CHAPTER 5

Sampling

Research Question: *How Can We Take a Valid Sample in an Educational Setting?*

Chapter Contents

- **Sample Planning**
- **Sampling Methods**
- **Sampling Distributions**

A common technique in journalism is to put a "human face" on a story. For instance, a *Boston Globe* correspondent (MacDonald, 2005) interviewed the mother of a 3-year-old daughter for a story about pre-kindergarten students being expelled from their schools for behavioral problems. Lisa Mathey of Ashburn, Virginia, was frustrated because her child, who has attention deficit and oppositional defiant disorders, was asked to leave one pre-K program, and this raised liability concerns at the next program she went to. Eventually, Lisa was able to get her daughter into a federally mandated program for children with disabilities. The reporter included a second similar story and also interviewed several educators and "experts."

The story provides a compelling rationale for not giving up on these particular "at-risk" children, for providing more consultant services to preschools, and for legislation to address this issue. However, we do not know if the children mentioned in the story are like most children in nursery schools or like most children who are expelled from their nursery schools—or if they are just two compelling stories that caught the eye of one reporter. In other words, we don't know how generalizable their stories are, and if we don't have confidence in generalizability, then the validity of this account of expulsion of preschool students is suspect. Because we don't know whether their situation is widely shared or unique, we cannot really judge what the account tells us about the educational world.

Sampling techniques tell us how to select cases that can lead to valid generalizations about a **population,** or the entire group you wish to learn about. In this chapter, we define the key components of sampling strategy and then present the types of sampling one may use in a research study, with the strengths and weaknesses of each.

By the chapter's end, you should understand which questions you need to ask to evaluate the generalizability of a study as well as what choices you need to make when designing a sampling strategy. You should also realize that it is just as important to select the "right" people or objects to study as it is to ask participants the right questions.

2Sample Planning

You have encountered the problem of generalizability in many of the studies you have read about in this book. Whether we are designing a sampling strategy or evaluating someone else's findings, we have to understand how and why researchers decide to sample and what the consequences of these decisions are for the generalizability of the study's findings.

Define Sample Components and the Population

Let's say that we are designing a survey about families of kindergarten children in a large city. We don't have the time or resources to study the entire kindergarten family population of the city, even though it comprises the set of individuals or other entities to which we wish to be able to generalize our findings. Even the school department, which closely tracks kindergarten enrollment because it predicts upper grade enrollment for subsequent years, does not have the resources to actually survey the families of the children they count. So instead, we resolve to study a **sample,** a subset of this population. The individual members of this sample are called **elements,** or elementary units.

In many studies, we sample directly from the elements in the population of interest. We may survey a sample of the entire population of students at a school, based on a list obtained from the principal's office. This list, from which the elements of the population are selected, is termed the **sampling frame.** The students who are selected and interviewed from that list are the elements.

Population: The entire set of individuals or other entities to which study findings are to be generalized.

Elements: The individual members of the population whose characteristics are to be measured.

Sampling frame: A list of all elements or other units containing the elements in a population.

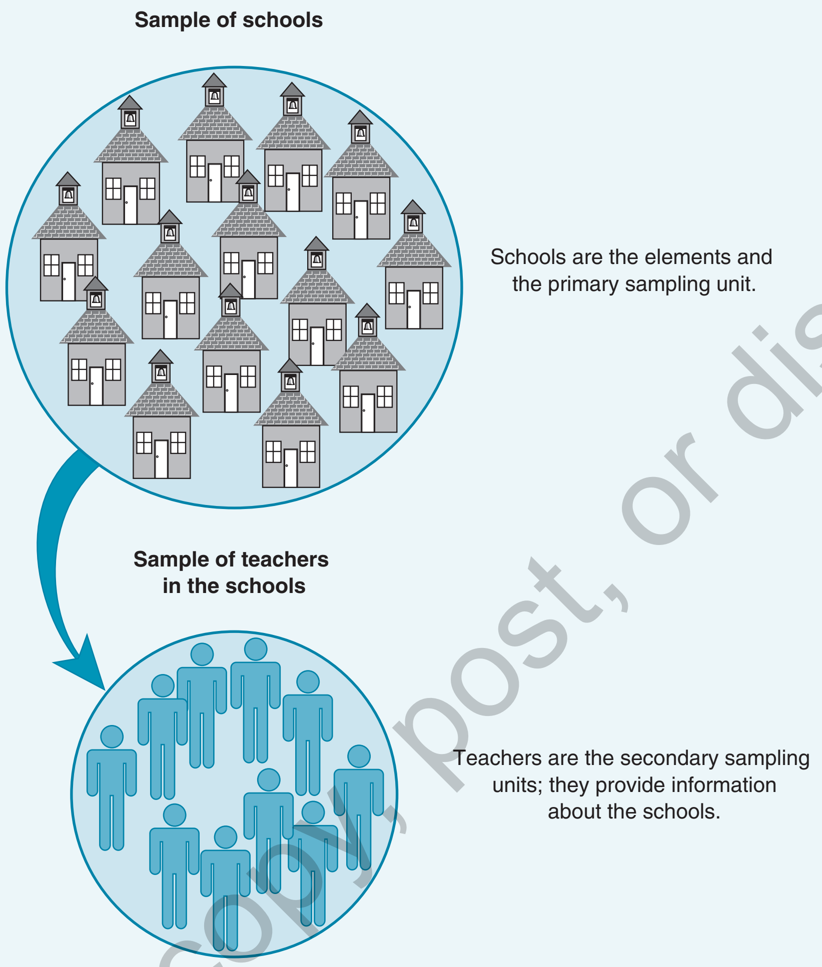
Enumeration units: Units that contain one or more elements and that are listed in a sampling frame.

Sampling units: Units listed at any stage of a multistage sampling design.

In some studies, the entities that can easily be reached are not the same as the elements from which we want information, but they include those elements. For example, we may have a list of households but not a list of the kindergarten-age children of a town, even though the children are the elements that we actually want to sample. In this situation, we could draw a sample of households so that we can then identify the kindergarten-age children in these households. The households are termed **enumeration units,** and the children in the households are the elements (Levy & Lemeshow, 1999, pp. 13–14).

Sometimes, the individuals or other entities from which we collect information are not actually the elements in our study. For example, a researcher might sample schools for a survey about educational practices and then interview a sample of teachers in each sampled school to obtain further information. Both the schools and the teachers are termed **sampling units** because we sample from both (Levy & Lemeshow, 1999, p. 22). The schools are selected in the first stage of the sample, so they are the *primary sampling units* (in this case, they are also the elements in the study). The teachers are *secondary sampling units* (but they are not elements, because they are used to provide information about the entire school) (see Exhibit 5.1).

Exhibit 5.1 Sample Components in a Two-Stage Study


Sample of schools
Schools are the elements and the primary sampling unit.
Sample of teachers in the schools
Teachers are the secondary sampling units; they provide information about the schools.

Source: Schutt (2009, p. 151). Based on information from Levy and Lemeshow (1999).

It is important to know exactly what population a sample can represent when you select or evaluate sample components. The population for a study is the aggregation of elements that we actually focus on and sample from, not some larger aggregation we really wish we could have studied. If we sample students in one high school, the population for our study is the student body of that school, not all high school students in the city or the state in which that high school is located.

Some populations, such as at-risk students, are not identified by a simple criterion such as a geographic boundary or an organizational membership. Clear definition of such a population is difficult (as we saw in the previous chapter) but quite necessary. Anyone should be able to determine just what population was actually studied, so we would have to define clearly the concept of "at-risk students" and specify how we determined their status.

A clear definition allows researchers in other locations or at other times to develop procedures for studying a comparable population. The more complete and explicit the definition is of the population from which a sample was selected, the more precise our generalizations can be.

Evaluate Generalizability

Once we have defined clearly the population from which we will sample, we need to determine the scope of the generalizations we will make from our sample.

Can the findings from a sample of the population be generalized to the population from which the sample was selected? Do a certain study's findings and theory apply to all the students in a given school? To the entire school population of a given state? Of all the students in the United States? This type of generalizability is defined as *sample generalizability.*

Can the findings from a study of one population be generalized to another, somewhat different population? Are the students of a small-town school system similar to students in other small towns? To students in large urban areas as well? To students across the United States? Do findings from a study of reading development in a large northeastern city differ from those that would be obtained in a rural area in the Southwest? This type of generalizability question is defined as cross-population generalizability.

This chapter focuses attention on the problem of sample generalizability: Can findings from a sample be generalized to the population from which the sample was drawn? This is really the most basic question to ask about a sample, and educational research methods provide many tools with which to address it.

Sample generalizability depends on sample quality, which is determined by the amount of **sampling error**—the difference between the characteristics of a sample and the characteristics of the population from which it was selected. The larger the sampling error, the less representative the sample—and thus the less generalizable the findings. To assess sample quality when you are planning or evaluating a study, ask yourself these questions:

- From what population were the cases selected?
- What method was used to select cases from this population?
- Do the cases that were studied represent, in the aggregate, the population from which they were selected?

