for such a survey!

Although this technique is more useful when you have very large national samples, it's possible to do something like this at a big university. Make a list of residence halls and fraternity/sorority houses. The first stage randomly selects from this cluster. Perhaps if the residences are large enough, you can then choose a second cluster from floors, then randomly select rooms, and finally pick roommates at random. Of course, such a sample would not include off-campus students. However, if your study is about residential life, then this random multistage method allows inferences to be made about the entire population of residential students, but not about all students at the university.

These probability sampling methods can be combined; perhaps a stratified systematic random sampling is useful for some studies, or stratified sampling can be used at the various stages in cluster sampling. The key point for any of these is that every element should have a specified (usually equal) chance of being selected into the final sample.

NONPROBABILITY SAMPLING

Sometimes it is just not practical, cost- or time-efficient, or necessary to do a true random sampling. However, it must always be kept in mind that, unless there is a random selection of units from the population, you cannot generalize to the entire population. With nonprobability methods—those for which every element does not have an equal chance of being selected for the study—you are limited to making conclusions about only those who have completed the survey.

There are several kinds of *nonprobability sampling*, including

- Convenience or accidental sampling
- Purposive or judgmental sampling
- Quota sampling
- Snowball sampling.

Convenience or Accidental Sampling

Everyone at one point fills out some sort of questionnaire, whether it's a warranty card for a product recently purchased or as part of an academic study in a class you took in college. Perhaps you have been stopped while shopping at a mall and asked to answer some questions. Maybe you wasted some money texting your vote for your favorite performer on a television show contest or answered a customer survey that popped up after you visited an Internet website. While you may have felt like a good citizen in helping out, what you may not be aware of is that you have been part of a nonprobability and typically nonscientific *accidental* or *convenience sample*, one from which it would be difficult to generate any reliable information about a population of people, despite misleading statements by others using these data to imply that these findings represent what "everybody" believes or does.

All that can be reliably said is that the data resulting from these studies apply only to those who were available and attending that particular psychology class the day surveys were distributed, who had time to text a message on a mobile phone, who had access to the Internet and linked to that webpage, or who were stopped on that day at that hour in that particular shopping mall. To say anything about all college students, all listeners of a radio show, all Internet users, all shoppers, or for that matter all citizens in the country is impossible. But too often these data are inadvertently misused in just that way, and sometimes they are used unethically on purpose.

Convenience samples are based on whoever just happens to be available at a particular moment or accidentally walking by the person distributing the survey. Everyone does not have an equal chance of getting selected. While it certainly is more convenient to go to an intro economics class and get 100 students at once than it is to generate 100 random numbers and stuff mailboxes with questionnaires, the results have limited generalizability.

Volunteer samples, sometimes referred to as "opt-in," or self-selected, respondents, are also samples of convenience. Those who respond to a sign asking for research subjects, participate because of some incentive (a course grade or money), or hear about a survey on some website may be different kinds of people from those who do not even see those announcements. Again, all people do not have an equal chance to select themselves into these studies. And even if they did see calls for survey respondents, those who volunteer may be a very different type of person from those who don't volunteer, such as they need money, have extra time to participate, or are interested in the particular topic of the study.

Purposive or Judgmental Sampling

Sometimes there is a specific reason to choose a unique sample on purpose, because of some characteristics of the units of analysis. *Purposive* or *judgmental sampling* involves designating a group of people for selection because you know they have some traits you want to study. For example, you choose a particular residence hall on campus for study because prior research has shown that it represented the feelings of most students on campus in the past. Many times, visitors to a shopping website get surveyed for the company to find out more about the characteristics of those people who bought something at or searched for information on that online outlet. Marketing researchers test products in a particular city because they have made the judgment that shoppers in that city represent a cross-section of potential buyers nationally. Although people in these kinds of samples are not randomly selected, prior research may have indicated that their patterns of buying were representative of the entire population and could be purposely chosen to make inferences about all potential buyers. In all these situations, every person does not have a specified chance of being selected for the study, only those who are available and purposely chosen.

Quota Sampling

Like stratified random sampling, researchers occasionally want to be sure that there is some representation in the final sample. An accidental sample might involve researchers stopping the first 100 people walking out of a movie theater to survey them about the film. Because it is not a random sample, the number of senior citizens and teenagers leaving first may not be in the same proportion as they were in the theater. If you wanted to guarantee that half are over the age of 65 and half are teenagers, a *quota sampling* would be done in which the first 50 elders and the first 50 teenagers coming out of the theater are surveyed. It is still an accidental sample, but one using a quota system. Breaking your sample into various strata can entail any number of categories, such as race/ethnicity, sex, political preference, and any other demographic characteristic used to screen respondents. Respondents are solicited until the number needed for each of the various criteria is met.

