Qualitative and Quantitative Sampling

Reasons for Sampling
Sampling Strategies
Conclusion

Sampling is a major problem for any type of research. We can't study every case of whatever we're interested in, nor should we want to. Every scientific enterprise tries to find out something that will apply to everything *of a certain kind by studying* a few examples, *the results of the study being, as we say, "generalizable."*

—Howard Becker, *Tricks of the Trade,* p. 67

In *Promises I Can Keep*, an in-depth study of low-income mothers, Edin and Kefalas (2005) first identified eight low-income neighborhoods in the Philadelphia, Pennsylvania, area through extensive qualitative fieldwork and quantitative analysis of census data. Each neighborhood met three selection criteria: at least 20 percent of householders were below the poverty line, at least 20 percent of all households had a single parent, and each had a large number of Black, White, and Hispanic residents. In each neighborhood, Edin and Kefalas recruited half of the mothers to interview through referrals from local experts (teachers, social workers, public nurses, clergy, business owners, and public housing officials) and half by posting fliers on public phone booths or personally contacting mothers on street corners. All mothers had incomes putting them below the poverty line in the previous year. Edin and Kefalas tried to get a mixture: 50 Whites, 50 African Americans, and 50 Puerto Ricans, and tried to get one-half over 25 and one-half under 25 years old. They eventually had 162 mothers, 52 whites, 63 African American, and 47 Puerto Rican. Only 40 were over 25 years old, but ages ranged from 15 to 56. They say, "The resulting sample is not random or representative but is quite heterogeneous" (238).

REASONS FOR SAMPLING

When we sample, we select some cases to examine in detail, and then we use what we learn from them to understand a much larger set of cases. Most, but not all, empirical studies use sampling. Depending on the study, the method we use for sampling can differ. Most books on sampling emphasize its use in quantitative research and contain applied mathematics and quantitative examples. The primary use of sampling in quantitative studies is to create a representative sample (i.e., a **sample**, a selected small collection of cases or units) that

Sample A small set of cases a researcher selects from a large pool and generalizes to the population.

closely reproduces or represents features of interest in a larger collection of cases, called the **population**.

We examine data in a sample in detail, and if we sampled correctly, we can generalize its results to the entire population. We need to use very precise sampling procedures to create representative samples in quantitative research. These procedures rely on the mathematics of probabilities and hence, are called *probability sampling*.

In most quantitative studies, we want to see how many cases of a population fall into various categories of interest. For example, we might ask how many in the population of all of Chicago's high school students fit into various categories (e.g., high-income family, single-parent family, illegal drug user, delinquent behavior arrestee, musically talented person). We use probability samples in quantitative research because they are very efficient. They save a lot of time and cost for the accuracy they deliver. A properly conducted probability sample may cost 1/1000 the cost and time of gathering information on an entire population, yet it will yield virtually identical results. Let us say we are interested in gathering data on the 18 million people in the United States diagnosed with diabetes. From a well-designed probability sample of 1,800, we can take what we learned and generalize it to all 18 million. It is more efficient to study 1,800 people to learn about 18 million than to study all 18 million people.

Probability samples can be highly accurate. For large populations, data from a well-designed, carefully executed probability sample are often equally if not more accurate than trying to reach every case in the population, but this fact confuses many people. Actually, when the U.S. government planned its 2000 census, all of the social researchers and statistically trained scientists agreed that probability sampling would produce more accurate data than the traditional census method of trying to count every person. A careful probability sample of 30,000 has a very tiny and known error rate. If we try to locate every single person of 300,000,000, systematic errors will slip in unless we take extraordinary efforts and expend huge amounts of time and money. By the way, the government actually conducted the census in the traditional way, but it was for political, not scientific, reasons.

Sampling proceeds differently in qualitative studies and often has a different purpose from quantitative studies. In fact, using the word *sampling* creates confusion in qualitative research because the term is closely associated with quantitative studies (see Luker, 2008:101). In qualitative studies, to allow us to make statements about categories in the population, we rarely sample to gather a small set of cases that is a mathematically accurate reproduction of the entire population. Instead, we sample to identify relevant categories at work in a few cases. In quantitative sampling, we select cases/units. We then treat them as carriers of aspects/features of the social world. A sample of cases/units "stands in" for the much larger population of cases/units. We pick a few to "stand in" for the many. In contrast, the logic of the qualitative sample is to sample aspects/features of the social world. The aspects/features of our sample highlight or "shine light into" key dimensions or processes in a complex social life. We pick a few to provide clarity, insight, and understanding about issues or relationships in the social world. In qualitative sampling, our goal is to deepen understanding about a larger process, relationship, or social scene. A sample gives us valuable information or new aspects. The aspects accentuate, enhance, or enrich key features or situations. We sample to open up new theoretical insights, reveal distinctive aspects of people or social settings, or deepen understanding of complex situations, events, or relationships. In qualitative research, "it is their relevance to the research topic rather than their representativeness which determines the way in which the people to be studied are selected" (Flick, 1998: 41).

We should not overdo the quantitative-qualitative distinction. In a few situations, a study that is primarily quantitative will use the qualitative-sampling

Population The abstract idea of a large group of many cases from which a researcher draws a sample and to which results from a sample are generalized.

strategy and vice versa. Nonetheless, most quantitative studies use probability or probability-like samples while most qualitative studies use a nonprobability method and nonrepresentative strategy.

SAMPLING STRATEGIES

We want to avoid two types of possible sampling mistakes. The first is to conduct sampling in a sloppy or improper manner; the second is to choose a type of sample inappropriate for a study's purpose. The first mistake reminds us to be very meticulous and systematic when we sample. To avoid the second mistake, we need a sampling strategy that matches our specific study's purpose and data. Sampling strategies fall into two broad types: a sample that will accurately represent the population of cases, and all others. We primarily use the first strategy in quantitative studies and the latter in qualitative studies.

Strategies When the Goal Is to Create a Representative Sample

In a representative sample, our goal is to create sample data that mirror or represent many other cases that we cannot directly examine. We can do this in two ways. The first is the preferred method and considered the "gold standard" for representative samples, the *probability sample*. It builds on more than a century of careful reasoning and applied mathematics plus thousands of studies in natural science and quantitative social science. With a probability sampling strategy, we try to create an accurate representative sample that has mathematically predictable errors (i.e., precisely known chances of being "off target"). This sampling approach is complex with several subtypes. Before we examine it, let us look at the second, simpler way to produce a representative sample: to use a nonprobability technique. It is a less accurate substitute when we want a representative sample; however, it is acceptable when probability sampling is impossible, too costly, time consuming, or impractical.

Convenience sampling A nonrandom sample in which the researcher selects anyone he or she happens to come across.

Nonprobability Sampling Techniques. Ideally, we would prefer probability samples when we want to create a representative sample, as a less demanding alternative there are two nonprobability alternatives: convenience and quota samples. In **convenience sampling** (also called *accidental, availability,* or *haphazard sampling*), our primary criteria for selecting cases are that they are easy to reach, convenient, or readily available. This sample type may be legitimate for a few exploratory preliminary studies and some qualitative research studies when our purpose is something other than creating a representative sample. Unfortunately, it often produces very nonrepresentative samples, so it is *not recommended* for creating an accurate sample to represent the population.