But researchers often project their theories onto groups or populations much larger than, or simply different from, those they have actually studied. The population to which generalizations are made in this way can be termed the **target population**—a set of elements larger than or different from the population that was sampled and to which the researcher would like to generalize any study findings. When we generalize findings to target populations, we must carefully consider the validity of claims that the findings can be applied to other groups, geographic areas, cultures, or times. Because the validity of cross-population generalizations cannot be tested empirically, except by conducting more research in other settings, we will not focus much attention on this problem here.

Assess the Diversity of the Population

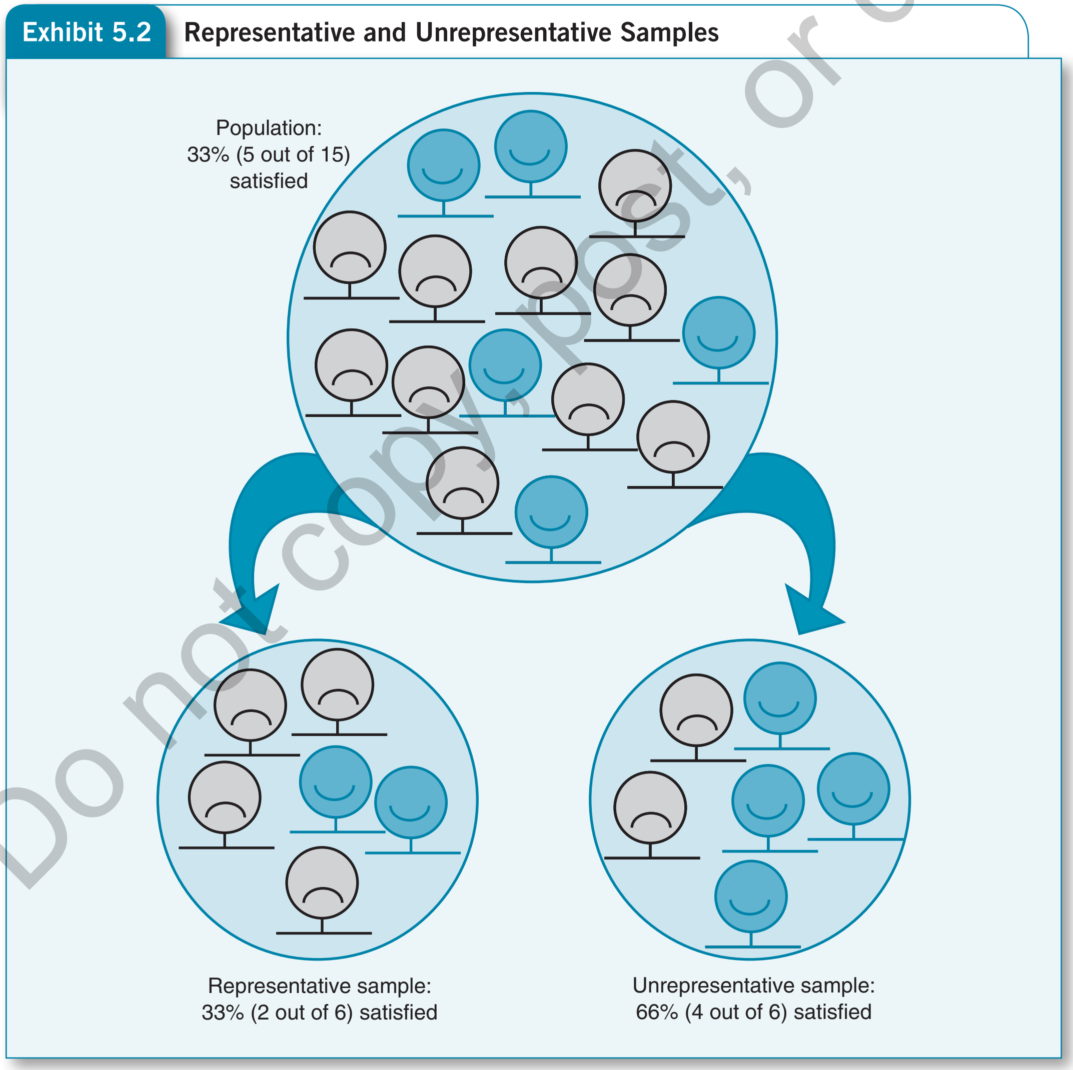
Sampling is unnecessary if all the units in the population are identical. Physicists don't need to select a representative sample of atomic particles to learn about basic physical processes. They can study a single atomic particle because it is identical to every other particle of its type.

What about people? Certainly all people are not identical (nor are other animals, in many respects). Nonetheless, if we are studying physical or psychological processes that are the same among all people, sampling is not needed to achieve generalizable findings. Psychologists and social psychologists often conduct experiments on college students to learn about processes that they think are identical across individuals. They believe that most people would have the same reactions as the college students if they experienced the same experimental conditions. But we must always bear in mind that we don't really know how generalizable our findings are to populations we haven't actually studied. This is particularly important when studying diverse school populations, where variables such as cultural background, language, income level, and family expectations can have strong effects on students' behavior and academic performance.

So we usually conclude that we must study the larger population in which we are interested, if we want to be able to make generalizations about it. For this purpose, we must obtain a **representative sample** of the population to which generalizations are sought (see Exhibit 5.2). The educational world and the people in it are just too diverse to be considered "identical units."

Representative sample: A sample that "looks like" the population from which it was selected in all respects that are potentially relevant to the study. The distribution of characteristics among the elements of a representative sample is the same as the distribution of those characteristics among the total population. In an unrepresentative sample, some characteristics are overrepresented or underrepresented.

Exhibit 5.2 **Representative and Unrepresentative Samples**



Source: Schutt (2009, p. 155).

Consider a Census

Census: Research in which information is obtained through the responses that all available members of an entire population give to questions.

In some circumstances, it may be feasible to skirt the issue of generalizability by conducting a **census**—studying the entire population of interest—rather than drawing a sample. This is what the federal government tries to do every 10 years with the U.S. census. Censuses also include studies of all the employees in small organizations, all the students in a class, or all the teachers in a school. Researchers don't often attempt to collect data from all the members of some large population because doing so would be too expensive and time-consuming. But fortunately, a well-designed sampling strategy can result in a representative sample of the same population at far less cost and in far less time.

In most survey situations, it is much better to survey only a limited number from the total population so that there are more resources for follow-up procedures that can overcome reluctance or indifference about participation.

2 Sampling Methods

Certain features of samples make them more or less likely to represent the population from which they are selected; the more representative the sample, the better. The crucial distinction about samples is whether they are based on a probability or a nonprobability sampling method. **Probability sampling methods** allow us to know in advance how likely it is that any element of a population will be selected. Sampling methods that do not let us know in advance the likelihood of selecting each element are termed **nonprobability sampling methods.**

Probability of selection: The likelihood that an element will be selected from the population for inclusion in the sample. In a census of all the elements of a population, the probability that any particular element will be selected is 1.0. If half of the elements in the population are sampled on the basis of chance (say, by tossing a coin), the probability of selection for each element is one half, or .5. As the size of the sample as a proportion of the population decreases, so does the probability of selection.

Probability sampling methods rely on a random, or chance, selection procedure, which is, in principle, the same as flipping a coin to decide which of two people "wins" and which one "loses." Heads and tails are equally likely to turn up in a coin toss, so both persons have an equal chance to win. Their **probability of selection** is 1 out of 2, or .5.

Flipping a coin is a fair way to select one of two people because the selection process harbors no **systematic bias.** You might win or lose the coin toss, but you know that the outcome was due simply to chance, not to bias. For the same reason, a roll of a six-sided die is a fair way to choose one of six possible outcomes (the odds of selection are 1 out of 6, or .17).

There is a natural tendency to confuse the concept of **random sampling,** in which cases are selected only on the basis of chance, with a haphazard method of sampling. On first impression, "leaving things up to chance" seems to imply not exerting any control over the sampling method. But to ensure that nothing but chance influences the selection of cases, the researcher must proceed very methodically, leaving nothing to chance except the selection of the cases themselves. The researcher must follow carefully controlled procedures if a purely random process is to occur.

Two problems are often cause for concern when drawing random samples:

1. If the sampling frame is incomplete, a sample selected randomly from that list will not really be a random sample of the population. You should always consider the adequacy of the sampling frame. Even for a simple population such as a university's student body, the registrar's list is likely to be at least a bit out of date at any given time. For example, some students will have dropped out, but their status will not yet be officially recorded.

2. Nonresponse is a major hazard in survey research because **nonrespondents** are likely to differ systematically from those who take the time to participate. If the response rate is low (say, below 65%), you should not assume that findings from even a random sample will be generalizable to the population from which the sample was selected.

