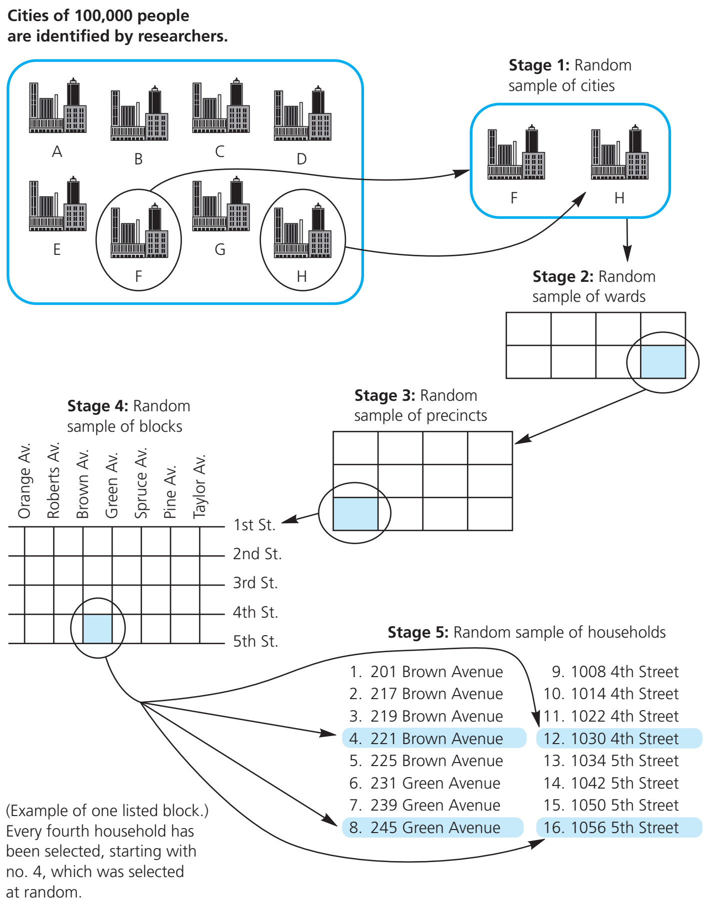
6 Sampling



MULTISTAGE SAMPLING

OBJECTIVES Studying this chapter should enable you to:

- Distinguish between a sample and a population.
- Explain what is meant by the term "representative sample."
- Explain how a target population differs from an accessible population.
- Explain what is meant by "random sampling," and describe briefly three ways of obtaining a random sample.
- Use a table of random numbers to select a random sample from a population.
- Explain how stratified random sampling differs from cluster random sampling.
- Explain what is meant by "systematic sampling," "convenience sampling," and "purposive sampling."
- Explain how the size of a sample can make a difference in terms of representativeness of the sample.
- Explain what is meant by the term "external validity."
- Distinguish between population generalizability and ecological generalizability and discuss when it is (and when it is not) appropriate to generalize the results of a study.

INTERACTIVE AND APPLIED LEARNING After, or while, reading this chapter:

Go to the Online Learning Center at www.mhhe.com/fraenkel8e to:

- Learn More About Sampling and Representativeness

Go to your online Student Mastery Activities book to do the following activities:

- Activity 6.1: Identifying Types of Sampling
- Activity 6.2: Drawing a Random Sample
- Activity 6.3: When Is It Appropriate to Generalize?
- Activity 6.4: True or False?
- Activity 6.5: Stratified Sampling
- Activity 6.6: Designing a Sampling Plan

Rosa Pak, a research professor at a large eastern university, wishes to study the effect of a new mathematics program on the mathematics achievement of students who are performing poorly in math in elementary schools throughout the United States. Because of a number of factors, of which time and money are only two, it is impossible for Rosa and her colleagues to try out the new program with the entire population of such students. They must select a *sample*. What is a sample anyway? Are there different kinds of samples? Are some kinds better than others to study? And just how does one go about obtaining a sample in the first place? Answers to questions like these are some of the things you will learn about in this chapter.

When we want to know something about a certain group of people, we usually find a few members of the group whom we know—or don't know—and study them. After we have finished "studying" these individuals, we usually come to some conclusions about the larger group of which they are a part. Many "commonsense" observations, in fact, are based on observations of relatively few people. It is not uncommon, for example, to hear statements such as: "Most female students don't like math"; "You won't find very many teachers voting Republican"; and "Most school superintendents are men."

What Is a Sample?

Most people, we think, base their conclusions about a group of people (students, Republicans, football players, actors, and so on) on the experiences they have with a fairly small number, or **sample**, of individual members. Sometimes such conclusions are an accurate representation of how the larger group of people acts or what they believe, but often they are not. It all depends on how representative (i.e., how similar) the sample is of the larger group.

One of the most important steps in the research process is the selection of the sample of individuals who will participate (be observed or questioned). **Sampling** refers to the process of selecting these individuals.

SAMPLES AND POPULATIONS

A sample in a research study is the group on which information is obtained. The larger group to which one hopes to apply the results is called the **population**.* All 700 (or whatever total number of) students at State University who are majoring in mathematics, for example, constitute a population; 50 of those students constitute a sample. Students who own automobiles make up another population, as do students who live in the campus dormitories. Notice that a group may be both a sample in one context and a population in another context. All State University students who own automobiles constitute the population of automobile owners at State, yet they also constitute a sample of all automobile owners at state universities across the United States.

When it is possible, researchers would prefer to study the entire population of interest. Usually, however, this is difficult to do. Most populations of interest are large, diverse, and scattered over a large geographic area. Finding, let alone contacting, all the members can be time-consuming and expensive. For that reason, of necessity,

*In some instances the sample and population may be identical.

researchers often select a sample to study. Some examples of samples selected from populations follow:

- A researcher is interested in studying the effects of diet on the attention span of third-grade students in a large city. There are 1,500 third-graders attending the elementary schools in the city. The researcher selects 150 of these third-graders, 30 each in five different schools, as a sample for study.
- An administrator in a large urban high school is interested in student opinions on a new counseling program in the district. There are six high schools and some 14,000 students in the district. From a master list of all students enrolled in the district schools, the administrator selects a sample of 1,400 students (350 from each of the four grades, 9–12) to whom he plans to mail a questionnaire asking their opinion of the program.
- The principal of an elementary school wants to investigate the effectiveness of a new U.S. history textbook used by some of the teachers in the district. Out of a total of 22 teachers who are using the text, she selects a sample of 6. She plans to compare the achievement of the students in these teachers' classes with those of another 6 teachers who are not using the text.

DEFINING THE POPULATION

The first task in selecting a sample is to define the population of interest. In what group, exactly, is the researcher interested? To whom does he or she want the results of the study to apply? The population, in other words, is the group of interest to the researcher, the group to whom the researcher would like to generalize the results of the study. Here are some examples of populations:

- All high school principals in the United States
- All elementary school counselors in the state of California
- All students attending Central High School in Omaha, Nebraska, during the academic year 2005–2006
- All students in Ms. Brown's third-grade class at Wharton Elementary School

The above examples reveal that a population can be any size and that it will have at least one (and sometimes several) characteristic(s) that sets it off from any other population. Notice that a population is always *all* of the individuals who possess a certain characteristic (or set of characteristics).

In educational research, the population of interest is usually a group of persons (students, teachers, or other individuals) who possess certain characteristics. In some cases, however, the population may be defined as a group of classrooms, schools, or even facilities. For example,

- All fifth-grade classrooms in Delaware (the hypothesis might be that classrooms in which teachers display a greater number and variety of student products have higher achievement)
- All high school gymnasiums in Nevada (the hypothesis might be that schools with "better" physical facilities produce more winning teams)

TARGET VERSUS ACCESSIBLE POPULATIONS

Unfortunately, the actual population (called the **target population**) to which a researcher would really like to generalize is rarely available. The population to which a researcher is *able* to generalize, therefore, is the **accessible population**. The former is the researcher's ideal choice; the latter, his or her realistic choice. Consider these examples:

Research problem to be investigated: The effects of computer-assisted instruction on the reading achievement of first- and second-graders in California.