When we select cases based on convenience, our sample can seriously misrepresent features in the entire population.[1] You may ask why, if this method is so bad and samples can be seriously nonrepresentative, anyone would use it. The reason is simple: convenience samples are easy, cheap, and quick to obtain. Another reason might be that people are ignorant about how to create a good representative sample. An example of such sampling is the person-on-the-street interview conducted by television programs. Television interviewers go out on the street with camera and microphone to talk to a few people who are convenient to interview. The people walking past a television studio in the middle of the day do not represent everyone. Likewise, television interviewers tend to pick people who look "normal" to them and avoid people who are unattractive, disabled, impoverished, elderly, or inarticulate. Another example is a newspaper that asks readers to clip a questionnaire and mail it in, a Web site that asks users to click on a choice, or a television program that asks viewers to call in their choices. Such samples may have entertainment value, but they easily yield highly misleading data that do not represent the population even when a large number of people respond.

Maybe you wonder what makes such a sample nonrepresentative. If you want to know about everyone in city XYZ that has a population of 1 million, only some read the newspaper, visit a Web site, or tuned into a program. Also, not everyone who is reading the newspaper, visiting the Web site, or has tuned in is equally interested in an issue. Some people will respond, and there may be many of them (e.g., 50,000), but they are self-selected. We cannot generalize accurately from self-selected people to the entire population. Many in the population do not read the newspaper, visit specific Web sites, or tune into certain television programs, and even if they did, they may lack the interest and motivation to participate. Two key ideas to remember about representative samples are that: (1) self-selection yields a nonrepresentative sample and (2) a big sample size alone is not enough to make a sample representative.

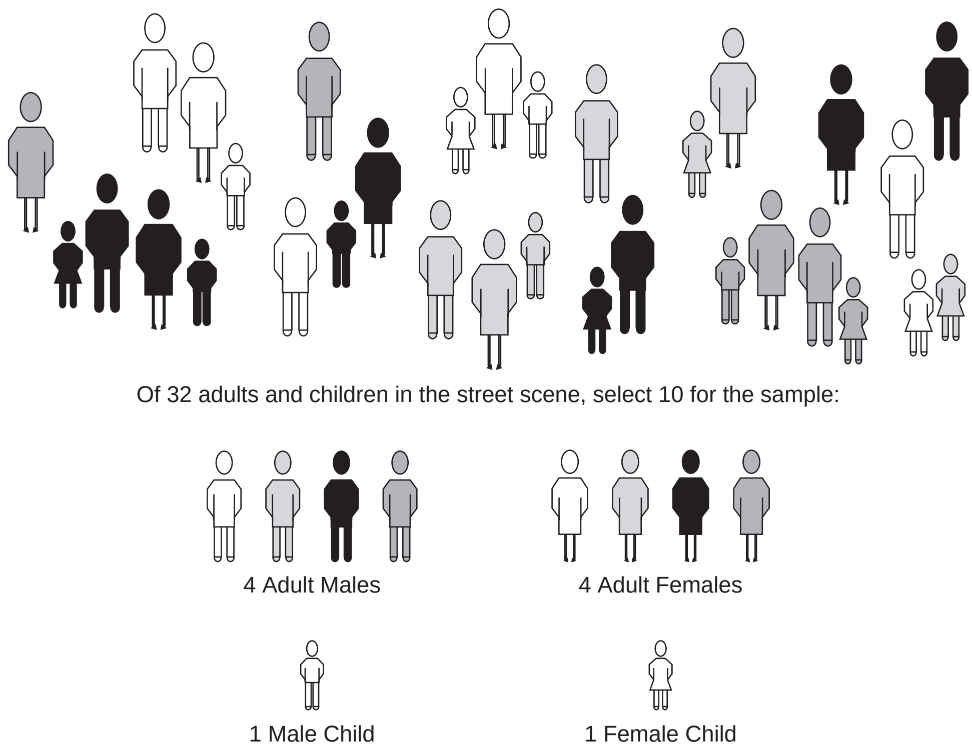
For many purposes, well-designed **quota sampling** is an acceptable nonprobability substitute method for producing a quasi-representative sample.[2] In quota sampling, we first identify relevant categories among the population we are sampling to capture diversity among units (e.g., male and female; or under age 30, ages 30 to 60, over age 60). Next we determine how many cases to get for each category—this is our "quota." Thus, we fix a number of cases in various categories of the sample at the start.

Let us return to the example of sampling residents from city XYZ. You select twenty-five males and twenty-five females under age 30 years of age, fifty males and fifty females aged 30 to 60, and fifteen males and fifteen females over age 60 for a 180-person sample. While this is a start as a population's diversity, it is difficult to represent all possible population characteristics accurately (see Figure 1). Nonetheless, quota sampling ensures that a sample has some diversity. In contrast, in convenience sampling, everyone in a sample might be of the same age, gender, or background. The description of sampling in the *Promises I Can Keep* study at the opening of this chapter used quota sampling (also see Example Box 1, Quota Samples).

Quota sampling is relatively easy. My students conducted an opinion survey of the undergraduate student body using quota sampling. We used three quota categories—gender, class, and minority/majority group status—and a convenience selection method (i.e., a student interviewer approached anyone in the library, a classroom, the cafeteria). We set the numbers to be interviewed in each quota category in advance: 50 percent males and 50 percent females; 35 percent freshman, 25 percent sophomores, 20 percent juniors, and 20 percent seniors; and 10 percent minority and 90 percent majority racially. We picked the proportions based on approximate representation in the student body according to university official records. In the study, a student interviewer approached a person, confirmed that he or she was a student, and verified his or her gender, class, and minority/majority status. If the person fit an unfilled quota (e.g., locate five freshman males who are racial-ethnic minorities), the person was included in the sample and the interviewer proceeded to ask survey questions. If the person did not fit the quota, the interviewer quickly thanked the person without asking survey questions and moved on to someone else.

Quota samples have three weaknesses. First, they capture only a few aspects (e.g., gender and age) of all population diversity and ignore others (e.g., race-ethnicity, area of residence in the city, income level). Second, the fixed number of cases in each category may not accurately reflect the proportion of cases in the total population for the category. Perhaps 20 percent of city residents are over 60 years old but are 10 percent of a quota. Lastly, we use convenience sampling selection for each

Quota sampling A nonrandom sample in which the researcher first identifies general categories into which cases or people will be placed and then selects cases to reach a predetermined number in each category.



Note: Shading indicates various skin tones.

FIGURE 1 Quota Sampling

quota category. For example, we include the first twenty-five males under age 30 we encounter—even if all twenty-five are high-income White lawyers who just returned from a seminar on financial investments. Nothing prevents us from sampling only "friendly"-acting people who want us to pick them.

Probability Sampling Techniques. Probability sampling is the "gold standard" for creating a representative sample. It has a specialized vocabulary that may make it difficult to understand until you learn it, so next we will review some of its vocabulary.

Sampling element The name for a case or single unit to be sampled.

The Language of Probability Sampling. You draw a sample from a large collection of cases/units. Each case/unit is your **sampling element**. It is the unit of analysis or a case in a population. It could be a person, a family, a neighborhood, a nation, an organization, a written document, a symbolic message (television commercial, display of a flag), or a social action (e.g., an arrest, a divorce, or a kiss).