Snowball Sampling

There may be times a study needs respondents who are difficult to find or identify. What if you wanted to do a survey of transgender people? A stratified random sample or a nonprobability quota sample would require that you identify everyone's identity before getting your participants. This is a question people might not want to answer when stopped accidentally in a shopping mall! So what researchers do to get such a sample is first to identify a few transgender people, perhaps through some personal contacts or organizations. Each person who volunteers then is asked to pass along a questionnaire to someone he or she knows who is also transgender, like a snowball rolling down the hill that becomes larger and larger as it picks up more snow. This is why this technique is called *snowball sampling*.

To take another example, let's say you wanted to study housekeepers. This could be an example of a purposive sample since you are purposely focusing on this one particular occupation. But finding respondents may be more difficult than simply doing a random sample of housekeepers. So you survey a few chosen through various channels and then ask them to provide names of other housekeepers they know who would be willing to participate in your study. It helps if people are contacted first by their friends and then give permission to provide their names. You survey them and ask for a list of more names, and so on and on. Clearly, this is not random, and the final sample is made up of networks of people who tend to be somewhat similar in other characteristics (social class, race/ethnicity, etc.), since people tend to pass along questionnaires to those who are their friends. Studies of friendship demonstrate that people's friends are similar to themselves in many basic characteristics like educational level, race/ethnicity, marital status, and income.

CROSS-SECTIONAL AND LONGITUDINAL STUDIES

When a survey is given at one point in time and only once to a particular sample of respondents, it is referred to as a *cross-sectional study*. These are not ideal for uncovering causal relationships that require demonstration of a time sequence for the independent and dependent variables, but they are easier to do and require less time commitment than *longitudinal studies* in which samples of respondents are followed over lengthy periods. One kind of longitudinal study involves following the same people and surveying them at different points in time. This is called a *panel study*, but it has the risk that people cannot be found later on due to moving, death, changing names, or simply lack of interest. Let's say you randomly select a group of 100 graduates from a city's high schools, and then every five years, you contact them to see what they are doing with their lives. But, as it happens, not everyone can be found years later. What may have begun as a random sample, now due to attrition, ends up as a biased one and composed of volunteers who want to participate.

Another version of a longitudinal study is a *cohort study*, which involves sampling, at various periods of time, different people who are similar to the first sample, usually in age. Perhaps researchers surveyed 20-year-olds in 1995, then studied a different group of 30-year-olds in 2005, interviewed a sample of 40-year-olds in 2015, and plan to survey 50-year-olds in 2025. Even though the samples are not composed of the same respondents over the years, they represent a cohort of people who were born the same year and have probably experienced the same major historical events and changes in the culture.

Like public opinion polls, some studies are interested in seeing how the same behaviors or attitudes change over time. Every year since 1975, the University of Michigan's *Monitoring the Future* study continues its survey of high school seniors' drug and alcohol use (see www.monitoringthefuture.org/). This is a *trend study* in which different seniors are surveyed about the same behaviors across different years in order to track changes. These are not the same people followed over time, and they are not those born in the same period or cohort. Figure 5.1 shows the differences in cross-sectional and longitudinal samples.

Figure 5.1 Cross-Sectional Versus Longitudinal Samples

	RESPONDENTS	TIMELINE
Cross-sectional	One set	One time
Longitudinal: panel	One set	Two or more time periods
Longitudinal: trends	Different people	Two or more time periods
Longitudinal: cohort	Different people, shared characteristics	Two or more time periods

SAMPLE SIZE

A common worry is how large a sample should be to have a reliable and valid study. Although mathematical models determining *sample size* are discussed in advanced statistics books (they involve "power analysis" and the calculation of confidence limits based on accuracy at, say, 95 percent or 99 percent confidence levels), there are some points to keep in mind for a beginning researcher. When a population is more homogeneous, fewer elements are required to get a representative sample. The more heterogeneous a population is on a variety of characteristics (such as race/ethnicity, sex, age), the larger a sample is needed to reflect that diversity. Stratified random sampling can achieve representation with a smaller number than simple random sampling would require to ensure representation of the diverse population.