Target population: All first- and second-grade children in California.

Accessible population: All first- and second-grade children in the Laguna Salada elementary school district of Pacifica, California.

Sample: Ten percent of the first- and second-grade children in the Laguna Salada district in Pacifica, California.

Research problem to be investigated: The attitudes of fifth-year teachers-in-training toward their student-teaching experience.

Target population: All fifth-year students enrolled in teacher-training programs in the United States.

Accessible population: All fifth-year students enrolled in teacher-training programs in the State University of New York.

Sample: Two hundred fifth-year students selected from those enrolled in the teacher-training programs in the State University of New York.

The more narrowly researchers define the population, the more they save on time, effort, and (probably) money, but the more they limit generalizability. It is essential that researchers describe the population and the sample in sufficient detail so that interested individuals

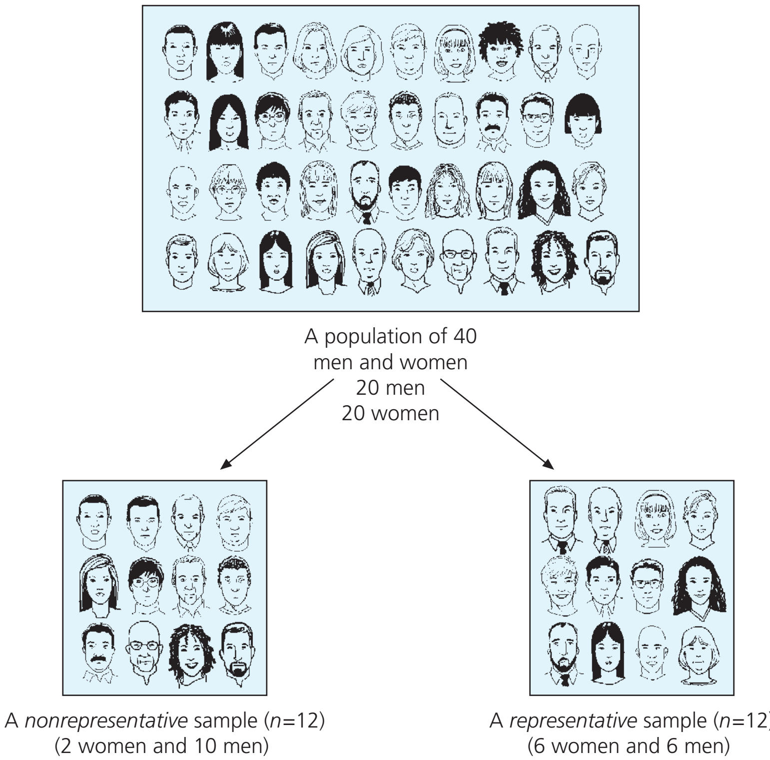


Figure 6.1
Representative Versus Nonrepresentative Samples

can determine the applicability of the findings to their own situations. Failure to define in detail the population of interest, and the sample studied, is one of the most common weaknesses of published research reports. It is important to note that the actual sample may be different from the sample originally selected because some subjects may refuse to participate, some subjects may drop out, data may be lost, and the like. We repeat, therefore, that it is very important to describe the characteristics of the actual sample studied in some detail.

RANDOM VERSUS NONRANDOM SAMPLING

Following is an example of each of the two main types of sampling.

Random sampling: The dean of a school of education in a large midwestern university wishes to find out how her faculty feel about the current sabbatical leave requirements at the university. She places all 150 names of the faculty in a hat, mixes them thoroughly, and then draws out the names of 25 individuals to interview.*

Nonrandom sampling: The president of the same university wants to know how his junior faculty feel about a promotion policy that he has recently introduced (with the advice of a faculty committee). He selects a sample of 30 from the total faculty of 1,000 to talk with. Five faculty members from each of the six schools that make up the university are chosen on the basis of the following criteria: They have taught at the university for less than five years, they are nontenured, they belong to one of the faculty associations on campus, and they have not been a member of the committee that helped the president draft the new policy.

In the first example, 25 names were selected from a hat after all the names had been mixed thoroughly. This is called **random sampling** because every member of the population (the 150 faculty members in the school) presumably had an equal chance of being selected. There are more sophisticated ways of drawing a random sample, but they all have the same intent—to select a *representative* sample from the population (Figure 6.1). The basic idea is that the group of individuals selected is very much like the entire population.

*A better way to do this will be discussed shortly, but this gives you the idea.

One can never be sure of this, of course, but if the sample is selected randomly and is sufficiently large, a researcher should get an accurate view of the larger group. The best way to ensure this is to see that no bias enters the selection process—that the researcher (or other factors) cannot consciously or unconsciously influence who gets chosen to be in the sample. We explain more about how to minimize bias later in this chapter.

In the second example, the president wants representativeness, but not as much as he wants to make sure there are certain kinds of faculty in his sample. Thus, he has stipulated that each of the individuals selected must possess all the criteria mentioned. Each member of the population (the entire faculty of the university) does *not* have an equal chance of being selected; some, in fact, have *no* chance. Hence, this is an example of **nonrandom sampling**, sometimes called purposive sampling (see p. 100). Here is another example of a random sample contrasted with a nonrandom sample.

Random: A researcher wishes to conduct a survey of all social studies teachers in a midwestern state to determine their attitudes toward the new state guidelines for teaching history in the secondary schools. There are a total of 725 social studies teachers in the state. The names of these teachers are obtained and listed alphabetically. The researcher then numbers the names on the list from 001 to 725. Using a table of random numbers, which he finds in a statistics textbook, he selects 100 teachers for the sample.

Nonrandom: The manager of the campus bookstore at a local university wants to find out how students feel about the services the bookstore provides. Every day for two weeks during her lunch hour, she asks every person who enters the bookstore to fill out a short questionnaire she has prepared and drop it in a box near the entrance before leaving. At the end of the two-week period, she has a total of 235 completed questionnaires.

In the second example, notice that all bookstore users did not have an equal chance of being included in the sample, which included only those who visited during the lunch hour. That is why the sample is not random. Notice also that some may not have completed the questionnaire.

Random Sampling Methods

After making a decision to sample, researchers try hard, in most instances, to obtain a sample that is representative of the population of interest—that means they prefer random sampling. The three most common ways of obtaining this type of sample are simple random sampling, stratified random sampling, and cluster sampling. A less common method is two-stage random sampling.

SIMPLE RANDOM SAMPLING

A **simple random sample** is one in which each and every member of the population has an equal and independent chance of being selected. If the sample is large, this method is the best way yet devised to obtain a sample representative of the population of interest. Let's take an example: Define a population as all eighth-grade students in school district Y. Imagine there are 500 students. If you were one of these students, your chance of being selected would be 1 in 500, if the sampling procedure were indeed random. Everyone would have the same chance of being selected.

The larger a random sample is in size, the more likely it is to represent the population. Although there is no guarantee of representativeness, of course, the likelihood of it is greater with large random samples than with any other method. Any differences between the sample and the population should be small and unsystematic. Any differences that do occur are the result of chance, rather than bias on the part of the researcher.

The key to obtaining a random sample is to ensure that each and every member of the population has an equal and independent chance of being selected. This can be done by using what is known as a **table of random numbers**—an extremely large list of numbers that has no order or pattern. Such lists can be found in the back of most statistics books. Table 6.1 illustrates part of a typical table of random numbers.

For example, to obtain a sample of 200 from a population of 2,000 individuals, using such a table, select a column of numbers, start anywhere in the column, and begin reading four-digit numbers. (Why four digits? Because the final number, 2,000, consists of four digits, and we must always use the same number of digits for each person. Person 1 would be identified as 0001; person 2, as 0002; person 635, as 0635; and so forth.) Then proceed to write down the first 200 numbers in the column that have a value of 2,000 or less.