The large collection is the population, but sometimes the word *universe* is used. To define the population, you specify the elements and identify

EXAMPLE BOX 1

Quota Samples

Two studies illustrate different uses of quota sampling in quantitative research. In a study, McMahon, McAlaney, and Edgar (2007) wanted to examine public views of binge drinking in the United Kingdom. They noted that most past research was on young adults and campaigns to curb binge drinking had been ineffective. The authors wanted to learn about public perceptions of binge drinking among the entire adult population. They developed a survey that asked how people defined binge drinking, the extent to which they saw it as a concern, and reasons for and solutions to it. They combined quota sampling with another sampling method to interview 586 people in one city (Inverclyde, Scotland). For quota sampling, interviewers approached potential participants in the streets surrounding a shopping center and invited them to take part in the survey. The quota was based on getting a balance of gender and six age categories. The other method was to go door-to-door in several low-income neighborhoods. The authors learned useful information about views on binge drinking across age groups in both genders in one city. They found wide variation in definitions of binge drinking and support for a "false consensus effect" in which a small number of the heaviest drinkers see their behavior as normal and socially accepted. Nonetheless, the sample is not representative, so findings on the extent of binge drinking in the public and views about it may not reflect the behaviors or views within the city's overall population.

A second study in China by Bian, Breiger, Davis, and Galaskiewicz (2005) employed a targeted use of quota sampling. Their interest was in the difference between the social networks and social ties (e.g., friends, family) among people in different social classes in major Chinese cities. They selected households in four of China's largest metropolitan areas (Shanghai, Shenzhen, Tianjin, and Wuhan), identified a set of neighborhoods in each, and then sampled 100 people per city. They had a list of thirteen occupational titles that represented the full range of the class system in China and 88 percent of all working people in the four cities. Their quota was to get an equal number in each city and a sufficient number of households in each of the thirteen occupational categories for careful analysis. Thus, only 4 percent of the people held the position as manager, but nearly 10 percent of the sample were managers, and 40 percent of people held an industrial worker occupation, but close to 10 percent of people in the sample were industrial workers. The study goal was to test hypotheses about whether a household's social ties are with others of similar or different social classes. They asked households to maintain a written log of social visits (in person or via phone) with other people and recorded the occupation of visitors. This process lasted a year, and researchers interviewed people every three months. The primary interest in the study was to compare patterns of social networks across the various social classes. For example, did managers socialize only with other managers or with people from a wide range of classes? Did industrial workers socialize with industrial workers as well as people in various lower occupations but not in higher occupations? Because the study goal was to compare social network patterns across the various classes, not to have a representative sample that described the Chinese population, it was a highly effective use of quota sampling.

---

its geographical and temporal boundaries as well as any other relevant boundaries.

Most probability studies with large samples of the entire U.S. population have several boundaries. They include adults over 18 who are residents of the forty-eight continental states and exclude the institutionalized population (i.e., people in hospitals, assisted living and nursing homes, military housing, prisons and jails, homeless and battered women's shelters, college dormitories). Ignoring people in Alaska, Hawaii, and Puerto Rico and excluding the institutionalized population can throw off statistics—for example, the unemployment rate would be higher if the millions of people in prison were included in calculations (see Western and Pettit, 2005). Many studies include only English speakers, yet as of 2007, roughly 5 percent of U.S. households were "linguistically isolated" (no one over 14 spoke English very well (U.S. Census Bureau, 2007).

To draw a probability sample we start with a population, but *population* is an abstract concept. We must conceptualize and define it more precisely in a process similar to conceptualization in the measurement process, for example, all people in Tampa, Florida, or all college students in the state of Nevada. A **target population** is the specific collection of elements we will study (e.g., noninstitutionalized persons 18 years of age and older with legal residences with the city limits of Tampa on May 15, 2011; students enrolled full-time in an accredited two- or four-year postsecondary educational facility in the state of Nevada in October 2010). In some ways, the target population is analogous to our use of a conceptual definition of the measurement process.

Populations are in constant motion, so we need a temporal boundary. For example, in a city at any given moment, people are dying, boarding or getting off airplanes, and driving across city boundaries in cars. Whom should we count? Do we exclude a long-time city resident who happens to be on vacation when the time is fixed? A population (e.g., persons over the age of 18 who are in the city limits of Milwaukee, Wisconsin, at 12:01 A.M. on March 1, 2011), is an abstract idea. It exists in the mind but is difficult to pinpoint concretely (see Example Box 2, Examples of Populations).

After we conceptualize our population, we need to create an operational definition for the abstract population idea in a way that is analogous to operationalization in the measurement process. We turn the abstract idea into an empirically concrete specific list that closely approximates all population elements. This is our **sampling frame**.

There are many types of sampling frames: telephone directories, tax records, driver's license records, and so on. Listing the elements in a population sounds simple, but it is often difficult because often there is no accurate, up-to-date list of all elements in a population.

A good sampling frame is crucial for accurate sampling. Any mismatch between a sampling frame and the conceptually defined population can create errors. Just as a mismatch between our theoretical and operational definitions of a variable weakens measurement validity, a mismatch between the abstract population and the sampling frame weakens our sampling validity. The most famous case in the history of sampling involved an issue of sampling frames.[3] (See Expansion Box 1, Sampling Frames and the History of Sampling.)

Let us say that our population is all adult residents in the Pacific coast region of the United States in 2010. We contact state departments of transportation to obtain lists of everyone with a driver's

Target population The concretely specified large group of many cases from which a researcher draws a sample and to which results from the sample are generalized.

Sampling frame A list of cases in a population, or the best approximation of them.

EXAMPLE BOX 2

Examples of Populations

1. All persons ages 16 or older living in Australia on December 2, 2009, who were not incarcerated in prison, asylums, and similar institutions
2. All business establishments employing more than 100 persons in Ontario Province, Canada, that operated in the month of July 2005
3. All admissions to public or private hospitals in the state of New Jersey between August 1, 1988, and July 31, 1993
4. All television commercials aired between 7:00 A.M. and 11:00 P.M. Eastern Standard Time on three major U.S. networks between November 1 and November 25, 2004
5. All currently practicing physicians in the United States who received medical degrees between January 1, 1960, and the present
6. All African American male heroin addicts in the Vancouver, British Columbia, or Seattle, Washington, metropolitan areas during 2004

EXPANSION BOX 1

Sampling Frames and the History of Sampling

A famous case in the history of sampling illustrates the limitations of quota sampling and of sampling frames. The *Literary Digest,* a major U.S. magazine, sent postcards to people before the 1920, 1924, 1928, and 1932 U.S. presidential elections. The magazine took the names for its sample from automobile registrations and telephone directories. People returned the postcards indicating for whom they would vote. The magazine correctly predicted all four election outcomes. The magazine's success with predictions was well known, and in 1936, it increased the sample from about 1 million to 10 million. 2.4 million people returned postcards they were sent. The magazine predicted a huge victory for Alf Landon over Franklin D. Roosevelt. But the *Literary Digest* was wrong; Roosevelt won by a landslide. Another random sample of 50,000 by George Gallup was accurate within 1 percent of the result.