Sample size also depends on what is being studied. If you want to find out information about something that occurs less frequently in the population you are studying, such as bird-watching among 16- to 20-year-olds, then you will need a large sample to find enough respondents. On the other hand, if the behavior and attitudes you are interested in are much more likely to occur (like participating in social media) or large differences or relationships are expected, then a smaller sample size is sufficient.

One simple answer is the larger the sample size, the better, but up to a point. You can cut the margin of error in half but you need to quadruple your sample size to do so. Sampling 1,000 people gives you a sampling error of around 3 percent; getting 4,000 respondents results in 1.5 percent margin of error. Is it worth the expense and time to do so?

Sampling error is greater when making inferences to a large population from a small sample with a large standard deviation, especially if the population is diverse (recall the M&Ms example; see Chapter 6 for more about sampling error and the concept of the *central limit theorem*). But remember, if you are using nonprobability techniques, it doesn't really matter whether 100,000 or 100 people text a message to voice their opinion or cast a vote; the results are limited to those who texted. Generalizing to the entire population is not possible, regardless of the sample size.

Another way of estimating sample size is to consider the kinds of analyses you plan to do. If you are interested in comparing subgroups, such as sex differences within five categories of ethnicity/race, then you will need enough people to fill ten different categories—five ethnicities by two genders equals ten categories: Asian/Pacific women, Asian/Pacific men, American Indian men, American Indian women, and so on. You are not going to be able to do much in the way of sophisticated statistical analysis if you do not have at least five people in each of those ten categories. Already you need 50 respondents, assuming they fall equally into these categories. If you use stratified random sampling you might be able to achieve this, but it may take hundreds of people selected by simple random sampling to get at least five in each category.

Given the limited time and money most of us have to do a short survey, especially for class projects or theses, consider giving out 100 surveys using random sampling techniques with the hope that half will reply (after some e-mail or postcard reminders) in order to get a small sample size of 50 respondents. Usually 20 to 30 percent of people who receive questionnaires return them right away. Reminders can bring that percentage up to 50 percent or more. Response rates under 60 or 70 percent may compromise the integrity of the random sample. See Dillman et al.'s (2014) popular "tailored design method" for ways to improve response rates by focusing on a total overview of implementing procedures and designing surveys to motivate people to participate.

Sometimes it is necessary to supplement randomly distributed questionnaires using nonprobability methods in order to increase sample size. However, you must be sure to code them in some way to see if the responses from this group are different from the randomly chosen ones. In any event, once you introduce nonprobability sampling, you are limited to making conclusions only about those who filled out and returned the survey. For most small surveys, this might be all you really need.

How you distribute your questionnaires can affect the quality of the sampling. Mailing them with return postage and envelopes requires some funding, but prepaid postage increases response rates. Giving them to people and asking them to send the surveys with their own postage results in fewer returned ones, especially for those on limited incomes. Dropping by in person and picking surveys up from respondents later might be a better way of getting them back if this is convenient to do and does not compromise anonymity or confidentiality. Online and e-mailed surveys are easier to distribute, but this type of distribution can also result in a nonrandom sample limited to those people with access to and comfort with using computers.

Following up with reminders (regular mail, e-mail, phone calls) one week after distribution, for example, and another reminder a week or ten days later, increases response rates. But when names or identifying numbers are not used, you have to send reminders to everyone, including those who already returned their surveys. It is essential that the reminder notice thank those who have already sent their forms back.

Distributing questionnaires in large settings, such as a classroom, and waiting for them to be completed generates a much better response, since nearly everyone fills them out in such settings, but the nonprobability volunteer sample may not be ideal. In general, it is considered good if you get around 50 percent of questionnaires completed in the first phase, but don't be surprised if the response rate is half that. After several reminders, it is very good if your response rate is over 70 percent.

At this point in the research journey, you should have a clear set of research questions or hypotheses guiding the development of a clearly written valid and reliable questionnaire, a sample ready and excited to take your survey, and a method for distributing and collecting the forms. Now you eagerly await their completion and make preparations for coding the data for statistical analyses (see Chapter 4). The next four chapters discuss various ways of presenting results and analyzing your data statistically.

REVIEW: WHAT DO THESE KEY TERMS MEAN?

Cohort studies
Convenience or accidental sampling
Cross-sectional studies
Longitudinal studies
Multistage or cluster sampling
Nonprobability sampling
Panel studies
Population
Population parameter
Probability sampling
Purposive or judgmental sampling
Quota sampling
Random digit dialing
Random sample (simple)
Sample size
Sample statistic
Sampling error or margin of error
Sampling frame
Snowball sampling
Stratified random sample
Systematic random sample
Trend studies
Unit of analysis or element
Volunteers

TEST YOURSELF

For each of the following studies, state what method of sampling was used.