TABLE 6.1 *Part of a Table of Random Numbers*

011723	223456	222167	032762	062281	565451
912334	379156	233989	109238	934128	987678
086401	016265	411148	251287	602345	659080
059397	022334	080675	454555	011563	237873
666278	106590	879809	899030	909876	198905
051965	004571	036900	037700	500098	046660
063045	786326	098000	510379	024358	145678
560132	345678	356789	033460	050521	342021
727009	344870	889567	324588	400567	989657
000037	121191	258700	088909	015460	223350
667899	234345	076567	090076	345121	121348
042397	045645	030032	657112	675897	079326
987650	568799	070070	143188	198789	097451
091126	021557	102322	209312	909036	342045

Let us take the first column of four numbers in Table 6.1 as an example. Reading only the first four digits, look at the first number in the column: It is 0117, so number 117 in the list of individuals in the population would be selected for the sample. Look at the second number: It is 9123. There is no 9123 in the population (because there are only 2,000 individuals in the entire population). So go on to the third number: It is 0864, hence number 864 in the list of individuals in the population would be chosen. The fourth number is 0593, so number 593 gets selected. The fifth number is 6662. There is no 6662 in the population, so go on to the next number, and so on, until reaching a total of 200 numbers, each representing an individual in the population who will be selected for the sample. Most researchers use computer-generated lists to obtain their samples randomly. This can be done quite easily using EXCEL software (see the box entitled "Using EXCEL to Draw a Random Sample" in Chapter 11 on p. 235).

The advantage of random sampling is that, if large enough, it is very likely to produce a representative sample. Its biggest disadvantage is that it is not easy to do. Each and every member of the population must be identified. In most cases, we must be able to contact the individuals selected. In all cases, we must know *who* 117 (for example) is.

Furthermore, simple random sampling is not used if researchers wish to *ensure* that certain subgroups are present in the sample in the same proportion as they are in the population. To do this, researchers must engage in what is known as stratified sampling.

STRATIFIED RANDOM SAMPLING

Stratified random sampling is a process in which certain subgroups, or *strata,* are selected for the sample in the same proportion as they exist in the population. Suppose the director of research for a large school district wants to find out student response to a new twelfth-grade American government textbook the district is considering adopting. She intends to compare the achievement of students using the new book with that of students using the more traditional text the district has purchased in the past. Since she has reason to believe that gender is an important variable that may affect the outcomes of her study, she decides to ensure that the proportion of males and females in the study is the same as in the population. The steps in the sampling process would be as follows:

1. She identifies the target (and accessible) population: all 365 twelfth-grade students enrolled in American government courses in the district.
2. She finds that there are 219 females (60 percent) and 146 males (40 percent) in the population. She decides to have a sample made up of 30 percent of the target population.
3. Using a table of random numbers, she then randomly selects 30 percent *from each stratum* of the population, which results in 66 female (30 percent of 219) and 44 male (30 percent of 146) students being selected from these subgroups. The proportion of males and females is the same in both the population and sample—40 and 60 percent (Figure 6.2).

The advantage of stratified random sampling is that it increases the likelihood of representativeness, especially if one's sample is not very large. It virtually ensures that key characteristics of individuals in the population are included in the same proportions in the sample. The disadvantage is that it requires more effort on the part of the researcher.

CLUSTER RANDOM SAMPLING

In both random and stratified random sampling, researchers want to make sure that certain kinds of individuals are included in the sample. But there are times when it is not possible to select a sample of individuals from a population. Sometimes, for example, a list of all members of the population of interest is not available. Obviously, then, simple random or stratified random sampling cannot be used. Frequently, researchers cannot select a sample of individuals due to administrative or other restrictions.

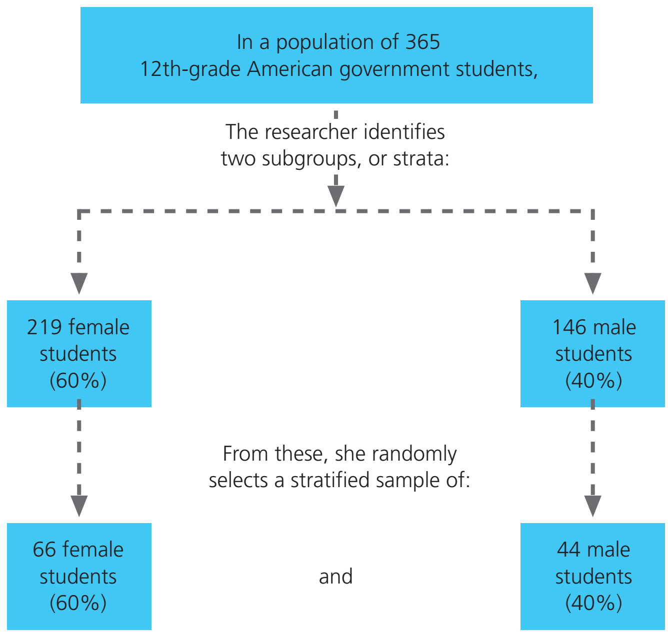


Figure 6.2 *Selecting a Stratified Sample*

This is especially true in schools. For example, if a target population were all eleventh-grade students within a district enrolled in U.S. history courses, it would be unlikely that the researcher could pull out randomly selected students to participate in an experimental curriculum. Even if it could be done, the time and effort required would make such selection difficult. About the best the researcher could hope for would be to study a number of intact classes—that is, classes already in existence. The selection of groups, or clusters, of subjects rather than individuals is known as **cluster random sampling**. Just as simple random sampling is more effective with larger numbers of individuals, cluster random sampling is more effective with larger numbers of clusters.

Let us consider another example of cluster random sampling. The superintendent of a large unified school district in a city on the East Coast wants to obtain some idea of how teachers in the district feel about merit pay. There are 10,000 teachers in all the elementary and secondary schools of the district, and there are 50 schools distributed over a large area. The superintendent does not have the funds to survey all teachers in the district, and he needs the information about merit pay quickly. Instead of randomly selecting a sample of teachers from every school, therefore, he decides to interview all the teachers in selected schools. The teachers in each school, then, constitute a cluster. The superintendent assigns a number to each school and then uses a table of random numbers to select 10 schools (20 percent of the population). All the teachers in the selected schools then constitute the sample. The interviewer questions all the teachers at each of these 10 schools, rather than having to travel to all the schools in the district. If these teachers do represent the remaining teachers in the district, then the superintendent is justified in drawing conclusions about the feelings of the entire population of teachers in his district about merit pay. It is possible that this sample is not representative, of course. Because the teachers to be interviewed all come from a small number of schools in the district, it might be the case that these schools differ in some ways from the other schools in the district, thereby influencing the views of the teachers in those schools with regard to merit pay. The more schools selected, the more likely the findings will be applicable to the population of teachers (Figure 6.3).

Cluster random sampling is similar to simple random sampling except that groups rather than individuals are randomly selected (that is, the sampling unit is a group rather than an individual). The advantages of cluster random sampling are that it can be used when it is difficult or impossible to select a random sample of individuals, it is often far easier to implement in schools, and it is frequently less time-consuming. Its disadvantage is that there is a far greater chance of selecting a sample that is not representative of the population.