The prediction was wrong for several reasons, but the sampling mistakes were central. Although the magazine sampled a very large number of people, its sampling frame did not accurately represent the target population (i.e., all voters). It excluded people without telephones or automobiles, a sizable percentage of the population in 1936. The frame excluded as much as 65 percent of the population, particularly a section of the voting population (lower income) that tended to favor Roosevelt. The magazine had been accurate in earlier elections because people with higher and lower incomes did not differ in the way they voted. Also, during earlier elections before the Great Depression, more low-income people could afford to have telephones and automobiles.

The *Literary Digest* mistake teaches us two lessons. First, an accurate sampling frame is crucial. Second, the size of a sample is less important than how accurately it represents the population. A representative sample of 50,000 can give more accurate predictions about the U.S. population than a nonrepresentative sample of 10 million or 50 million.

license in California, Oregon, and Washington. We know some people do not have driver's licenses, although some people drive illegally without them or do not drive. The lists of people with licenses, even if updated regularly, quickly goes out of date as people move into or out of a state. This example shows that before we use official records, such as driver's licenses, as a sampling frame, we must know how officials produce such records. When the state of Oregon instituted a requirement that people show a social security number to obtain a driver's license, the number applying for licenses dropped by 10 percent (from 105,000 issued over three months of 2007 to 93,000 in the same three months of 2008). Thus, thousands disappeared from official records. We could try income tax records, but not everyone pays taxes. Some people cheat and do not pay, others have no income and do not have to file, others have died or have not begun to pay taxes, and still others have entered or left the area since taxes were due. Voter registration records exclude as much as half of the population. In the United States between 53 and 77 percent of eligible voters are registered (Table 401, Statistical Abstract of the United States, 2009). Telephone directories are worse. Many people are not listed in a telephone directory, some people have unlisted numbers, and others have recently moved. With a few exceptions (e.g., a list of all students enrolled at a university), it is difficult to get a perfectly accurate sampling frame. A sampling frame can include those outside the target population (e.g., a telephone directory that lists people who have moved away) or it may omit those within it (e.g., those without telephones). (See Example Box 3, Sampling Frame.)

The ratio of a sample size to the size of the target population is the **sampling ratio**. If the target

Sampling ratio The number of cases in the sample divided by the number of cases in the population or the sampling frame, or the proportion of the population in a sample.

EXAMPLE BOX 3

Sampling Frame

A study by Smith, Mitchell, Attebo, and Leeder (1997) in Australia shows how different sampling frames can influence a sample. The authors examined 2,557 people aged 49 and over living in a defined post code area recruited from a door-to-door census. Of all addresses, people in 80.9 percent were contacted and 87.9 percent of the people responded. The authors searched the telephone directory and the electoral roll for each person. The telephone directory listed 82.2 percent and the electoral roll contained 84.3 percent. Younger people, those who did not own their own homes, and those born outside of Australia were significantly less likely to be included in either sampling frame. The telephone directory was also likely to exclude people with higher occupational prestige while the electoral roll was likely to exclude unmarried persons and males.

population has 50,000 people and the sample has 150, then the sampling ratio is 150/50,000 = 0.003, or 0.3 percent. For a target population of 500 and sample of 100, the sampling ratio is 100/500 = 0.20, or 20 percent. Usually, we use the number of elements in a sampling frame as our best estimate of the size of the target population.

Parameter A characteristic of the entire population that is estimated from a sample.

Statistic A word with several meanings including a numerical estimate of a population parameter computed from a sample.

Except for small specialized populations (e.g., all students in a classroom), when we do not need to sample, we use data from a sample to estimate features in the larger population. Any characteristic of a population (e.g., the percentage of city residents who smoke cigarettes, the average height of all women over the age of 21, the percent of people who believe in UFOs) is a population **parameter**. It is the true characteristic of the population. We do not know the parameter with absolute certainty for large populations (e.g., an entire nation), so we can estimate it by using sample data. Information in the sample used to estimate a population parameter is called a **statistic**. (See Figure 2.)

Random Sampling

In applied mathematics, probability theory relies on random processes. The word *random* has several meanings. In daily life, it can mean unpredictable, unusual, unexpected, or haphazard. In mathematics, random has a specific meaning: a selection process without any pattern. In mathematics, random processes mean that each element will have an equal probability of being selected. We can mathematically calculate the probability of outcomes over many cases with great precision for true random processes.

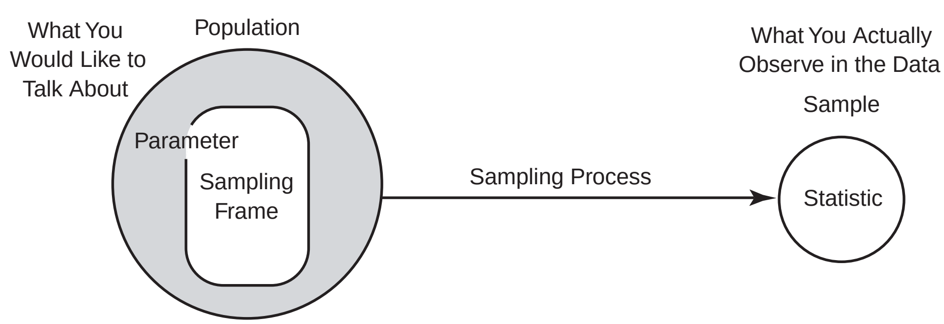


FIGURE 2 **A Model of the Logic of Sampling**

Random samples yield samples most likely to truly represent the entire population. They also allow us to calculate statistically the relationship between the sample and the population—that is, the size of the **sampling error**. The sampling error is the deviation between what is in the sample data and an ideal population parameter due to random processes.

Probability samples rely on random selection processes. Random selection for sampling requires more precision, time, and effort than samples with nonrandom selection. The formal mathematical procedure specifies exactly which person to pick for the sample, and it may be very difficult to locate that specific person! In sampling, random is not anyone, nor does it mean thoughtless or haphazard. For example, if we are using true random sampling in a telephone survey, we might have to call back six or seven times at different times of the days and on different days, trying to get a specific person whom the mathematically random process identified.[4]

This chapter does not cover all technical and statistical details of random sampling. Instead, we discuss the fundamentals of how probability sampling works, the difference between good and bad samples, how to draw a sample, and basic principles of sampling in social research. If you plan to pursue a career in quantitative research, you will need more mathematical and statistical background on probability and sampling than space permits here.

Five Ways to Sample Randomly

Simple Random. All probability samples are modeled on the **simple random sample** that first specifies the population and target population and identifies its specific sampling elements (e.g., all households in Prescott, Arizona, in March 2011). Next we create an accurate sampling frame and we then use a true random process (discussed later) to pick elements from the sampling frame. Beyond creating an accurate sampling frame, the next difficulty is that we must locate the specific sampled element selected by a random process. If the sampled element is a household, we may have to revisit or call back five times to contact that specific selected household.

To select elements from a sampling frame, we will need to create a list of random numbers that will tell us which elements on it to select. We will need as many random numbers as there are elements to be sampled. The random numbers should range from 1 (the first element on the sampling frame) to the highest number in our sampling frame. If the sampling frame lists 15,000 households, and we want to sample 150 from it, we need a list of 150 random numbers (i.e., numbers generated by a true random process, from 1 to 15,000).

There are two main ways to obtain a list of random numbers. The "old-fashioned" way is to use a **random-number table**. Such tables are available in most statistics and research methods books including this one (see Appendix). The numbers are generated by a pure random process so that any number has an equal probability of appearing in any position. Today most people use computer programs to produce lists of random numbers. Such programs are readily available and often free.