Probability Sampling Methods

Probability sampling methods are those in which the probability of selection is known and is not zero (so there is some chance of selecting each element). These methods randomly select elements and therefore have no systematic bias; nothing but chance determines which elements are included in the sample. This feature of probability samples makes them much more desirable than nonprobability samples when the goal is to generalize to a larger population.

Bias: Sampling bias occurs when some population characteristics are over- or underrepresented in the sample because of particular features of the method of selecting the sample.

Even a randomly selected sample will have some sampling error—some deviation from the characteristics of the population—due to chance. The probability of selecting a head is .5 in a single toss of a coin and in 20, 30, or however many tosses of a coin you like. But it is perfectly possible to toss a coin twice and get a head both times. The random "sample" of the two sides of the coin is selected in an unbiased fashion, but it still is unrepresentative. In general, both the size of the sample and the homogeneity (sameness) of the population affect the degree of error due to chance. Despite what you might think, the proportion of the population that the sample represents does not affect the sample representativeness, unless that proportion is very large—it is the number of cases in the sample that is important. To elaborate:

- The larger the sample, the more confidence we can have in the sample's representativeness. If we randomly pick 5 people to represent the entire population of our city, our sample is unlikely to be very representative of the entire population in terms of age, gender, race, attitudes, and so on. But if we randomly pick 100 people, the odds of having a representative sample are much better; with a random sample of 1,000, the odds become very good indeed.
- The more homogeneous the population, the more confidence we can have in the representativeness of a sample of any particular size. That's why blood testing works—blood is homogeneous in any specific individual's body. Or let's say we plan to draw samples of 50 people from each of two communities to estimate mean family income. One community is very diverse, with family incomes varying from $12,000 to $85,000. In the other, more homogeneous community, family incomes are concentrated in a narrow range, from $41,000 to $64,000. The estimated mean family income based on the sample from the homogeneous community is more likely to be representative than is the estimate based on the sample from the more heterogeneous community. With less variation to represent, fewer cases are needed to represent the homogeneous community.
- The fraction of the total population that a sample contains does not affect the sample's representativeness unless that fraction is large. This isn't obvious, but it is mathematically true. The raw number of cases matters more than the proportion of the population. Other things being equal, a sample of 1,000 from a population of 1 million (with a sampling fraction of 0.001, or 0.1%) is much better than a sample of 100 from a population of 10,000 (although the sampling fraction is 0.01, or 1%, which is 10 times higher). The larger size of the samples makes representativeness more likely, not the proportion of the whole that the sample represents. We can regard any sampling fraction under 2% with about the same degree of confidence (Sudman, 1976, p. 184). In fact, sample representativeness is not likely to increase much until the sampling fraction is quite a bit higher.

Because they do not disproportionately exclude or include particular groups within the population, random samples that are successfully implemented avoid systematic bias. Random error can still be considerable, however, and different types of random samples vary in their ability to minimize it. The four most common types of random samples are simple random sampling, systematic random sampling, cluster sampling, and stratified random sampling.

Simple Random Sampling

Simple random sampling identifies cases strictly on the basis of chance. As you know, flipping a coin or rolling a die can be used to identify cases strictly on the basis of chance, but these procedures are not very efficient tools for drawing a sample. A **random number table** simplifies the process considerably. The researcher numbers all the elements in the sampling frame and then uses a systematic procedure for picking corresponding numbers from the random number table. (Practice Exercise 1 at the end of this chapter explains the process step-by-step.) Alternatively, a researcher may use a lottery procedure. Each case number is written on a small card, and then the cards are mixed up and the sample is selected from the cards. A computer program can also easily generate a random sample of any size.

The probability of selection in a true simple random sample is equal for each element. If a sample of 500 is selected from a population of 17,000 (i.e., a sampling frame of 17,000), then the probability of selection for each element is 500/17,000, or .03. Every element has an equal chance of being selected, just like the odds in a toss of a coin (1/2) or a roll of a die (1/6). Thus, simple random sampling is an "equal probability of selection method," or EPSEM.

Systematic Random Sampling

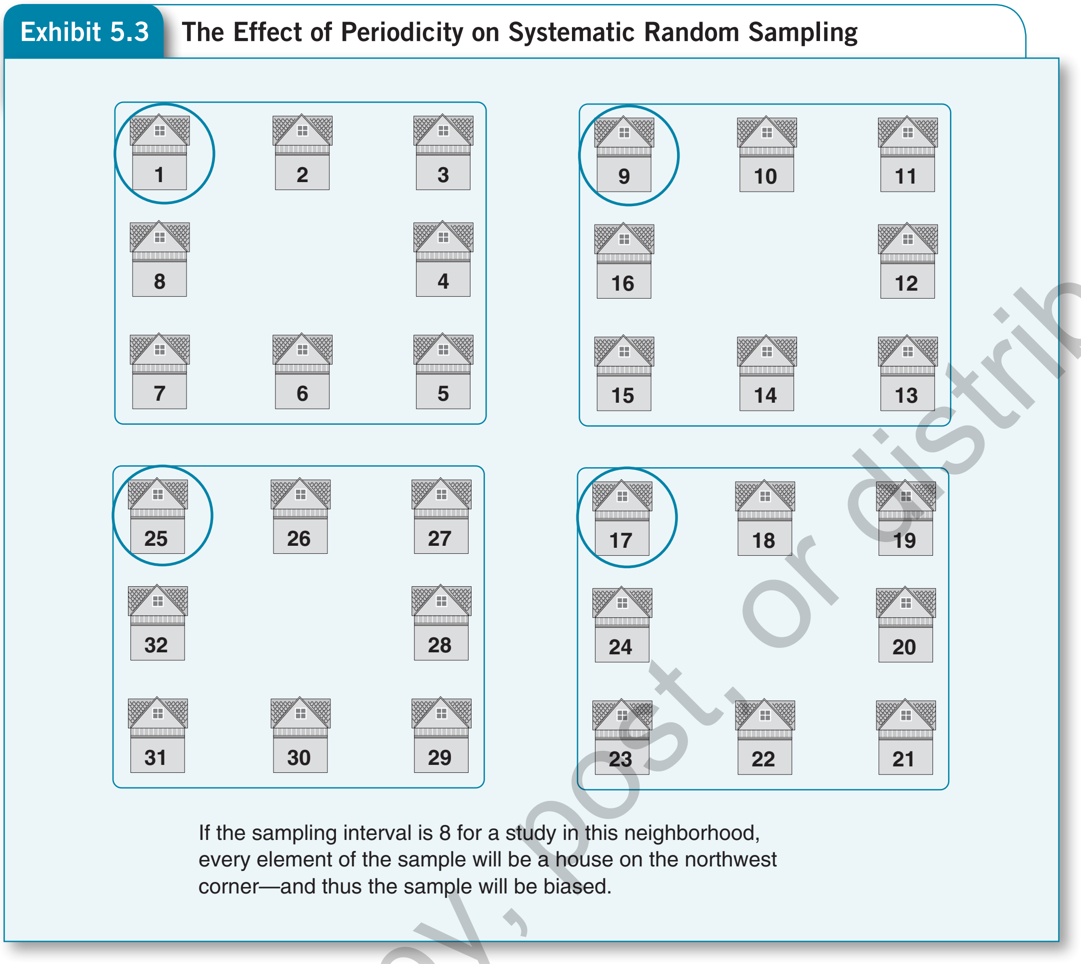
Systematic random sampling is a variant of simple random sampling. The first element is selected randomly from a list or from sequential files, and then every *n*th element is selected. This is a convenient method for drawing a random sample when the population elements are arranged sequentially. It is particularly efficient when the elements are not actually printed (i.e., there is no sampling frame) but instead are represented by folders in filing cabinets.

Systematic random sampling requires three steps:

1. The total number of cases in the population is divided by the number of cases required for the sample. This division yields the **sampling interval,** the number of cases from one sampled case to another. If 50 cases are to be selected out of 1,000, the sampling interval is 20; every 20th case is selected.
2. A number from 1 to 20 (or whatever the sampling interval is) is selected randomly. This number identifies the first case to be sampled, counting from the first case on the list or in the files.
3. After the first case is selected, every *n*th case is selected for the sample, where *n* is the sampling interval. If the sampling interval is not a whole number, the size of the sampling interval is varied systematically to yield the proper number of cases for the sample. For example, if the sampling interval is 30.5, the sampling interval alternates between 30 and 31.

In almost all sampling situations, systematic random sampling yields what is essentially a simple random sample. The exception is a situation in which the sequence of elements is affected by **periodicity**—that is, the sequence varies in some regular, periodic pattern. For example, the houses in a new development with the same number of houses on each block (e.g., eight) may be listed by block, starting with the house in the northwest corner of each block and continuing clockwise. If the sampling interval is 8, the same as the periodic pattern, all the cases selected will be in the same position (see Exhibit 5.3). But in reality, periodicity and the sampling interval are rarely the same, so this usually isn't a problem.