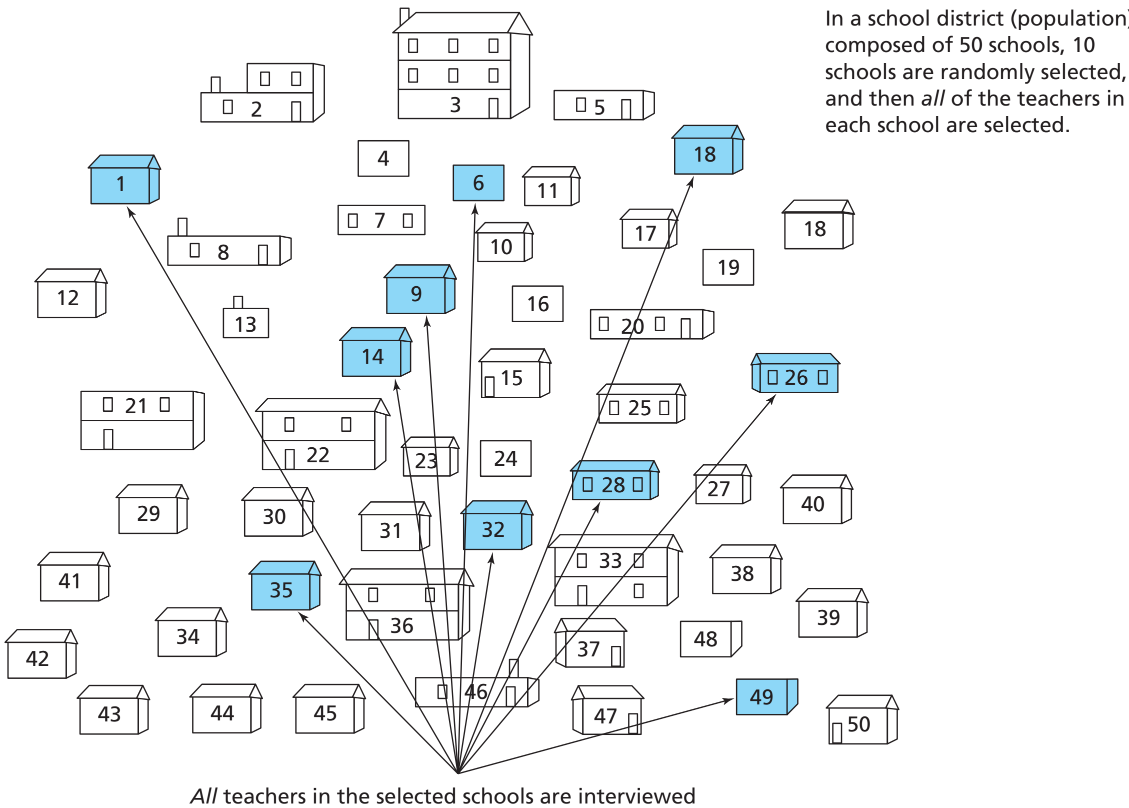


Figure 6.3 *Cluster Random Sampling*

Many beginning researchers make a common error with regard to cluster random sampling: randomly selecting only *one* cluster as a sample and then observing or interviewing all individuals within that cluster. Even if there is a large number of individuals within the cluster, it is the cluster that has been randomly selected, rather than individuals; hence the researcher is not entitled to draw conclusions about a target population of such individuals. Yet some researchers do draw such conclusions. We repeat, they should not.

TWO-STAGE RANDOM SAMPLING

It is often useful to combine cluster random sampling with individual random sampling. This is accomplished by **two-stage random sampling**. Rather than randomly selecting 100 students from a population of 3,000 ninth-graders located in 100 classes, the researcher might decide to select 25 classes randomly from the population of 100 classes and then randomly select 4 students from each class. This is much less time-consuming than visiting most of the 100 classes. Why would this be better than using all the students in four randomly selected classes? Because four classes would be too few to ensure representativeness, even though they were selected randomly.

Figure 6.4 illustrates the different random sampling methods we have discussed.

Nonrandom Sampling Methods

SYSTEMATIC SAMPLING

In **systematic sampling**, every *n*th individual in the population list is selected for inclusion in the sample. For example, in a population list of 5,000 names, to select a sample of 500, a researcher would select every tenth name on the list until reaching a total of 500 names. Here is an example of this type of sampling: The principal of a large middle school (grades 6–8) with 1,000 students wants to know how students feel about the new menu in the school cafeteria. She obtains an alphabetical list of all students in the school and selects every tenth student on the list to be in the sample.

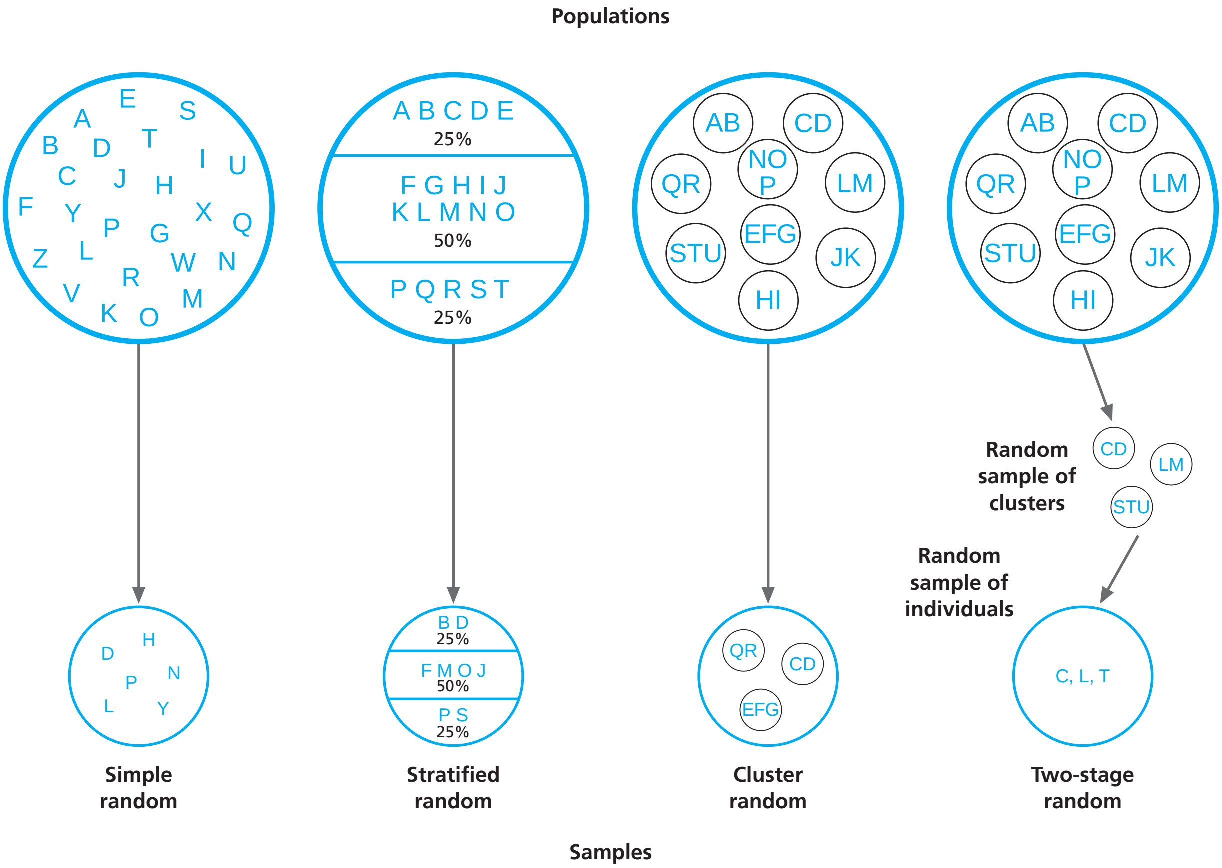


Figure 6.4 *Random Sampling Methods*

To guard against bias, she puts the numbers 1 to 10 into a hat and draws one out. It is a 3. So she selects the students numbered 3, 13, 23, 33, 43, and so on until she has a sample of 100 students to be interviewed.

The above method is technically known as systematic sampling with a **random start**. In addition, there are two terms that are frequently used when referring to systematic sampling. The **sampling interval** is the distance in the list between each of the individuals selected for the sample. In the example given above, it was 10. A simple formula to determine it is:

$$\frac{\text{Population size}}{\text{Desired sample size}}$$

The **sampling ratio** is the proportion of individuals in the population that is selected for the sample. In the example above, it was .10, or 10 percent. A simple way to determine the sampling ratio is:

$$\frac{\text{Sample size}}{\text{Population size}}$$

There is a danger in systematic sampling that is sometimes overlooked. If the population has been ordered systematically—that is, if the arrangement of individuals on the list is in some sort of pattern that accidentally coincides with the sampling interval—a markedly biased sample can result. This is sometimes called **periodicity**. Suppose that the middle school students in the preceding example had not been listed alphabetically but rather by homeroom and that the homeroom teachers had previously listed the students in their rooms by grade point average, high to low. That would mean that the better students would be at the top of each homeroom list. Suppose also that each homeroom had 30 students. If the principal began her selection of every tenth student with the first or second or third student on the list, her sample would consist of the better students in the school rather than a representation of the entire student body. (Do you see why? Because in each homeroom, the poorest students would be those who were numbered between 24 and 30, and they would never get chosen.)