You may ask, once we select an element from the sampling frame, do we then return it to the sampling frame, or do we keep it separate? Unrestricted random sampling is called "random sampling with replacement"—that is, replacing an element after sampling it so it has a chance to be selected again. In simple random sampling without replacement, we "toss out" or ignore elements

Random sample A sample using a mathermatically random method, such as a random-number table or computer program, so that each sampling element of a population has an equal probablity of being selected into the sample.

Sampling error How much a sample deviates from being representative of the population.

Simple random sample A random sample in which a researcher creates a sampling frame and uses a pure random process to select cases so that each sampling element in the population will have an equal probability of being selected.

Random-number table A list of numbers that has no pattern and that researchers use to create a random process for selecting cases and other randomization purposes.

already selected for the sample. For almost all practical purposes in social science, random sampling is without replacement.

We can see the logic of simple random sampling with an elementary example: sampling marbles from a jar. Let us say I have a large jar full of 5,000 marbles, some red and some white. The marble is my sampling element, the 5,000 marbles are my population (both target and ideal), and my sample size is 100. I do not need a sampling frame because I am dealing with small physical objects. The population parameter I want to estimate is the percentage of red marbles in the jar.

I need a random process to select 100 marbles. For small objects, this is easy; I close my eyes, shake the jar, pick one marble, and repeat the procedure 100 times. I now have a random sample of marbles. I count the number of red marbles in my sample to estimate the percentage of red versus white marbles in the population. This is a lot easier than counting all 5,000 marbles. My sample has 52 white and 48 red marbles.

Does this mean that the population parameter is exactly 48 percent red marbles? Maybe or maybe not; because of random chance, my specific sample might be off. I can check my results by dumping the 100 marbles back in the jar, mixing the marbles, and drawing a second random sample of 100 marbles. On the second try, my sample has 49 white marbles and 51 red ones. Now I have a problem. Which is correct? You might ask how good this random sampling business is if different samples from the same population can yield different results. I repeat the procedure over and over until I have drawn 130 different samples of 100 marbles each (see Chart 1 for results). Most people might find it easier to empty the jar and count all 5,000 marbles, but I want to understand the process of sampling. The results of my 130 different samples reveal a clear pattern. The most common mix of red and white marbles is 50/50. Samples that are close to that split are more frequent than those with more uneven splits. The population parameter appears to be 50 percent white and 50 percent red marbles.

Sampling distribution A distribution created by drawing many random samples from the same population.

Central limit theorem A mathematical relationship that states when many random samples are drawn from a population, a normal distribution is formed, and the center of the distribution for a variable equals the population parameter.

Mathematical proofs and empirical tests demonstrate that the pattern found in Chart 1 always appears. The set of many different samples is my **sampling distribution**. It is a distribution of different samples. It reveals the frequency of different sample outcomes from many separate random samples. This pattern appears if the sample size is 1,000 instead of 100, if there are 10 colors of marbles instead of 2, if the population has 100 marbles or 10 million marbles instead of 5,000, and if the sample elements are people, automobiles, or colleges instead of marbles. In fact, the "bell-shaped" sampling distribution pattern becomes clearer as I draw more and more independent random samples from a population.

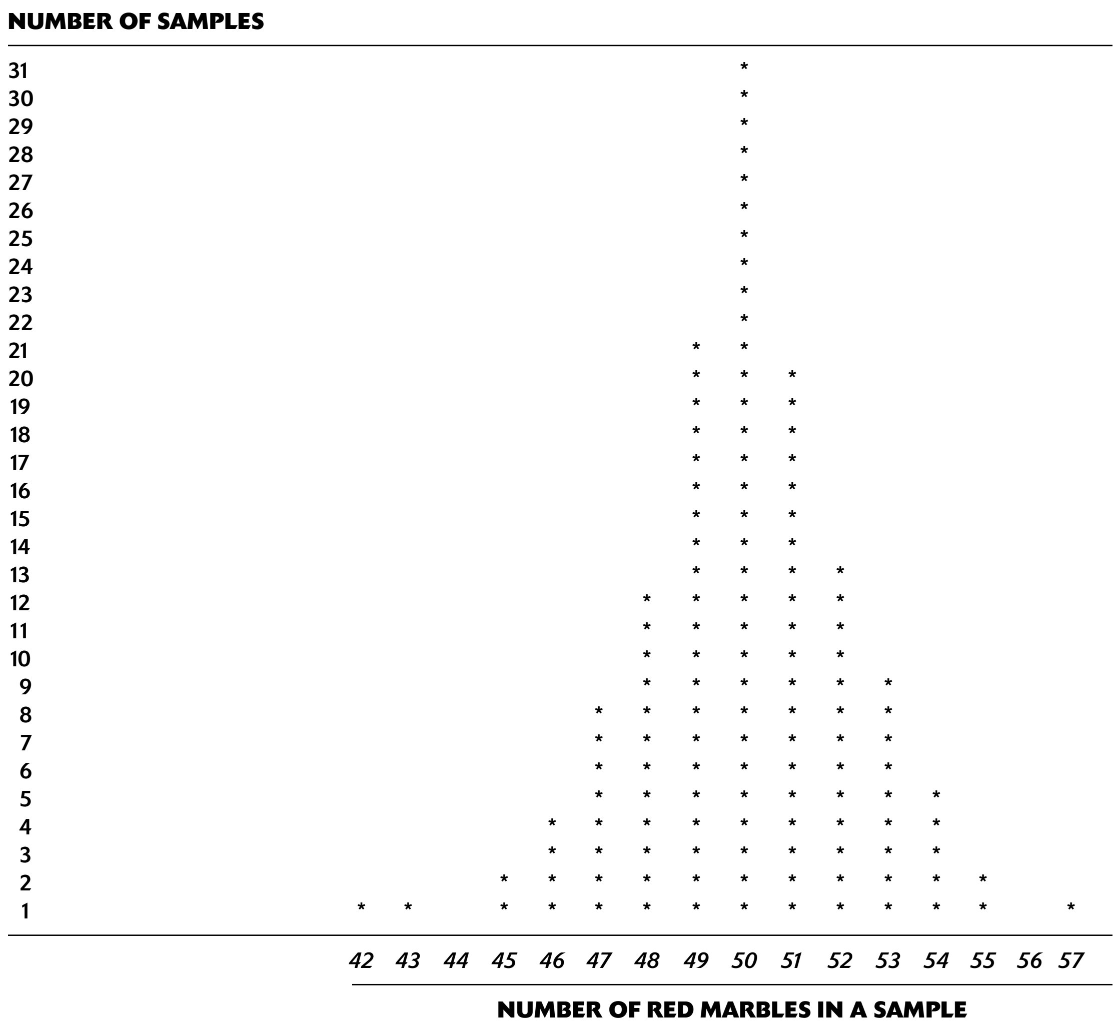
The sampling distribution pattern tells us that over many separate samples, the true population parameter (i.e., the 50/50 split in the preceding example) is more common than any other outcome. Some samples may deviate from the population parameter, but they are less common. When we plot many random samples as in the graph in Chart 1, the sampling distribution always looks like a normal or bell-shaped curve. Such a curve is theoretically important and is used throughout statistics. The area under a bell-shaped curve is well known or, in this example, we can quickly figure out the odds that we will get a specific number of marbles. If the true population parameter is 50/50, standard statistical charts tell what the odds of getting 50/50 or a 40/50 or any other split in a random sample are.