Exhibit 5.3 The Effect of Periodicity on Systematic Random Sampling



If the sampling interval is 8 for a study in this neighborhood, every element of the sample will be a house on the northwest corner—and thus the sample will be biased.

Source: Schutt (2009, p. 163).

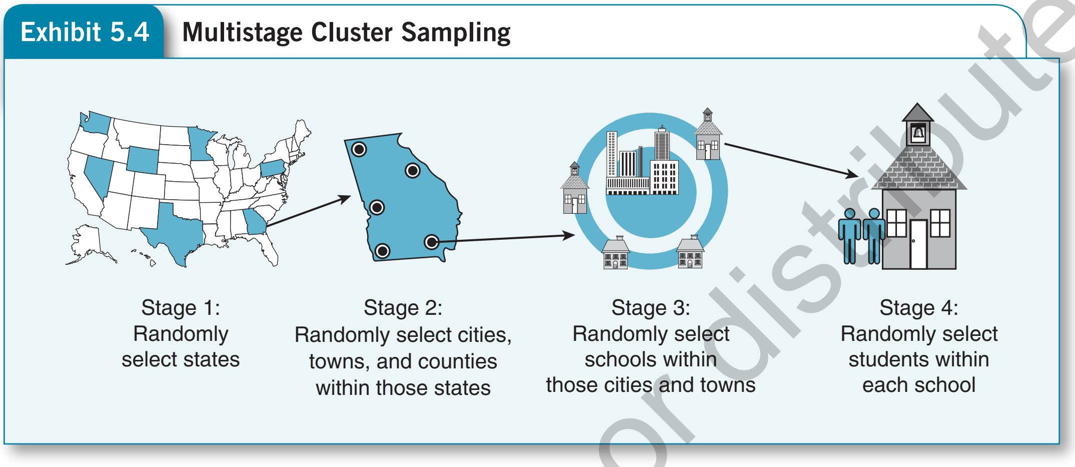
Cluster Sampling

Cluster sampling is useful when a sampling frame—a definite list—of elements is not available, as often is the case for large populations spread out across a wide geographic area or among many different organizations. We don't have a good list of all the Catholics in America, all the businesspeople in Arizona, or all the waiters in New York. A **cluster** is a naturally occurring, mixed aggregate of elements of the population, with each element (person, for instance) appearing in one, and only one, cluster. Schools could serve as clusters for sampling students, blocks could serve as clusters for sampling city residents, counties could serve as clusters for sampling the general population, and restaurants could serve as clusters for sampling waiters.

Cluster sampling is, at least, a two-stage procedure. First, the researcher draws a random sample of clusters. (A list of clusters should be much easier to obtain than a list of all the individuals in each cluster in the population.) Next, the researcher draws a random sample of elements within each selected cluster. Because only a fraction of the total clusters is involved, obtaining the sampling frame at this stage should be much easier.

Cluster samples often involve multiple stages, with clusters within clusters. A national study of middle school students, for example, might involve first sampling states, then geographic units (cities, towns, counties) within those states, then schools within those units, and, finally, students within each selected school (see Exhibit 5.4). In multistage cluster sampling, the clusters at the first stage of sampling are termed the *primary sampling units* (Levy & Lemeshow, 1999, p. 228).

Exhibit 5.4 **Multistage Cluster Sampling**



Source: Chambliss and Schutt (2010, p. 119).

Multistage Cluster Sampling

How many clusters and how many individuals within clusters should be selected? As a general rule, the more clusters you select, with the fewest individuals in each, the more representative your sampling will be. Unfortunately, this strategy also maximizes the time needed and cost for the sample. Remember, too, that the more internally homogeneous the clusters, the fewer cases needed per cluster. Homogeneity within a cluster is good. So if you set out to draw a cluster sample, be sure to consider how similar individuals are within the clusters, as well as, given the time and resources available, how many clusters you can afford to include.

Cluster sampling is a very popular method among survey researchers, but it has one general drawback: Sampling error is greater in a cluster sample than in a simple random sample because there are two steps involving random selection rather than just one. This sampling error increases as the number of clusters decreases, and it decreases as the homogeneity of cases per cluster increases. This is another way of restating the points above: It's better to include as many clusters as possible in a sample, and it's more likely that a cluster sample will be representative of the population if cases are relatively similar within clusters.

Stratified Random Sampling

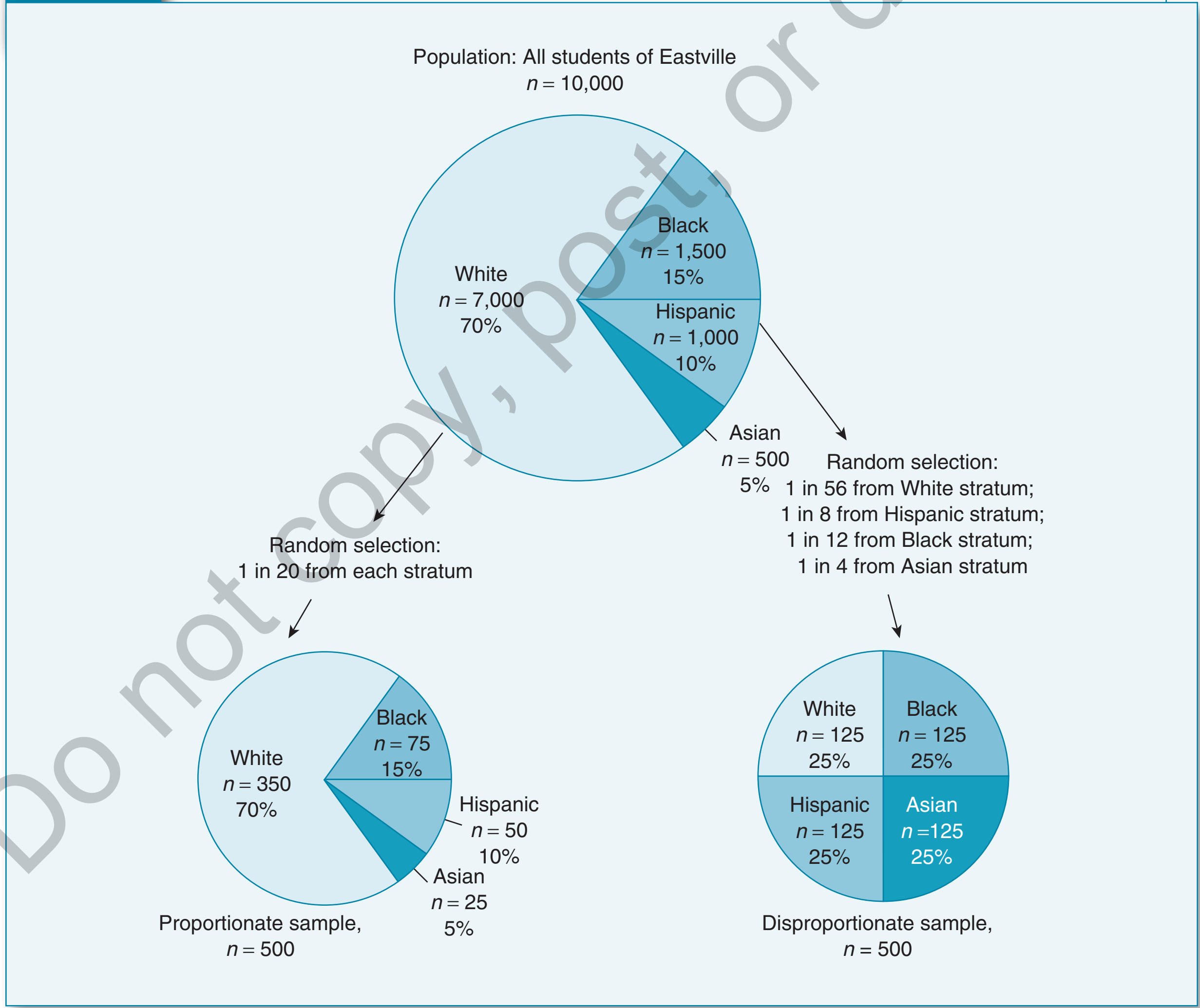
Suppose you want to survey personnel in the Eastville school system to determine their attitude toward a proposed change in the length of the school day. Simple random sampling would produce large numbers of teachers—the most numerous employees—but very few, if any, principals or higher administrators. But you want administrators in your sample. **Stratified random sampling** ensures that various groups will be included.

First, all elements in the population (i.e., in the sampling frame) are distinguished according to their value on some relevant characteristic (administrator, teacher, instructional aide, custodian, etc.). That characteristic determines the sampling strata. Next, elements are sampled randomly from within these strata: so many administrators, so many teachers, and so on. Of course, using this method requires more information prior to sampling than is the case with simple random sampling. Each element must belong to one and only one stratum.