When planning to select a sample from a list of some sort, therefore, researchers should carefully examine the list to make sure there is no cyclical pattern present. If the list has been arranged in a particular order, researchers should make sure the arrangement will not bias the sample in some way that could distort the results. If such seems to be the case, steps should be taken to ensure representativeness—for example, by randomly selecting individuals from each of the cyclical portions. In fact, if a population list is randomly ordered, a systematic sample drawn from the list is a random sample.

CONVENIENCE SAMPLING

Many times it is extremely difficult (sometimes even impossible) to select either a random or a systematic nonrandom sample. At such times, a researcher may use **convenience sampling**. A convenience sample is a group of individuals who (conveniently) are available for study (Figure 6.5). Thus, a researcher might decide to study two third-grade classes at a nearby elementary school because the principal asks for help in evaluating the effectiveness of a new spelling textbook. Here are some examples of convenience samples:

- To find out how students feel about food service in the student union at an East Coast university, the manager stands outside the main door of the cafeteria one Monday morning and interviews the first 50 students who walk out of the cafeteria.
- A high school counselor interviews all the students who come to him for counseling about their career plans.
- A news reporter for a local television station asks passersby on a downtown street corner their opinions about plans to build a new baseball stadium in a nearby suburb.
- A university professor compares student reactions to two different textbooks in her statistics classes.

In each of the above examples, a certain group of people was chosen for study because they were available. The obvious advantage of this type of sampling is

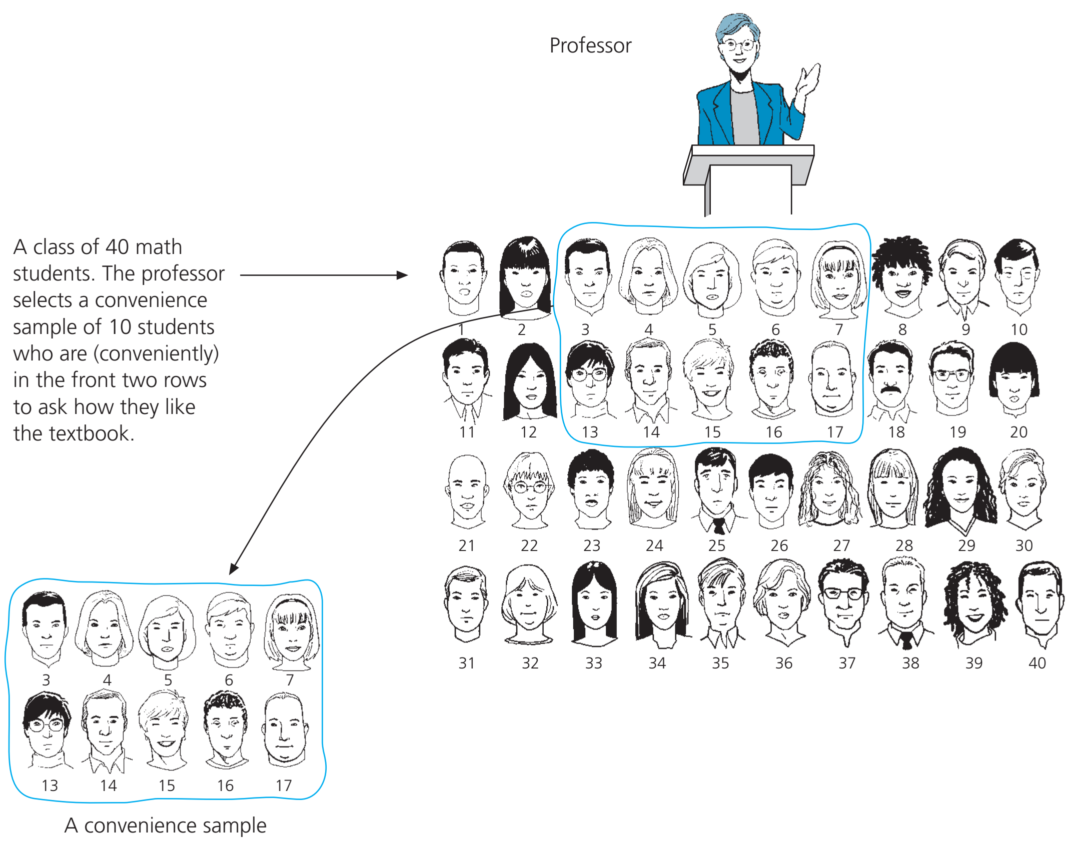


Figure 6.5 *Convenience Sampling*

convenience. But just as obviously, it has a major disadvantage in that the sample will quite likely be biased. Take the case of the TV reporter who is interviewing passersby on a downtown street corner. Many possible sources of bias exist. First of all, of course, anyone who is not downtown that day has no chance to be interviewed. Second, those individuals who are unwilling to give their views will not be interviewed. Third, those who agree to be interviewed will probably be individuals who hold strong opinions one way or the other about the stadium. Fourth, depending on the time of day, those who are interviewed quite possibly will be unemployed or have jobs that do not require them to be indoors. And so forth.

In general, convenience samples cannot be considered representative of any population and should be avoided if at all possible. Unfortunately, sometimes they are the only option a researcher has. When such is the case, the researcher should be especially careful to include information on demographic and other characteristics of the sample studied. The study should also be *replicated,* that is, repeated, with a number of similar samples to decrease the likelihood that the results obtained were simply a one-time occurrence. We will discuss replication in more depth later in the chapter.

PURPOSIVE SAMPLING

On occasion, based on previous knowledge of a population and the specific purpose of the research, investigators use personal judgment to select a sample. Researchers assume they can use their knowledge of the population to judge whether or not a particular sample will be representative. Here are some examples:

- An eighth-grade social studies teacher chooses the 2 students with the highest grade point averages in her class, the 2 whose grade point averages fall in the middle of the class, and the 2 with the lowest grade point averages to find out how her class feels about including a discussion of current events as a regular part of classroom activity. Similar samples in the past have represented the viewpoints of the total class quite accurately.
- A graduate student wants to know how retired people age 65 and over feel about their "golden years." He has been told by one of his professors, an expert on aging and the aged population, that the local Association of Retired Workers is a representative cross section of retired people age 65 and over. He decides to interview a sample of 50 people who are members of the association to get their views.

In both of these examples, previous information led the researcher to believe that the sample selected would be representative of the population. There is a second form of purposive sampling in which it is not expected that the persons chosen are themselves representative of the population, but rather that they possess the necessary information *about* the population. For example:

- A researcher is asked to identify the unofficial power hierarchy in a particular high school. She decides to interview the principal, the union representative, the principal's secretary, and the school custodian because she has prior information that leads her to believe they are the people who possess the information she needs.
- For the past five years, the leaders of the teachers' association in a midwestern school district have represented the views of three-fourths of the teachers in the district on most major issues. This year, therefore, the district administration decides to interview just the leaders of the association rather than select a sample from all the district's teachers.

Purposive sampling is different from convenience sampling in that researchers do not simply study whoever is available but rather use their judgment to select a sample that they believe, based on prior information, will provide the data they need. The major disadvantage of purposive sampling is that the researcher's judgment may be in error—he or she may not be correct in estimating the representativeness of a sample or their expertise regarding the information needed. In the second example above, this year's leaders of the teachers' association may hold views markedly different from those of their members. Figure 6.6 illustrates the methods of convenience, purposive, and systematic sampling.

A Review of Sampling Methods

Let us illustrate each of the previous sampling methods using the same hypothesis: "Students with low self-esteem demonstrate lower achievement in school subjects."

Target population: All eighth-graders in California.
Accessible population: All eighth-graders in the San Francisco Bay Area (seven counties).
Feasible sample size: $n = 200 - 250$.