The **central limit theorem** from mathematics tells us that as the number of different random samples in a sampling distribution increases toward infinity, the pattern of samples and of the population parameter becomes increasingly predictable. For a huge number of random samples, the sampling distribution always forms a normal curve, and the midpoint of the curve will be the population parameter.

CHART 1 Example of Sampling Distribution

RED	WHITE	NUMBER OF SAMPLES
42	58	1
43	57	1
45	55	2
46	54	4
47	53	8
48	52	12
49	51	21
50	50	31
51	49	20
52	48	13
53	47	9
54	46	5
55	45	2
57	43	1
	Total	130

Number of red and white marbles that were randomly drawn from a jar of 5,000 marbles with 100 drawn each time, repeated 130 times for 130 independent random samples.



You probably do not have the time or energy to draw many different samples and just want to draw one sample. You are not alone. We rarely draw many random samples except to verify the central limit theorem. We draw only one random sample, but the central limit theorem lets us generalize from one sample to the population. The theorem is about many samples, but it allows us to calculate the probability that a particular sample is off from the population parameter. We will not go into the calculations here.

The important point is that random sampling does not guarantee that every random sample perfectly represents the population. Instead, it means that most random samples will be close to the population parameter most of the time. In addition, we can calculate the precise probability that a particular sample is inaccurate. The central limit theorem lets us estimate the chance that a particular sample is unrepresentative or how much it deviates from the population parameter. It lets us estimate the size of the sampling error. We do this by using information from one sample to estimate the sampling distribution and then combine this information with knowledge of the central limit theorem and area under a normal curve. This lets us create something very important, **confidence intervals**.

The confidence interval is a simple but very powerful idea. When television or newspaper polls are reported, you may hear about what journalists call the "margin of error" being plus or minus 2 percentage points. This is a version of confidence interval, which is a range around a specific point that we use to estimate a population parameter.

Confidence intervals A range of values, usually a little higher and lower than a specific value found in a sample, within which a researcher has a specified and high degree of confidence that the population parameter lies.

Systematic sampling A random sample in which a researcher selects every *k*th (e.g., third or twelfth) case in the sampling frame using a sampling interval.

Sampling interval The inverse of the sampling ratio that is used when selecting cases in systematic sampling.

We use a range because the statistics of random processes are based on probability. They do not let us predict an exact point. They do allow us to say with a high level of confidence (e.g., 95 percent) that the true population parameter lies within a certain range (i.e., the confidence interval). The calculations for sampling errors or confidence intervals are beyond the level of the discussion here. Nonetheless, the sampling distribution is the key idea that tells us the sampling error and confidence interval. Thus, we cannot say, "This sample gives a perfect measure of the population parameter," but we can say, "We are 95 percent certain that the true population parameter is no more than 2 percent different from what was have found in the sample." (See Expansion Box 2, Confidence Intervals.)

Going back to the marble example, I cannot say, "There are precisely 2,500 red marbles in the jar based on a random sample." However, I can say, "I am 95 percent certain that the population parameter lies between 2,450 and 2,550." I combine the characteristics of my sample (e.g., its size, the variation in it) with the central limit theorem to predict specific ranges around the population parameter with a specific degree of confidence.

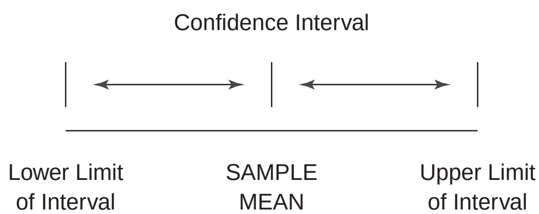
Systematic Sampling. **Systematic sampling** is a simple random sampling with a shortcut selection procedure. Everything is the same except that instead of using a list of random numbers, we first calculate a **sampling interval** to create a quasi-random selection method. The sampling interval (i.e., 1 in *k*, where *k* is some number) tells us how to select elements from a sampling frame by skipping elements in the frame before selecting one for the sample.

For instance, we want to sample 300 names from 900. After a random starting point, we select every third name of the 900 to get a sample of 300. The sampling interval is 3. Sampling intervals are easy to compute. We need the sample size and the population size (or sampling frame size as a best estimate). We can think of the sampling interval as the inverse of the sampling ratio. The sampling ratio for 300 names out of 900 is 300/900 = .333 = 33.3 percent. The sampling interval is 900/300 = 3.

EXPANSION BOX 2

Confidence Intervals

The confidence interval is a simple and very powerful idea; it has excellent mathematics behind it and some nice formulas. If you have a good mathematics background, this concept could be helpful. If you are nervous about complex mathematical formulas with many Greek symbols, here is a simple example with a simple formula (a minimum of Greek). The interval is a range that goes above and below an estimate of some characteristic of the population (i.e., population parameters), such as its average or statistical mean. The interval has an upper and lower limit. The example illustrates a simplified way to calculate a confidence interval and shows how sample size and sample homogeneity affect it.



Let us say you draw a sample of nine 12-year-old children. You weigh them and find that their average weight, the mean, is 90 pounds with a standard deviation of 36 pounds. You want to create a confidence interval around your best estimate of the population parameter (the mean weight for the population of all 12-year-olds). You symbolize the population parameter with the Greek letter μ.

Here is how to figure out a confidence interval for the population mean based on a simple random sample. You estimate a confidence level around μ by adding and subtracting a range above and below the sample mean, your best estimate of μ.

To calculate the confidence interval around the sample mean, you first calculate something called the *standard error of the mean*. Call it standard error for short. It is your estimate of variability in the sampling distribution. You use another Greek letter, σ, to symbolize the standard deviation and add the letter *m* as a subscript to it, indicating that it is your estimate of the standard deviation in the sampling distribution. Thus, the standard error comes from the standard deviation in the sampling distribution of all possible random samples from the population.

You estimate the standard deviation of the sampling distribution by getting the standard deviation of your sample and adjusting it slightly. To simplify this example, you skip the adjustment and assume that it equals the sample standard deviation. To get the standard error, you adjust it for your sample size symbolized by the letter *N*. The formula for it is:

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

Let us make the example more concrete. For the example, let us look at weight among nine 12-year-olds. For the sampling distribution of the mean you use a mean of 90 pounds and a standard deviation of 36/3 = 12 (note the square root of 9 = 3). The confidence interval has a low and upper limit. Here are formulas for them.

Lower limit $M - Z_{.95}\sigma_m$
Upper limit $M + Z_{.95}\sigma_m$

In addition to the σ_m there are two other symbols now:

M in the formula stands for mean in your sample.
$Z_{.95}$ stands for the z-score under a bell-shaped or normal curve at a 95 percent level of confidence (the most typical level). The z-score for a normal curve is a standard number (i.e., it is always the same for 95 percent level of confidence, and it happens to be 1.96). We could pick some confidence level other than 95 percent, but it is the most typical one used.