For "proportionate to size" sampling, the size of each stratum in the population must be known. This method efficiently draws an appropriate representation of elements across strata. Imagine that you plan to draw a sample of 500 students from the Eastville school system, which is ethnically diverse and has 10,000 students. The student population is 15% Black, 10% Hispanic, 5% Asian, and 70% White. If you drew a simple random sample, you might end up with somewhat disproportionate numbers of each group. But if you created sampling strata based on race and ethnicity, you could randomly select cases from each stratum: 75 Blacks (15% of the sample), 50 Hispanics (10%), 25 Asians (5%), and 350 Whites (70%). By using **proportionate stratified sampling,** you would eliminate any possibility of sampling error in the sample's distribution of ethnicity. Each stratum would be represented exactly in proportion to its size in the population from which the sample was drawn (see Exhibit 5.5).

In **disproportionate stratified sampling,** the proportion of each stratum that is included in the sample is intentionally varied from what it is in the population. In the case of the sample stratified by ethnicity, you

Exhibit 5.5 Stratified Random Sampling



Source: Chambliss and Schutt (2010, p. 120).

might select equal numbers of cases from each racial or ethnic group: 125 Blacks (25% of the sample), 125 Hispanics (25%), 125 Asians (25%), and 125 Whites (25%). In this type of sample, the probability of selection of every case is known but unequal between strata. You know what the proportions are in the population, and so you can easily adjust your combined sample statistics to reflect these true proportions. For instance, if you want to combine the ethnic groups and estimate the average reading test score of the total population, you would have to "weight" each case in the sample. The weight is a number you multiply by the value of each case based on the stratum it is in. For example, you would multiply the scores of all Blacks in the sample by 0.6 (75/125), the scores of all Hispanics by 0.4 (50/125), and so on. Weighting in this way reduces the influence of the oversampled strata and increases the influence of the undersampled strata to what they would have been if pure probability sampling had been used.

Researchers sometimes include one element of *dis*proportionate random sampling in their otherwise proportionate random sampling strategy. Why would anyone select a sample that is so unrepresentative in the first place? The most common reason is to ensure that cases from smaller strata are included in the sample in sufficient numbers to allow separate statistical estimates and to facilitate comparisons between strata. Remember that one of the determinants of sample quality is sample size. The same is true for subgroups within samples. If a key concern in a research project is to describe and compare the reading scores of students from different racial and ethnic groups, then it is important that the researchers base the mean reading score of each group on enough cases to be a valid representation. If few members of a particular minority group are in the population, they need to be oversampled.

Nonprobability Sampling Methods

Nonprobability sampling methods are often used in qualitative research; they also are used in quantitative studies when researchers are unable to use probability selection methods. There are four common nonprobability sampling methods: availability sampling, quota sampling, purposive sampling, and snowball sampling. Because these methods do not use a random selection procedure, we cannot expect a sample selected with any of these methods to yield a representative sample. Nonetheless, these methods are useful when random sampling is not possible, when a research question calls for an intensive investigation of a small population, or when a researcher is performing a preliminary, exploratory study.

Availability Sampling

Elements are selected for **availability sampling** (sometimes called "haphazard" or "convenience" sampling) because they're available or easy to find. For example, sometimes people stand outside stores in shopping malls asking passersby to answer a few questions about their shopping habits. That may make sense, but asking the same people for their views on the economy doesn't. In important respects, regular mall shoppers are not representative members of the total population. The people who happen to be available in any situation are unlikely to be just like those who are unavailable. We can't be at all certain that what we learn can be generalized with any confidence to a larger population of concern.

An availability sample is often appropriate in educational research—for example, when a field researcher is exploring a new setting and trying to get some sense of prevailing attitudes or when a survey researcher conducts a preliminary test of a new set of questions. A participant observation study of a group may require no more sophisticated approach. A teacher-researcher studying the effects of homework assignments on sixth graders may decide to use the students in her own sixth-grade classroom because she sees them every day and assigns them homework. These students became the availability sample. However, they are in no way representative of all sixth graders in her school, her school system, or the sixth-grade population nationally.

Availability sampling often masquerades as a more rigorous form of research. Popular magazines periodically survey their readers by printing a questionnaire for readers to fill out and mail in. A follow-up article then appears in the magazine under a title such as "What You Think About Intimacy in Marriage?" If the magazine's circulation is large, a large sample can be achieved in this way. The problem is that usually only a tiny fraction of readers return the questionnaire, and these respondents are probably unlike other readers who did not have the interest or time to participate. So the survey is based on an availability sample. Even though the follow-up article may be interesting, we have no basis for thinking that the results describe the readership as a whole—much less the population at large.

Do you see now why availability sampling differs so much from random sampling methods, which require that "nothing but chance" affect the actual selection of cases? What makes availability sampling "haphazard" is precisely that a great many things other than chance can affect the selection of cases, ranging from the prejudices of the research staff to the work schedules of potential respondents. To truly leave the selection of cases up to chance, we have to design the selection process very carefully so that other factors are not influential. There's nothing "haphazard" about selecting cases randomly.

Quota Sampling

Quota sampling is intended to overcome the most obvious flaw of availability sampling—that the sample will just consist of whoever or whatever is available, without any concern for its similarity to the population of interest. In this approach, quotas are set to ensure that the sample represents certain characteristics in proportion to their prevalence in the population.

Suppose that you wish to sample adult residents of Eastville in a study of support for a tax increase to improve its schools. You know from the town's annual report what the proportions of town residents are in terms of gender, race, age, and number of children. You think that each of these characteristics might influence support for new school taxes, so you want to be sure that the sample includes men, women, Whites, Blacks, Hispanics, Asians, older people, younger people, big families, small families, and childless families in proportion to their numbers in the town population.

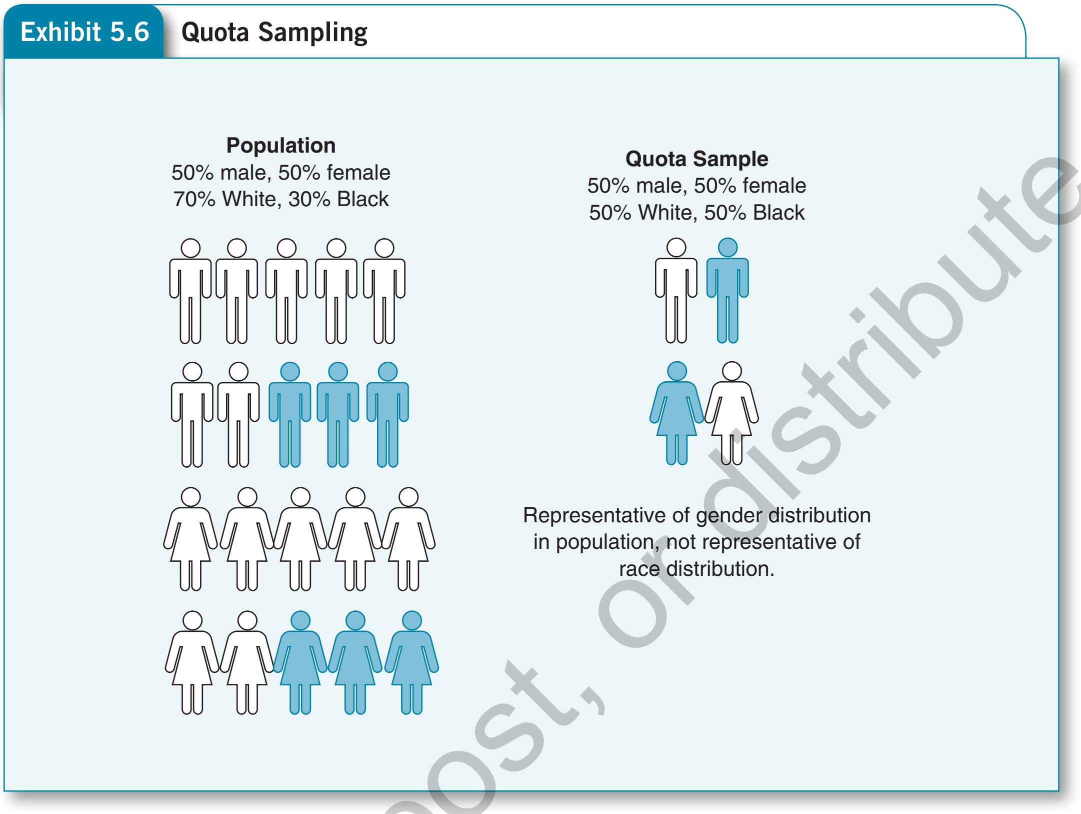
This is where quotas come in. Let's say that 48% of the town's adult residents are men and 52% are women and that 60% are employed, 5% are unemployed, and 35% are out of the labor force. These percentages and the percentages corresponding to the other characteristics become the quotas for the sample. If you plan to include a total of 500 residents in your sample, 240 must be men (48% of 500), 260 must be women, 300 must be employed, and so on. You may even set more refined quotas, such as certain numbers of employed women, employed men, unemployed men, and so on. With the quota list in hand, you (or your research staff) can now go out into the community looking for the right number of people in each quota category.