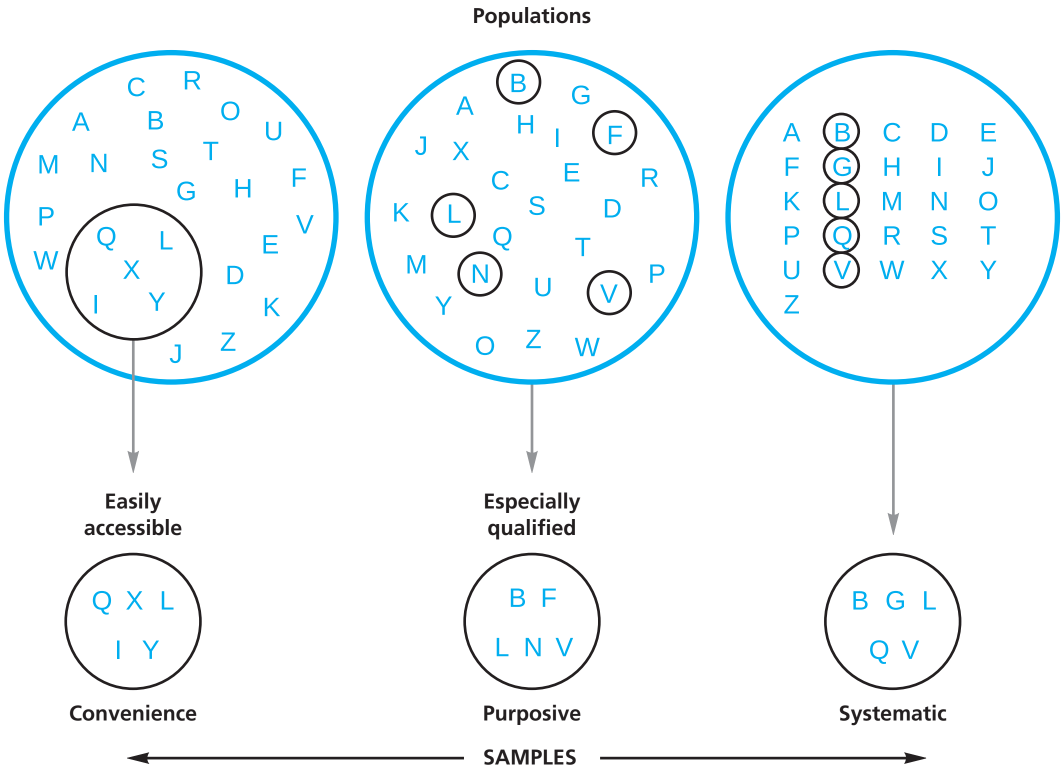


Figure 6.6 *Nonrandom Sampling Methods*

Simple random sampling: Identify all eighth-graders in all public and private schools in the seven counties (estimated number of eighth-grade students = 9,000). Assign each student a number, and then use a table of random numbers to select a sample of 200. The difficulty here is that it is time-consuming to identify every eighth-grader in the Bay Area and to contact (probably) about 200 different schools in order to administer instruments to one or two students in those schools.

Cluster random sampling: Identify all public and private schools having an eighth grade in the seven counties. Assign each of the schools a number, and then randomly select four schools and include all eighth-grade classes in each school. (We would estimate 2 classes per school × 30 students per class × 4 schools = a total of 240 students.) Cluster random sampling is much more feasible than simple random sampling to implement, but it is limited because of the use of only four schools, even though they are to be selected randomly. For example, the selection of only four schools may exclude the selection of private-school students.

Stratified random sampling: Obtain data on the number of eighth-grade students in public versus private schools and determine the proportion of each type (e.g., 80 percent public, 20 percent private). Determine the number from each type to be sampled: public = 80 percent of 200 = 160; private = 20 percent of 200 = 40. Randomly select samples of 160 and 40 students from respective subpopulations of public and private students. Stratification may be used to ensure that the sample is representative on other variables as well. The difficulty with this method is that stratification requires that the researcher know the proportions in each stratum of the population, and it also becomes increasingly difficult as more variables are added. Imagine trying to stratify not only on the public-private variable but also (for example) on student ethnicity, gender, and socioeconomic status, and on teacher gender and experience.

Two-stage random sampling: Randomly select 25 schools from the accessible population of schools, and then randomly select 8 eighth-grade students from each school ($n = 8 \times 25 = 200$). This

CONTROVERSIES IN RESEARCH

Sample or Census?

Samples look at only part of the population. A *census* tries to look at the entire population. The U.S. Census Bureau, charged with conducting the United States Census every 10 years, estimated that the 2000 census missed 1–2 percent of the population—3.4 million people, including 1.2 percent of the African Americans who lived largely in the inner city. The procedure for taking a census consists of sending out mailings and following up with door-to-door canvassing of non-respondents.

Some statisticians have proposed augmenting the headcount by surveying a separate representative sample and using these data to estimate the size and demographics of non-respondents. Supporters of the idea argue that this would provide a better picture of the population; opponents say that the assumptions involved, along with processing errors, would produce more error.

It can be argued that a sizable random sample of the entire population accompanied by more extensive follow-up would provide more accurate data than the current procedure at no greater expense, but this is precluded by the Constitution. (For more on this topic, search the Internet for national census sampling.)

method is much more feasible than simple random sampling and more representative than cluster sampling. It may well be the best choice in this example, but it still requires permission from 25 schools and the resources to collect data from each.

Convenience sampling: Select all eighth-graders in four schools to which the researcher has access (again, we estimate two classes of 30 students per school, so $n = 30 \times 4 \times 2 = 240$). This method precludes generalizing beyond these four schools, unless a strong argument with supporting data can be made for their similarity to the entire group of accessible schools.

Purposive sampling: Select 8 classes from throughout the seven counties on the basis of demographic data showing that they are representative of all eighth-graders. Particular attention must be paid to self-esteem and achievement scores. The problem is that such data are unlikely to be available and, in any case, cannot eliminate possible differences between the sample and the population on other variables—such as teacher attitude and available resources.

Systematic sampling: Select every 45th student from an alphabetical list for each school.

$$\frac{\text{200 students in sample}}{\text{9,000 students in population}} = \frac{1}{45}$$

This method is almost as inconvenient as simple random sampling and is likely to result in a biased sample, since the 45th name in each school is apt to be in the last third of the alphabet (remember there are an estimated 60 eighth-graders in each school), introducing probable ethnic or cultural bias.

Sample Size

Drawing conclusions about a population after studying a sample is never totally satisfactory, since researchers can never be sure that their sample is perfectly representative of the population. Some differences between the sample and the population are bound to exist, but if the sample is randomly selected and of sufficient size, these differences are likely to be relatively insignificant and incidental. The question remains, therefore, as to what constitutes an adequate, or sufficient, size for a sample.

Unfortunately, there is no clear-cut answer to this question. Suppose a target population consists of 1,000 eighth-graders in a given school district. Some sample sizes, of course, are obviously too small. Samples with 1 or 2 or 3 individuals, for example, are so small that they cannot possibly be representative. Probably any sample that has less than 20 to 30 individuals is too small, since that would only be 2 or 3 percent of the population. On the other hand, a sample can be too large, given the amount of time and effort the researcher must put into obtaining it. In this example, a sample of 250 or more individuals would probably be needlessly large, as that would constitute a quarter of the population. But what about samples of 50 or 100? Would these be sufficiently large? Would a sample of 200 be too large? At what point, exactly, does a sample stop being too small and become sufficiently large? The best answer is that a sample should be as large as the researcher can obtain with a reasonable expenditure of time and energy. This, of course, is not as much help as one would like, but it suggests that researchers should try to obtain as large a sample as they reasonably can.

MORE ABOUT RESEARCH

The Difficulty in Generalizing from a Sample

In 1936 the *Literary Digest,* a popular magazine of the time, selected a sample of voters in the United States and asked them for whom they would vote in the upcoming presidential election—Alf Landon (Republican) or Franklin Roosevelt (Democrat). The magazine editors obtained a sample of 2,375,000 individuals from lists of automobile and telephone owners in the United States (about 20 percent returned the mailed postcards). On the basis of their findings, the editors predicted that Landon would win by a landslide. In fact, it was Roosevelt who won the landslide victory. What was wrong with the study?