You now have everything you need to calculate upper and lower limits of the confidence interval. You compute them by adding and subtracting 1.96 standard deviations to/from the mean of 90 as follows:

Lower limit $90 - (1.96)(12) = 66.48$
Upper limit $90 + (1.96)(12) = 113.52$

(*continued*)

EXPANSION BOX 2
(continued)

This says you can be 95 percent confident that the population parameter lies somewhere between 66.48 and 113.52 pounds. You determined the upper and lower limits by adding and subtracting an amount to the sample mean (90 pounds in your example). You use 1.96 because it is the z-score when you want to be 95 percent confident. You calculated 12 as the standard error of the mean based on your sample size and the standard deviation of your sample.

You might see the wide range of 66 to 113 pounds and think it is large, and you might ask why is the sample small, with just nine children?

Here is how increasing the sample size affects the confidence interval. Let us say that instead of a sample of nine children you had 900 12-year-olds (luckily the square root of 900 is easy to figure out: 30). If everything remained the same, your σ_m with a sample of 900 is $36/30 = 1.2$. Now your confidence interval is as follows

Lower limit $90 - (1.96)(1.2) = 87.765$
Upper limit $90 + (1.96)(1.2) = 92.352$

With the much larger sample size, you can be 95 percent confident that the population parameter of average weight is somewhere between 87.765 and 92.352 pounds.

Here is how having a very homogeneous sample affects the confidence interval. Let us say that you had a standard deviation of 3.6 pounds, not 36 pounds. If everything else remained the same, your σ_m with a standard deviation of 3.6 is $3.6/9 = 0.4$

Now your confidence interval is as follows

Lower limit $90 - (1.96)(0.4) = 89.215$
Upper limit $90 + (1.96)(0.4) = 90.784$

With the very homogeneous sample, you can be 95 percent confident that the population parameter of average weight is somewhere between 89.215 and 90.784 pounds.

Let us review the confidence intervals as sample size and standard deviation change:

Sample size = 9, standard deviation = 36. Confidence interval is 66 to 113 pounds.
Sample size = 900, standard deviation = 36. Confidence interval is 87.765 to 92.352 pounds.
Sample size = 9, standard deviation = 3.6 pounds. Confidence interval is 89.215 to 90.784 pounds.

In most cases, a simple random sample and a systematic sample yield equivalent results. One important situation in which systematic sampling cannot be substituted for simple random sampling occurs when the elements in a sample are organized in some kind of cycle or pattern. For example, our sampling frame is organized as a list of married couples with the male first and the female second (see Table 1). Such a pattern gives us an unrepresentative sample if systematic sampling is used. Our systematic sample can be nonrepresentative and include only wives because of the organization of the cases. When our sample frame is organized as couples, even-numbered sampling intervals result in samples with all husbands or all wives.

Figure 3 illustrates simple random sampling and systematic sampling. Notice that different names were drawn in each sample. For example, H. Adams appears in both samples, but C. Droullard

TABLE 1 Problems with Systematic Sampling of Cyclical Data

CASE	
1	Husband
2[a]	Wife
3	Husband
4	Wife
5	Husband
6[a]	Wife
7	Husband
8	Wife
9	Husband
10[a]	Wife
11	Husband
12	Wife

Random start = 2; Sampling interval = 4.
[a]Selected into sample.

FIGURE 3 How to Draw Simple Random and Systematic Samples

1. Number each case in the sampling frame in sequence. The list of 40 names is in alphabetical order, numbered from 1 to 40.
2. Decide on a sample size. We will draw two 25 percent (10-name) samples.
3. For a *simple random sample,* locate a random-number table (see excerpt to this figure). Before using the random-number table, count the largest number of digits needed for the sample (e.g., with 40 names, two digits are needed; for 100 to 999, three digits; for 1,000 to 9,999, four digits). Begin anywhere on the random-number table (we will begin in the upper left) and take a set of digits (we will take the last two). Mark the number on the sampling frame that corresponds to the chosen random number to indicate that the case is in the sample. If the number is too large (over 40), ignore it. If the number appears more than once (10 and 21 occurred twice in the example), ignore the second occurrence. Continue until the number of cases in the sample (10 in our example) is reached.
4. For a *systematic sample,* begin with a random start. The easiest way to do this is to point blindly at the random-number table, then take the closest number that appears on the sampling frame. In the example, 18 was chosen. Start with the random number and then count the sampling interval, or 4 in our example, to come to the first number. Mark it, and then count the sampling interval for the next number. Continue to the end of the list. Continue counting the sampling interval as if the beginning of the list were attached to the end of the list (like a circle). Keep counting until ending close to the start, or on the start if the sampling interval divides evenly into the total of the sampling frame.

No.	*Name (Gender)*	*Simple Random*	*Systematic*	*No.*	*Name (Gender)*	*Simple Random*	*Systematic*
01	Abrams, J. (M)			21	Hjelmhaug, N. (M)	Yes*	
02	Adams, H. (F)	Yes	Yes (6)	22	Huang, J. (F)	Yes	Yes (1)
03	Anderson, H. (M)			23	Ivono, V. (F)		
04	Arminond, L. (M)			24	Jaquees, J. (M)		
05	Boorstein, A. (M)			25	Johnson, A. (F)		
06	Breitsprecher, P. (M)	Yes	Yes (7)	26	Kennedy, M. (F)		Yes (2)
07	Brown, D. (F)			27	Koschoreck, L. (F)		
08	Cattelino, J. (F)			28	Koykkar, J. (M)		
09	Cidoni, S. (M)			29	Kozlowski, C. (F)	Yes	
10	Davis, L. (F)	Yes*	Yes (8)	30	Laurent, J. (M)		Yes (3)
11	Droullard, C. (M)	Yes		31	Lee, R. (F)		
12	Durette, R. (F)			32	Ling, C. (M)		
13	Elsnau, K. (F)	Yes		33	McKinnon, K. (F)		
14	Falconer, T. (M)		Yes (9)	34	Min, H. (F)	Yes	Yes (4)
15	Fuerstenberg, J. (M)			35	Moini, A. (F)		
16	Fulton, P. (F)			36	Navarre, H. (M)		
17	Gnewuch, S. (F)			37	O'Sullivan, C. (M)		
18	Green, C. (M)		START, Yes (10)	38	Oh, J. (M)		Yes (5)
19	Goodwanda, T. (F)	Yes		39	Olson, J. (M)		
20	Harris, B. (M)			40	Ortiz y Garcia, L. (F)		

Excerpt from a Random-Number Table (for Simple Random Sample)

15010	18590	00102	42210	94174	22099
90122	38221	21529	00013	04734	60457
67256	13887	94119	11077	01061	27779
13761	23390	12947	21280	44506	36457
81994	66611	16597	44457	07621	51949
79180	25992	46178	23992	62108	43232
07984	47169	88094	82752	15318	11921

*Numbers that appeared twice in random numbers selected.

is in only the simple random sample. This is because it is rare for any two random samples to be identical.

The sampling frame contains twenty males and twenty females (gender is in parentheses after each name). The simple random sample yielded three males and seven females, and the systematic sample yielded five males and five females. Does this mean that systematic sampling is more accurate? No. To check this, we draw a new sample using different random numbers, taking the first two digits and beginning at the end (e.g., 11 from 11921 and then 43 from 43232). Also, we draw a new systematic sample with a different random start. The last time the random start was 18, but we now try a random start of 11. What did we find? How many of each gender?[5]

Stratified Sampling. When we use **stratified sampling**, we first divide the population into subpopulations (strata) on the basis of supplementary information.[6] After dividing the population into strata, we draw a random sample from each subpopulation. In stratified sampling, we control the relative size of each stratum rather than letting random processes control it. This guarantees representativeness or fixes the proportion of different strata within a sample. Of course, the necessary information about strata is not always available.