The problem is that even when we know that a quota sample is representative of the particular characteristics for which quotas have been set, we have no way of knowing if the sample is representative in terms of any other characteristics. In Exhibit 5.6, for example, quotas have been set for gender only. Under the circumstances, it's no surprise that the sample is representative of the population only in terms of gender, not in terms of race.

Of course, you must know the characteristics of the entire population to set the right quotas. In most cases, researchers know what the population looks like in terms of no more than a few of the characteristics relevant to their concerns—and in some cases, they have no such information on the entire population. Exhibit 5.7 summarizes the differences between quota sampling and stratified random sampling. The key difference, of course, is quota sampling's lack of random selection.

If you're now feeling skeptical of quota sampling, you've gotten the drift of our remarks. Nonetheless, in situations where you can't draw a random sample, it may be better to establish quotas than to have no parameters at all.

Exhibit 5.6 Quota Sampling



Source: Schutt (2009, p. 172).

Exhibit 5.7 Comparison of Stratified and Quota Sampling Methods

Feature	Stratified	Quota
Unbiased (random) selection of cases	Yes	No
Sampling frame required	Yes	No
Ensures representation of key strata	Yes	Yes

Source: Schutt (2009, p. 173).

Purposive Sampling

In **purposive sampling,** each sample element is selected for a purpose, usually because of the unique position of the sample elements. Purposive sampling may involve studying the entire population of some limited group (middle school principals) or a subset of a population (high school guidance counselors who deal with ninth graders). Or a purposive sample may be a "key informant survey," which targets individuals who are particularly knowledgeable about the issues under investigation.

Herbert Rubin and Irene Rubin (1995) suggest three guidelines for selecting informants when designing any purposive sampling strategy. Informants should be:

- "Knowledgeable about the cultural arena or situation or experience being studied."
- "Willing to talk."
- "Represent[ative of] the range of points of view." (p. 66)

In addition, Rubin and Rubin (1995) suggest continuing to select interviewees until you can pass two tests:

- Completeness. "What you hear provides an overall sense of the meaning of a concept, theme, or process." (p. 72)
- Saturation. "You gain confidence that you are learning little that is new from subsequent interview[s]." (p. 73)

Adhering to these guidelines will help to ensure that a purposive sample adequately represents the setting or issues studied.

Of course, purposive sampling does not produce a sample that represents some larger population, but it can be exactly what is needed in a case study of a school, community, or some other clearly defined and relatively limited group. In an intensive case study of a large high school, a purposive sample of school leaders might be complemented with a probability sample of all school staff and students.

Snowball Sampling

Snowball sampling is useful for hard-to-reach or hard-to-identify populations for which there is no sampling frame, but the members of which are somewhat interconnected (at least some members of the population know each other). It can be used to sample members of such groups as closeted gay educators, informal organizational leaders, and students who have suffered from bullying. It also may be used for charting the relationships among members of some group (a sociometric study), exploring the population of interest prior to developing a formal sampling plan, and developing what becomes a census of informal leaders of small organizations or communities. However, researchers using snowball sampling normally cannot be confident that their sample represents the total population of interest, so generalizations must be tentative.

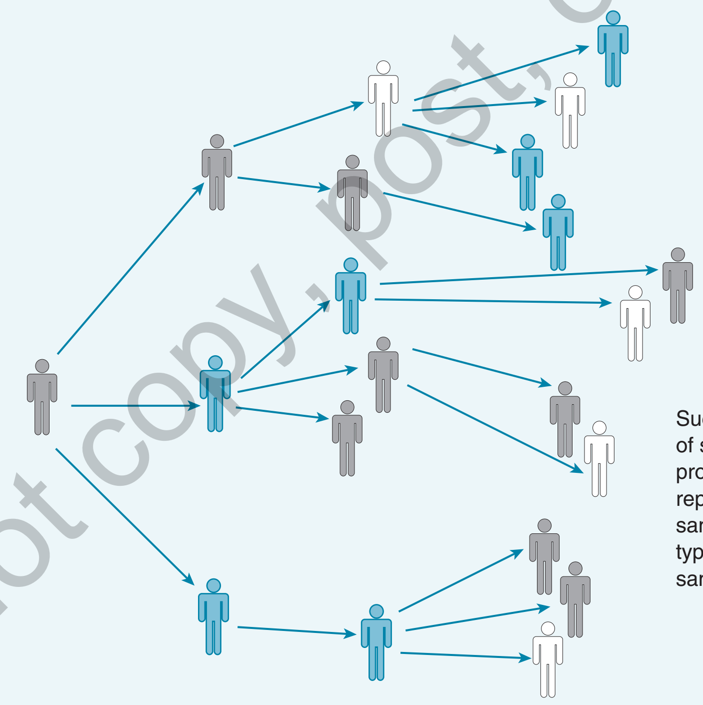
Yvonne Spicer (2004) used both purposive and snowball sampling to study a population that was hard to identify and had members who were known to each other: African American female K–12 public school principals in Massachusetts. Her review of previous research led her to conclude that

> much of the research literature on educational leadership is dominated by the experience of men and White women. When African American women are discussed. . . , this tends to be in the context of White women and minorities as a whole. (p. 77)

Spicer (2004) initially assumed that local school districts, the state department of education, or the state principals' organization would have a list of African American women principals. This proved not to be the case, a circumstance that suggested to her that, as the literature predicted, these women leaders were institutionally invisible. To locate her target population, Spicer created a purposive sample "based on the recommendations of nominators from professional organizations, colleagues, and district data" (p. 77). Once she had contacted subjects identified by the nominators, Spicer used snowball sampling to increase her list, asking the recommended principals to put her in contact with others who met her criteria. Eventually she located 65 such women across the entire state (Spicer, 2004, p. 77).

Spicer's target population had a strong reason for assisting her in locating as many respondents as possible: They felt invisible as a group and wanted to help her get the word out that they existed and were succeeding in their leadership roles. Such an enthusiastic response is not always the case. Snowball sampling can sometimes be problematic because the initial contacts may shape the entire sample and foreclose access to some members of the population of interest. More systematic versions of snowball sampling can reduce the potential for bias. When the sampling is repeated through several waves, with new respondents bringing in more peers, the composition of the sample converges on a more representative mix of characteristics than would occur with uncontrolled snowball sampling. Exhibit 5.8 shows how the sample spreads out through successive recruitment waves to an increasingly diverse pool (Heckathorn, 1997, p. 178).

Exhibit 5.8 **Respondent-Driven Sampling**



Successive waves of sampling gradually produce a more representative sample than is typical of snowball sampling.

Instructions to respondents:
"We'll pay you $5 each for up to three names, but only one of those names can be somebody from your own town. The others have to be from somewhere else."

Source: Based on Heckathorn (1997, p. 178).

Lessons About Sample Quality

Some lessons are implicit in our evaluations of the samples in this chapter. You should keep these lessons in mind when you read articles that use sampling, as well as when conducting your own research:

- We can't evaluate the quality of a sample if we don't know what population it is supposed to represent. If the population is unspecified because the researchers were never clear about the population they were trying to sample, then we can safely conclude that the sample itself is no good.
- We can't evaluate the quality of a sample if we don't know how cases in the sample were selected from the population. If the method was specified, we then need to know whether cases were selected in a systematic fashion and on the basis of chance. In any case, we know that a haphazard method of sampling (as in person-on-the-street interviews) undermines generalizability.
- Sample quality is determined by the sample actually obtained, not just by the sampling method itself. If many of the people selected for our sample are nonrespondents or people (or other entities) who do not participate in the study even though they have been selected for the sample, the quality of our sample is undermined—even if we chose the sample in the best possible way.
- We need to be aware that even researchers who obtain very good samples may talk about the implications of their findings for some group that is larger than, or just different from, the population they actually sampled. For example, findings from a representative sample of students in one university often are discussed as if they tell us about university students in general. And maybe they do; we just don't know for sure.