Certainly not the size of the sample. The most frequent explanations have been that the data were collected too far ahead of the election and that *a lot* of people changed their minds, and/or that the sample of voters was heavily biased in favor of the more affluent, and/or that the 20 percent return rate introduced a major bias. What do you think?

A common misconception among beginning researchers is illustrated by the following statement: "Although I obtained a random sample only from schools in San Francisco, I am entitled to generalize my findings to the entire state of California because the San Francisco schools (and hence my sample) reflect a wide variety of socioeconomic levels, ethnic groups, and teaching styles." The statement is incorrect because variety is not the same thing as representativeness. In order for the San Francisco schools to be representative of all the schools in California, they must be very similar (ideally, identical) with respect to characteristics such as the ones mentioned. Ask yourself: Are the San Francisco schools representative of the entire state with regard to ethnic composition of students? The answer, of course, is that they are not.

There are a few guidelines that we would suggest with regard to the *minimum* number of subjects needed. For descriptive studies, we think a sample with a minimum number of 100 is essential. For correlational studies, a sample of at least 50 is deemed necessary to establish the existence of a relationship. For experimental and causal-comparative studies, we recommend a minimum of 30 individuals per group, although sometimes experimental studies with only 15 individuals in each group can be defended if they are very tightly controlled; studies using only 15 subjects per group should probably be replicated, however, before too much is made of any findings.* In qualitative studies, the number of participants in a sample is usually somewhere between 1 and 20.

External Validity: Generalizing from a Sample

As indicated earlier in this chapter, researchers generalize when they apply the findings of a particular study to people or settings that go beyond the particular people or settings used in the study. The whole notion of science is built on the idea of **generalizing**. Every science seeks to find basic principles or laws that can be applied to a great variety of situations and, in the case of the social sciences, to a great many people. Most researchers wish to generalize their findings to appropriate populations. But when is generalizing warranted? When can researchers say with confidence that what they have learned about a sample is also true of the population? Both the nature of the sample and the environmental conditions—the setting—within which a study takes place must be considered in thinking about generalizability. The extent to which the results of a study can be generalized determines the **external validity** of the study. In the next two chapters, we also discuss how the concept of validity is applied to instruments (instrument validity) and to the internal design of a study.

POPULATION GENERALIZABILITY

Population generalizability refers to the degree to which a sample represents the population of interest. If the results of a study only apply to the group being studied and if that group is fairly small or is narrowly defined, the usefulness of any findings is seriously limited. This is why trying to obtain a representative sample is so important. Because conducting a study takes a considerable amount of time, energy, and (frequently) money, researchers usually want the results of an investigation to be as widely applicable as possible.

*More specific guidelines are provided in the Research Tips box on page 234 in Chapter 11.

When we speak of **representativeness**, however, we are referring only to the essential, or relevant, characteristics of a population. What do we mean by *relevant?* Only that the characteristics referred to might possibly be a contributing factor to any results that are obtained. For example, if a researcher wished to select a sample of first- and second-graders to study the effect of reading method on pupil achievement, such characteristics as height, eye color, or jumping ability would be judged to be irrelevant—that is, we would not expect any variation in them to have an effect on how easily a child learns to read, and hence we would not be overly concerned if those characteristics were not adequately represented in the sample. Other characteristics, such as age, gender, or visual acuity, on the other hand, might (logically) have an effect and hence should be appropriately represented in the sample.

Whenever purposive or convenience samples are used, generalization is made more plausible if data are presented to show that the sample is representative of the intended population on at least some relevant variables. This procedure, however, can never guarantee representativeness on all relevant variables.

One aspect of generalizability that is often overlooked in "methods" or "treatment" studies pertains to the teachers, counselors, administrators, or others who administer the various treatments. We must remember that such studies involve not only a sample of students, clients, or other recipients of the treatments but also a sample of those who implement the various treatments. Thus, a study that randomly selects students but not teachers is only entitled to generalize the outcomes to the population of students—*if* they are taught by the same teachers. To generalize the results to other teachers, the sample of teachers must also be selected randomly and must be sufficiently large.

Finally, we must remember that the sample in any study is the group about whom data are actually obtained. The best sampling plan is of no value if information is missing on a sizable portion of the initial sample. Once the sample has been selected, every effort must be made to ensure that the necessary data are obtained on each person in the sample. This is often difficult to do, particularly with questionnaire-type survey studies, but the results are well worth the time and energy expended. Unfortunately, there are no clear guidelines as to how many subjects can be lost before representativeness is seriously impaired. Any researchers who lose over 10 percent of the originally selected sample would be well advised to acknowledge this limitation and qualify their conclusions accordingly.

Do researchers always want to generalize? The only time researchers are not interested in generalizing beyond the confines of a particular study is when the results of an investigation are of interest only as applied to a particular group of people at a particular time, and where all of the members of the group are included in the study. An example might be the opinions of an elementary school faculty on a specific issue such as whether to implement a new math program. This might be of value to that faculty for decision making or program planning, but not to anyone else.

WHEN RANDOM SAMPLING IS NOT FEASIBLE

As we have shown, sometimes it is not feasible or even possible to obtain a random sample. When this is the case, researchers should describe the sample as thoroughly as possible (detailing, for example, age, gender, ethnicity, and socioeconomic status) so that interested others can judge for themselves the degree to which any findings apply, and to whom and where. This is clearly an inferior procedure compared to random sampling, but sometimes it is the only alternative one has.

There is another possibility when a random sample is impossible to obtain: It is called **replication**. The researcher (or other researchers) repeats the study using different groups of subjects in different situations. If a study is repeated several times, using different groups of subjects and under different conditions of geography, socioeconomic level, ability, and so on, and if the results obtained are essentially the same in each case, a researcher may have additional confidence about generalizing the findings.

In the vast majority of studies that have been done in education, random samples have not been used. There seem to be two reasons for this. First, educational researchers may be unaware of the hazards involved in generalizing when one does not have a random sample. Second, in many studies it is simply not feasible for a researcher to invest the time, money, or other resources necessary to obtain a random sample. For the results of a particular study to be applicable to a larger group, then, the researcher must argue convincingly that the sample employed, even though not chosen randomly, is in fact representative of the target population. This is difficult, however, and always subject to contrary arguments.

ECOLOGICAL GENERALIZABILITY

Ecological generalizability refers to the degree to which the results of a study can be extended to other settings or conditions. Researchers must make clear the nature of the environmental conditions—the setting—under which a study takes place. These conditions must be the same in all important respects in any new situation in which researchers wish to assert that their findings apply. For example, it is not justifiable to generalize from studies on the effects of a new reading program on third-graders in a large urban school system to teaching mathematics, even to those students in that system. Research results from urban school environments may not apply to suburban or rural school environments; results obtained with transparencies may not apply to those with textbooks. What holds true for one subject, or with certain materials, or under certain conditions, or at certain times may not generalize to other subjects, materials, conditions, or times.

An example of inappropriate ecological generalizing occurred in a study that found that a particular method of instruction applied to map reading resulted in greater transfer to general map interpretation on the part of fifth-graders in several schools. The researcher accordingly recommended that the method of instruction be used in other content areas, such as mathematics and science, overlooking differences in content, materials, and skills involved, in addition to probable differences in resources, teacher experience, and the like. Improper ecological generalizing such as this remains the bane of much educational research.

Unfortunately, application of the powerful technique of random sampling is virtually never possible with respect to ecological generalizing. While it is conceivable that a researcher could identify "populations" of organization patterns, materials, classroom conditions, and so on, and then randomly select a sizable number of combinations from all possible combinations, the logistics of doing so boggle the mind. Therefore, researchers must be cautious about generalizing the results from any one study. Only when outcomes have been shown to be similar through replication across specific environmental conditions can we generalize across those conditions. Figure 6.7 illustrates the difference between population and ecological generalizing.