1. Researcher surveyed the first 25 sophomores and the first 25 seniors who entered the cafeteria during lunch.

2. As part of their requirements for a course, all the students in "Introduction to Social Work" completed a questionnaire.
3. The Human Resources director divided the number of people she wanted to survey in the organization by the total number of people in the population and then chose every *n*th person to get the survey.
4. Members of the local motorcycle club completed questionnaires and were then asked to give questionnaires to other people they knew who owned motorcycles, who were also asked to give some questionnaires to their motorcycle friends.

INTERPRET: WHAT DO THESE REAL EXAMPLES TELL US?

For the following examples from actual research, discuss what sampling strategy is used and what the limitations and strengths might be with each type of sample:

1. For a study of "smartphone addiction" and academic performance in Lebanon, researchers created a sample "based on voluntary participation of university students without any gender, socioeconomic, or nationality restrictions. Systematic random sampling was implemented by randomly picking the first student from the student population ordered by student identification number and then selecting each 3rd student from the list." (Hawi and Samaha 2016: 83–84)
2. A study of body image perception of adolescents in Nigeria used multistage sampling to first randomly select one of three local government areas in Benin City. In that area, "schools were stratified into two groups to ensure appropriate representation of students from both public and private schools." Then, for Stage 2, 14 schools were selected randomly proportionate to the ratio of public to private schools. At Stage 3, "The number of students selected from each school was determined by the population of the school using equal sampling ratio" and at Stage 4, "The final sampling unit was arrived at by a systematic random sampling, an equal probability method" where every *n*th person was selected after randomly choosing the first student. (Otakpor and Ehimigbai 2016: 72–73)
3. For a study on ethnic identity as a predictor of attitudes toward blacks, whites, and Hispanic lesbian, gay, and bisexual people, researchers used quota sampling: "e-mail invitations were sent out nationally to participants who were offered $2 in exchange for their participation . . . . Quota sampling in the context of this study refers to a sampling method in which a predetermined number of participants are selected, the number being 300 for each racial group (i.e., Black, White, and Hispanic) for this survey. Once the number of participants exceeded approximately 300 for each group, the survey was terminated." (Elias et al. 2017: 10)

4. In order to interview fans of the World Cup and why they supported teams when living outside their country of origin, researchers developed this sampling strategy: "Beginning from our personal networks at a large research university that attracts people from all over the world, we conducted a snowball sample, our recruitment materials seeking 'people in diaspora who are interested in football/soccer in general and/or the World Cup in particular.' With this purposive sample, we recruited people self-identified as living outside their country of origin." (Stanfill and Valdivia 2017: 105)
5. To investigate whether participation in extracurricular activities relates to academic motivation among disadvantaged high school students, "a stratified sample of 3000 grade 7 to grade 10 high school students from disadvantaged neighborhoods in the province of Quebec, Canada, was formed based on sex, grade level, and region." (Denault and Guay 2017: 97)

CONSULT: WHAT COULD BE DONE?

A study was completed at a university focusing on students' use of alcohol and drugs. Questionnaires were left in everyone's mailboxes, and students were told to return them to a particular location by a certain date. Almost 25 percent of the students returned the surveys, and the results were printed in the campus newspaper. The author of the article concluded that an extremely high percentage of students at the school drank alcohol or smoked marijuana in the past month, despite anecdotal evidence that the rates were not that extreme. You are asked to consult and explain more about the sampling.

1. What might be going on here?
2. What could you do to remedy any problems?
3. Describe two other sampling methods you would recommend instead and the strengths and weaknesses of each.

DECIDE: WHAT DO YOU DO NEXT?

For your study on how diverse people develop and maintain friendships, especially on social media, respond to the following items:

1. Design a sampling strategy that involves a purposive method. Whom would you survey and why?
2. How would you go about getting a random sample of people in your local community and online?

3. Describe several ways of getting an accidental or convenience sample for this study.
4. In each case, list the strengths and weaknesses of the sampling design.
5. If you were to do a study of friendships as depicted in television sitcoms, and you wanted to do a content analysis of the shows themselves, describe how you would go about designing a random sample of television shows.
6. If you are doing your own research project, select a sampling method, describe how you plan to survey respondents, and generate an actual list of respondents, mailbox numbers, or other relevant information.