Generalizability in Qualitative Research

Qualitative research often focuses on populations that are hard to locate or very limited in size. In consequence, nonprobability sampling methods such as availability sampling and snowball are often used. Janet Ward Schofield (2002) suggests ways of increasing the generalizability of the samples obtained in such situations:

> *Studying the Typical.* Choosing sites on the basis of their fit with a typical situation is far preferable to choosing on the basis of convenience. (p. 181)
>
> *Performing Multisite Studies.* A finding emerging repeatedly in the study of numerous sites would appear to be more likely to be a good working hypothesis about some as yet unstudied site than a finding emerging from just one or two sites. . . . Generally speaking, a finding emerging from the study of several very heterogeneous sites would be more . . . likely to be useful in understanding various other sites than one emerging from the study of several very similar sites. (p. 184)

The effort of some qualitative researchers to understand the particulars of a situation in depth, as an important object of inquiry in itself, also leads some to question the value of generalizability as most researchers understand it. In the words of researcher Norman Denzin,

> The interpretivist rejects generalization as a goal and never aims to draw randomly selected samples of human experience. . . . Every instance of social interaction . . . represents a slice from the life world that is the proper subject matter for interpretive inquiry. (cited in Schofield, 2002, p. 173)

2Sampling Distributions

Sampling distribution: A hypothetical distribution of a statistic (e.g., proportion, mean) across an infinite number of random samples that could be drawn from a population.

A well-designed probability sample is one that is likely to be representative of the population from which it was selected. But as you've seen, random samples are subject to sampling error due just to chance. To deal with that problem, educational researchers take into account the properties of a sampling distribution, a hypothetical distribution of a statistic across an infinite number of random samples that could be drawn from a population. Any single random sample can be thought of as just one of an infinite number of random samples that, in theory, could have been selected from the population. If we had the finances of Bill Gates, had all the time in the world, and were able to draw an infinite number of samples, and we calculated the same type of statistic for each of these samples, we would then have a sampling distribution. Understanding sampling distributions is the foundation for understanding how statisticians can estimate sampling error.

What does a sampling distribution look like? Because a sampling distribution is based on some statistic calculated for different samples, we need to choose a statistic. Let's focus on the arithmetic average, or mean. We will explain the calculation of the mean in Chapter 13, but you may already be familiar with it: You add up the values of all the cases and divide by the total number of cases. Let's say you draw a random sample of 500 families and find that their average (mean) family income is $58,239. Imagine that you then draw another random sample. That sample's mean family income might be $60,302. Imagine marking these two means on graph paper and then drawing more random samples and marking their means on the graph. The resulting graph would be a sampling distribution of the mean.

Estimating Sampling Error

Inferential statistics: A mathematical tool for estimating how likely it is that a statistical result based on data from a random sample is representative of the population from which the sample is assumed to have been selected.

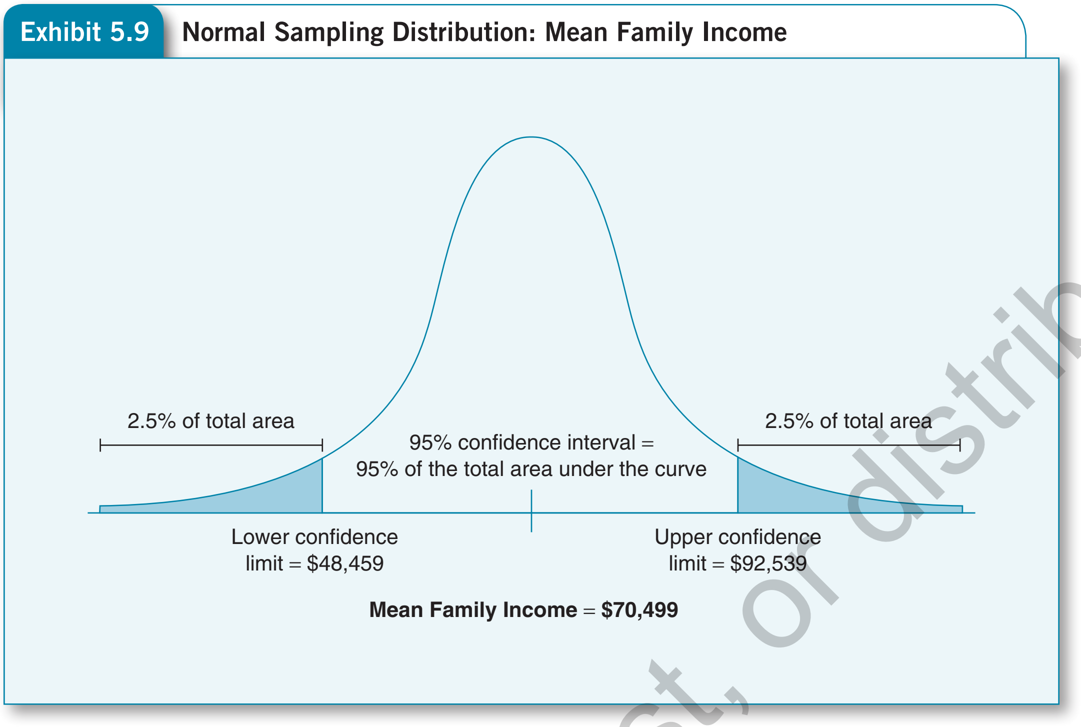
Random sampling error (chance sampling error): Differences between the population and the sample that are due only to chance factors (random error), not to systematic sampling error. Random sampling error may or may not result in an unrepresentative sample. The magnitude of sampling error due to chance factors can be estimated statistically.

We don't actually observe sampling distributions in real research. It would take too much time and money. Instead, researchers just draw the best sample they can. (Now you understand why it is important to have a sample that is representative of the population.) A sampling distribution usually remains a hypothetical or theoretical distribution. We can use the properties of sampling distributions, however, to calculate the amount of sampling error that was likely with the actual random sample used in a study. The tool for calculating sampling error is called **inferential statistics.**

Sampling distributions for many statistics, including the mean, have a "normal" shape. A graph of a normal distribution looks like a bell, with one "hump" in the middle, centered on the population mean, and a number of cases tapering off on both sides of the mean. Note that a normal distribution is symmetrical: If you fold it in half at its center (at the population mean), the two halves will match perfectly. This shape is produced by **random sampling error**—variation due purely to chance. The value of the statistic varies from sample to sample because of chance, so higher and lower values are equally likely.

Exhibit 5.9 shows what the sampling distribution of family incomes would look like if it formed a perfectly normal distribution—this would occur if, rather than 50 random samples, we had selected thousands of random samples.

Exhibit 5.9 **Normal Sampling Distribution: Mean Family Income**



Source: Schutt (2009, p. 179).

As we have stated earlier, in research, our goal is to generalize what we observe in our sample to the population of interest. In statistical terminology, a sample statistic is an estimate of the population parameter we want to estimate. The properties of a sampling distribution facilitate this process of statistical inference. In the sampling distribution, the most frequent value of the **sample statistic**—the statistic (such as the mean) computed from sample data—is identical to the **population parameter**—the statistic computed for the entire population. In other words, we can have a lot of confidence that the value at the peak of the bell curve represents the norm for the entire population.

In a normal distribution, a predictable proportion of cases falls within certain ranges under the curve. Inferential statistics takes advantage of this feature and allow us to estimate how likely it is that, given a particular sample, the true population value will be within some range of the statistic. For example, a statistician might conclude from a sample of 30 families that *we can be 95% confident that the true mean family income in the total population is between $48,459 and $92,539.* The interval from $48,459 to $92,539 would then be called the "95% confidence interval for the mean." The lower ($48,459) and upper ($92,539) bounds of this interval are termed the *confidence limits.* Exhibit 5.9 marks such confidence limits, indicating the range that encompasses 95% of the area under the normal curve; 95% of all sample means would fall within this range.

Although all normal distributions have these same basic features, they differ in the extent to which they cluster around the mean. A sampling distribution is more compact when it is based on larger samples. Stated another way, we can be more confident in estimates based on larger random samples because we know that a larger sample creates a more compact sampling distribution. This should make intuitive sense. If you want to estimate the average family income of residents of New York City and you select a sample of only 30 families, the mean you estimate from this sample of 30 is less likely to reflect New York City residents' family income than a sample of, say, 3,000.

In most social science disciplines, including education, researchers typically rely on 95% or 99% confidence intervals. In fact, every time you read the results of an opinion poll in the newspaper or hear about one on a news broadcast, you are really being given a confidence interval. For example, when a newspaper reports that 30% of high school seniors have used marijuana in the past 6 months, the reporter will also usually add the phrase "plus or minus four percentage points." This is a confidence interval. These conventional confidence limits reflect the conservatism inherent in classical statistical inference: Don't make an inferential statement unless you are very confident (at least 95% confident) that it is correct.

We will explain how to calculate confidence intervals in Chapter 13. You will find it easier to understand this procedure after you have learned some of the basic statistics we introduce in that chapter. If you have already completed a statistics course, you might want to turn now to Chapter 13's confidence interval section for a quick review. In any case, you should now have a sense of how researchers make inferences from a random sample of a population.

Determining Sample Size

You have learned that more confidence can be placed in the generalizability of statistics from larger samples, so you may be eager to work with random samples that are as large as possible. Unfortunately, researchers often cannot afford to sample a very large number of cases. Therefore, they try to determine during the design phase of their study how large a sample they must have to achieve their purposes. They have to consider the degree of confidence desired, the homogeneity of the population, the complexity of the analysis they plan, and the expected strength of the relationships they will measure.