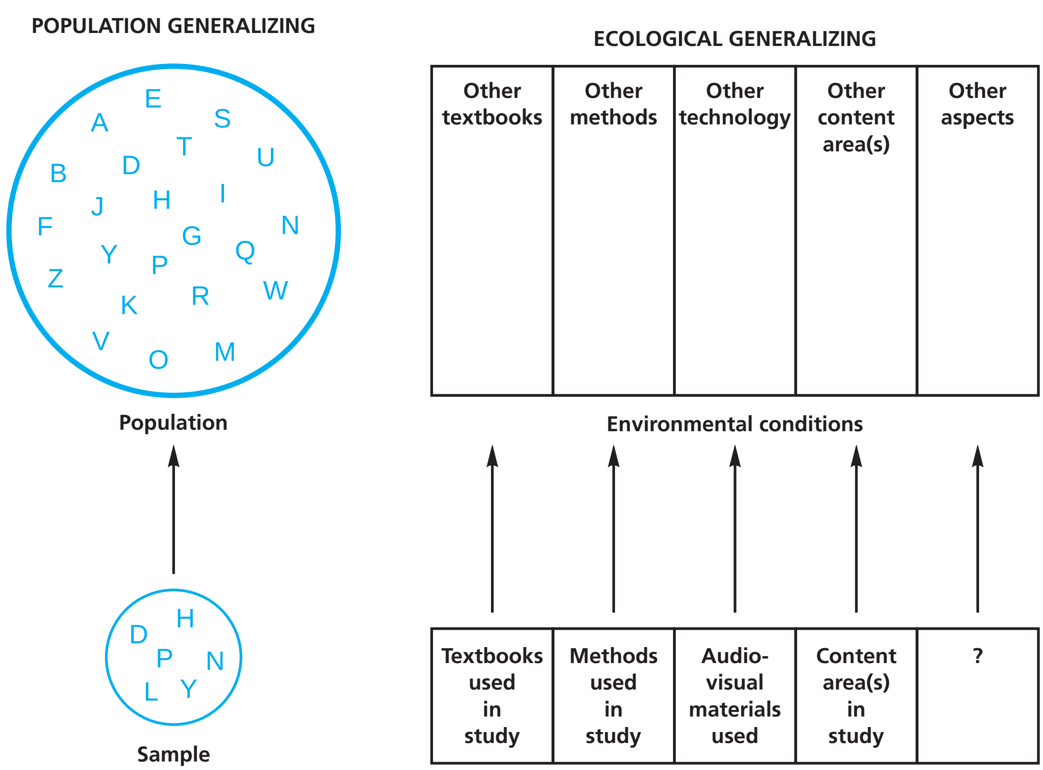


Figure 6.7 *Population as Opposed to Ecological Generalizing*

Go back to the **INTERACTIVE AND APPLIED LEARNING** feature at the beginning of the chapter for a listing of interactive and applied activities. Go to the **Online Learning Center** at **www.mhhe.com/fraenkel8e** to take quizzes, practice with key terms, and review chapter content.

Main Points

SAMPLES AND SAMPLING

- The term *sampling,* as used in research, refers to the process of selecting the individuals who will participate (e.g., be observed or questioned) in a research study.
- A sample is any part of a population of individuals on whom information is obtained. It may, for a variety of reasons, be different from the sample originally selected.

SAMPLES AND POPULATIONS

- The term *population,* as used in research, refers to all the members of a particular group. It is the group of interest to the researcher, the group to whom the researcher would like to generalize the results of a study.
- A target population is the actual population to whom the researcher would like to generalize; the accessible population is the population to whom the researcher is entitled to generalize.
- A representative sample is a sample that is similar to the population on all characteristics.

RANDOM VERSUS NONRANDOM SAMPLING

- Sampling may be either random or nonrandom. Random sampling methods include simple random sampling, stratified random sampling, cluster random sampling, and two-stage random sampling. Nonrandom sampling methods include systematic sampling, convenience sampling, and purposive sampling.

RANDOM SAMPLING METHODS

- A simple random sample is a sample selected from a population in such a manner that all members of the population have an equal chance of being selected.
- A stratified random sample is a sample selected so that certain characteristics are represented in the sample in the same proportion as they occur in the population.
- A cluster random sample is one obtained by using groups as the sampling unit rather than individuals.
- A two-stage random sample selects groups randomly and then chooses individuals randomly from these groups.
- A table of random numbers lists and arranges numbers in no particular order and can be used to select a random sample.

NONRANDOM SAMPLING METHODS

- A systematic sample is obtained by selecting every *n*th name in a population.
- A convenience sample is any group of individuals that is conveniently available to be studied.
- A purposive sample consists of individuals who have special qualifications of some sort or are deemed representative on the basis of prior evidence.

SAMPLE SIZE

- Samples should be as large as a researcher can obtain with a reasonable expenditure of time and energy. A recommended minimum number of subjects is 100 for a descriptive study, 50 for a correlational study, and 30 in each group for experimental and causal-comparative studies.

EXTERNAL VALIDITY (GENERALIZABILITY)

- The term *external validity,* as used in research, refers to the extent that the results of a study can be generalized from a sample to a population.
- The term *population generalizability* refers to the extent to which the results of a study can be generalized to the intended population.
- The term *ecological generalizability* refers to the extent to which the results of a study can be generalized to conditions or settings other than those that prevailed in a particular study.

REPLICATION

- When a study is replicated, it is repeated with a new sample and sometimes under new conditions.

Key Terms

For Discussion

1. A team of researchers wants to determine student attitudes about the recreational services available in the student union on campus. The team stops the first 100 students it meets on a street in the middle of the campus and asks each of them questions about the union. What are some possible ways that this sample might be biased?

2. Suppose a researcher is interested in studying the effects of music on learning. He obtains permission from a nearby elementary school principal to use the two third-grade classes in the school. The ability level of the two classes, as shown by standardized tests, grade point averages, and faculty opinion, is quite similar. In one class, the researcher plays classical music softly every day for a semester. In the other class, no music is played. At the end of the semester, he finds that the class in which the music was played has a markedly higher average in arithmetic than the other class, although they do not differ in any other respect. To what population (if any) might the results of this study be generalized? What, exactly, could the researcher say about the effects of music on learning?
3. When, if ever, might a researcher not be interested in generalizing the results of a study? Explain.
4. "The larger a sample, the more justified a researcher is in generalizing from it to a population." Is this statement true? Why or why not?
5. Some people have argued that no population can *ever* be studied in its entirety. Would you agree? Why or why not?
6. "The more narrowly researchers define the population, the more they limit generalizability." Is this always true? Discuss.
7. "The best sampling plan is of no value if information is missing on a sizable proportion of the initial sample." Why is this so? Discuss.
8. "The use of random sampling is almost never possible with respect to ecological generalizing." Why is this so? Can you think of a possible study for which ecological generalizing would be possible? If so, give an example.

Research Exercise 6: Sampling Plan

Use Problem Sheet 6 to describe, as fully as you can, your sample—that is, the subjects you will include in your study. Describe the type of sample you plan to use and how you will obtain the sample. Indicate whether you expect your study to have population generalizability. If so, to what population? if not, why not? Then indicate whether the study would have ecological generalizability. If so, to what settings? if not, why would it not?

Problem Sheet 6

Sampling Plan

1. My intended sample (participants in your study) consists of (*state who and how many*):
____

2. Key demographics (characteristics of the sample) are as follows (e.g., age range, sex distribution, ethnic breakdown, socioeconomic status, location [where are these subjects located?], etc.): _______________
____

3. State what type of sample you plan to use (i.e., convenience, purposive, simple random, stratified random, cluster, systematic). _______________
____

4. I will gain access to and/or get contact information for my sample through the following steps: _______________
____

5. What, if any, are the inclusion/exclusion criteria for participation in your study?
____

6. External validity:
   a. To whom do you think you can generalize the results of your study? Explain.
   ____
   b. If applicable, to what settings/conditions could you generalize the results of your study (ecological validity)?
   ____
   c. If results are not generalizable, why not?
   ____

An electronic version of this Problem Sheet that you can fill in and print, save, or e-mail is available on the Online Learning Center at www.mhhe.com/fraenkel8e.