- The less sampling error desired, the larger the sample size must be.
- Samples of more homogeneous populations can be smaller than samples of more diverse populations. Stratified sampling uses prior information on the population to create more homogeneous population strata from which the sample can be selected, so stratified samples can be smaller than simple random samples.
- If the only analysis planned for a survey sample is to describe the population in terms of a few variables, a smaller sample is required than if a more complex analysis involving sample subgroups is planned. If much of the analysis will focus on estimating the characteristics of subgroups within the sample, it is the size of the subgroups that must be considered, not the size of the total sample (Levy & Lemeshow, 1999, p. 74).
- When the researchers will be testing hypotheses and expect to find very strong relationships among the variables, they will need a smaller sample to detect these relationships than if they expect weaker relationships.

Researchers can make more precise estimates of the sample size required through a method called "statistical power analysis" (Hedges & Rhoads, 2010). Statistical power analysis requires a good advance estimate of the strength of the hypothesized relationship in the population. In addition, the math is complicated, so it helps to have some background in mathematics or to be able to consult a statistician. For these reasons, many researchers do not conduct formal power analyses when deciding how many cases to sample.

2Conclusions

The question that began this chapter—"How can we take a valid sample in an educational setting?"—was not linked to a particular research study. Rather, it addressed the overall issue of how we can make valid generalizations in educational research. Generalizing from smaller samples to larger populations is important because frequently we lack both the time and the money to undertake surveys involving thousands or tens of thousands of people.

Fortunately, sampling is a powerful tool for educational research. Probability sampling methods allow a researcher to use the laws of chance, or probability, to draw samples from which population parameters can be estimated with a high degree of confidence. A sample of just 1,000 or 1,500 individuals can be used to estimate reliably the characteristics of the population of a nation comprising millions of individuals.

But researchers do not come by representative samples easily. Well-designed samples require careful planning, some advance knowledge about the population to be sampled, and adherence to systematic selection procedures—all so that the selection procedures are not biased. And even after the sample data are collected, the researcher's ability to generalize from the sample findings to the population is not completely certain. The best that he or she can do is to perform additional calculations that state the degree of confidence that can be placed in the sample statistic.

The alternatives to random, or probability-based, sampling methods are almost always much less palatable for quantitative studies, even though they typically are much cheaper. Without a method of selecting cases likely to represent the population in which the researcher is interested, research findings will have to be carefully qualified. Qualitative researchers whose goal is to understand a small group or setting in depth may necessarily have to use unrepresentative samples, but they must keep in mind that the generalizability of their findings will not be known. Additional procedures for sampling in qualitative studies will be introduced in Chapter 8.

Educational researchers often seek to generalize their conclusions from the population that they studied to some larger target population. The validity of generalizations of this type is necessarily uncertain, because having a representative sample of a particular population does not at all ensure that what we find will hold true in other populations. Nonetheless, as you will see in Chapter 13, the cumulation of findings from studies based on local or otherwise unrepresentative populations can provide important information about broader populations.

Key Terms

Availability sampling	102
Census	96
Cluster	99
Cluster sampling	99
Disproportionate stratified sampling	101
Element	92
Enumeration units	92
Inferential statistics	108
Nonprobability sampling method	96
Nonrespondent	97
Periodicity	98
Population	91
Population parameter	109
Probability of selection	96
Probability sampling method	96
Proportionate stratified sampling	101
Purposive sampling	104
Quota sampling	103
Random number table	98
Random sampling	96
Random sampling error	108
Representative sample	95
Sample	92
Sample statistic	109
Sampling error	94
Sampling frame	92
Sampling interval	98
Sampling unit	92
Simple random sampling	98
Snowball sampling	105
Stratified random sampling	100
Systematic bias	96
Systematic random sampling	98
Target population	94

Highlights

- Sampling theory focuses on the generalizability of descriptive findings to the population from which the sample was drawn. It also considers whether statements can be generalized from one population to another.
- Sampling is unnecessary when the elements that would be sampled are identical, but the complexity of the educational world makes it difficult to argue very often that different elements are identical. Conducting a complete census of a population also eliminates the need for sampling, but the resources required for a complete census of a large population are usually prohibitive.
- Nonresponse undermines sample quality: It is the obtained sample, not the desired sample, that determines sample quality.
- Probability sampling methods rely on a random selection procedure to ensure no systematic bias in the selection of elements. In a probability sample, the odds of selecting elements are known, and the method of selection is carefully controlled.
- A sampling frame (a list of elements in the population) is required in most probability sampling methods. The adequacy of the sampling frame is an important determinant of sample quality.
- Simple random sampling and systematic random sampling are equivalent probability sampling methods in most situations.
- Stratified random sampling uses prior information about a population to make sampling more efficient. Stratified sampling may be either proportionate or disproportionate.
- Cluster sampling is less efficient than simple random sampling but is useful when a sampling frame is unavailable. It is also useful for large populations spread out across a wide area or among many organizations.
- Nonprobability sampling methods can be useful when random sampling is not possible, when a research question does not concern a larger population, and when a preliminary exploratory study is appropriate. However, the representativeness of nonprobability samples cannot be determined.
- The likely degree of error in an estimate of a population characteristic based on a probability sample decreases when the size of the sample and the homogeneity of the population from which the sample was selected increase. Sampling error is not affected by the proportion of the population that is sampled, except when that proportion is large.

Student Study Site

To assist in completing the web exercises, please access the study site at www.sagepub.com/check, where you will find the web exercise with accompanying links. You'll find other useful study materials such as self-quizzes and e-flashcards for each chapter, along with a group of carefully selected articles from research journals that illustrate the major concepts and techniques.

Discussion Questions

1. When, if ever, is it reasonable to assume that a sample is not needed because "everyone is the same"—that is, the population is homogeneous?
2. What increases sampling error in probability-based sampling designs? Stratified rather than simple random sampling? Disproportionate (rather than proportionate) stratified random sampling? Stratified rather than cluster random sampling? Why do researchers select disproportionate (rather than proportionate) stratified samples?

Practice Exercises

1. Select a random sample using the table of random numbers in Appendix D. Compute a statistic based on your sample, and compare it to the corresponding figure for the entire population. Here's how to proceed:

a. First, select a very small population for which you have a reasonably complete sampling frame. One possibility would be the list of asking prices for houses advertised in your local paper. Another would be the listing of some characteristic of states in a U.S. Census Bureau publication, such as average income or population size.

b. The next step is to create your sampling frame, a numbered list of all the elements in the population. If you are using a complete listing of all elements, as from a U.S. Census Bureau publication, the sampling frame is the same as the list. Just number the elements (states). If your population is composed of housing ads in the local paper, your sampling frame will be those ads that contain a housing price. Identify these ads, and then number them sequentially, starting with 1.

c. Decide on a method of picking numbers out of the random number table in Appendix D, such as taking every number in each row, row by row (or you may move down or diagonally across the columns). Use only the first (or last) digit in each number if you need to select 1 to 9 cases or only the first (or last) two digits if you want fewer than 100 cases.

d. Pick a starting location in the random number table. It's important to pick a starting point in an unbiased way, perhaps by closing your eyes and then pointing to some part of the page.

e. Record the numbers you encounter as you move from the starting location in the direction you decided on in advance, until you have recorded as many random numbers as the number of cases you need in the sample. If you are selecting states, 10 might be a good number. Ignore numbers that are too large (or small) for the range of numbers used to identify the elements in the population. Discard duplicate numbers.

f. Calculate the average value in your sample for some variable that was measured—for example, population size in a sample of states or housing price for the housing ads. Calculate the average by adding up the values of all the elements in the sample and dividing by the number of elements in the sample.

g. Go back to the sampling frame and calculate this same average for all the elements in the list. How close is the sample average to the population average?

h. Estimate the range of sample averages that would be likely to include 90% of the possible samples.

2. From professional journals, select four articles that describe research using a sample drawn from some population. Identify the type of sample used in each study, and note any strong and weak points in how the sample was actually drawn. Did the researchers have a problem due to nonresponse? Considering the sample, how confident are you in the validity of generalizations about the population based on the sample?

Web Exercises

1. What can you learn about sampling on the Web? Conduct a search on "sampling" and "population" and select a few of these sites. List a few new points that you learn about sampling.

2. Check out the "people" section of the U.S. Census Bureau website: www.census.gov. Based on some of the data you find there, write a brief summary of some aspect of the current characteristics of the American school-age population.

Developing a Research Proposal

Consider the possibilities for sampling.

1. Propose a sampling design that would be appropriate if you were to survey students on your campus only or in the school in which you work. Define the population, identify the sampling frame(s), and specify the elements and any other units at different stages. Indicate the exact procedure for selecting people to be included in the sample.

2. Propose a different sampling design for conducting your survey in a larger population, such as your city, state, or the entire nation.