In general, if the stratum information is accurate, stratified sampling produces samples that are more representative of the population than those of simple random sampling. A simple example illustrates why this is so. Imagine a population that is 51 percent female and 49 percent male; the population parameter is a gender ratio of 51 to 49. With stratified sampling, we draw random samples among females and among males so that the sample contains a 51 to 49 percent gender ratio. If we had used simple random sampling, it would be possible for a random sample to be off from the true gender ratio in the population. Thus, we have fewer errors representing the population and a smaller sampling error with stratified sampling.

We use stratified sampling when a stratum of interest is a small percentage of a population and random processes could miss the stratum by chance. For example, we draw a sample of 200 from 20,000 college students using information from the college registrar's office. It indicates that 2 percent of the 20,000 students, or 400, are divorced women with children under the age of 5. For our study, this group is important to include in the sample. There would be four such students (2 percent of 200) in a representative sample, but we could miss them by chance in one simple random sample. With stratified sampling, we obtain a list of the 400 such students from the registrar and randomly select four from it. This guarantees that the sample represents the population with regard to the important strata (see Example Box 4, Illustration of Stratified Sampling).

In special situations, we may want the proportion of a stratum in a sample to differ from its true proportion in the population. For example, the population contains 0.5 percent Aleuts, but we want to examine Aleuts in particular. We oversample so that Aleuts make up 10 percent of the sample. With this type of disproportionate stratified sample, we cannot generalize directly from the sample to the population without special adjustments.

In some situations, we want the proportion of a stratum or subgroup to differ from its true proportion in the population. For example, Davis and Smith (1992) reported that the 1987 General Social Survey oversampled African Americans. A random sample of the U.S. population yielded 191 Blacks. Davis and Smith conducted a separate sample of African Americans to increase it to 544. The 191 Black respondents are about 13 percent of the random sample, roughly equal to the percentage of Blacks in the U.S. population. The 544 Blacks are 30 percent of the disproportionate sample. The researcher who wants to use the entire sample must adjust it to reduce the number of sampled African Americans before generalizing to the U.S. population. Disproportionate sampling helps the researcher who wants to focus on issues

Stratified sampling A random sample in which the researcher first identifies a set of mutually exclusive and exhaustive categories, divides the sampling frame by the categories, and then uses random selection to select cases from each category.

EXAMPLE BOX 4

Illustration of Stratified Sampling

Sample of 100 Staff of General Hospital, Stratified by Position

POSITION	POPULATION *N*	POPULATION *Percent*	SIMPLE RANDOM SAMPLE *n*	STRATIFIED SAMPLE *n*	ERRORS COMPARED TO THE POPULATION
Administrators	15	2.88	1	3	−2
Staff physicians	25	4.81	2	5	−3
Intern physicians	25	4.81	6	5	+1
Registered nurses	100	19.23	22	19	+3
Nurse assistants	100	19.23	21	19	+2
Medical technicians	75	14.42	9	14	+5
Orderlies	50	9.62	8	10	−2
Clerks	75	14.42	5	14	+1
Maintenance staff	30	5.77	3	6	−3
Cleaning staff	25	4.81	3	5	−2
Total	520	100.00	100	100	

Randomly select 3 of 15 administrators, 5 of 25 staff physicians, and so on.

Note: Traditionally, *N* symbolizes the number in the population and *n* represents the number in the sample.

The simple random sample overrepresents nurses, nursing assistants, and medical technicians but underrepresents administrators, staff physicians, maintenance staff, and cleaning staff. The stratified sample gives an accurate representation of each position.

most relevant to a subpopulation. In this case, he or she can more accurately generalize to African Americans using the 544 respondents instead of a sample of only 191. The larger sample is more likely to reflect the full diversity of the African American subpopulation.

Cluster Sampling. We use **cluster sampling** to address two problems: the lack of a good sampling frame for a dispersed population and the high cost to reach a sampled element.[7] For example, there is no single list of all automobile mechanics in North America. Even if we had an accurate sampling frame, it would cost too much to reach the sampled mechanics who are geographically spread out. Instead of using a single sampling frame, we use a sampling design that involves multiple stages and clusters.

A *cluster* is a unit that contains final sampling elements but can be treated temporarily as a sampling element itself. First we sample clusters, and then we draw a second sample from within the clusters selected in the first stage of sampling. We randomly sample clusters and then randomly sample elements from within the selected clusters. This has a significant practical advantage when we can create a good sampling frame of clusters even if it is impossible to create one for sampling elements. Once we have a sample of clusters, creating a sampling frame for elements within each cluster becomes manageable. A second advantage for geographically dispersed populations is that elements within each cluster are physically closer to one another, which can produce a savings in locating or reaching each element.

Cluster sampling A type of random sample that uses multiple stages and is often used to cover wide geographic areas in which aggregated units are randomly selected and then samples are drawn from the sampled aggregated units or clusters.

We draw several samples in stages in cluster sampling. In a three-stage sample, stage 1 is a random sampling of large clusters; stage 2 is a random sampling of small clusters within each selected large cluster; and the last stage is a sampling of elements from within the sampled small clusters. For example, we want a sample of individuals from Mapleville. First, we randomly sample city blocks, then households within blocks, and then individuals within households (see Chart 2). Although there is no accurate list of all residents of Mapleville, there is an accurate list of blocks in the city. After selecting a random sample of blocks, we count all households on the selected blocks to create a sample frame for each block. Then we use the list of households to draw a random sample at the stage of sampling households. Finally, we choose a specific individual within each sampled household.

Cluster sampling is usually less expensive than simple random sampling, but it is less accurate. Each stage in cluster sampling introduces sampling errors, so a multistage cluster sample has more sampling errors than a one-stage random sample.[8]

When we use cluster sampling, we must decide the number of clusters and the number of elements within clusters. For example, in a two-stage cluster sample of 240 people from Mapleville, we could randomly select 120 clusters and select 2 elements from each or randomly select two clusters and select 120 elements in each. Which is better? A design with more clusters is better because elements within clusters (e.g., people living on the same block) tend to be similar to each other (e.g., people on the same block tend to be more alike than those on different blocks). If few clusters are chosen, many similar elements could be selected, which would be less representative of the total population. For example, we could select two blocks with relatively wealthy people and draw 120 people from each block. This would be less representative than a sample with 120 different city blocks and 2 individuals chosen from each.

When we sample from a large geographical area and must travel to each element, cluster sampling significantly reduces travel costs. As usual, there is a trade-off between accuracy and cost. For example, Alan, Ricardo, and Barbara each personally interview a sample of 1,500 students who represent the population of all college students in North America. Alan obtains an accurate sampling frame of all students and uses simple random sampling. He travels to 1,000 different locations to interview one or two students at each. Ricardo draws a random sample of three colleges from a list of all 3,000 colleges and then visits the three and selects 500 students from each. Barbara draws a random sample of 300 colleges. She visits the 300 and selects 5 students at each. If travel costs average $250 per location, Alan's travel bill is $250,000, Ricardo's is $750, and Barbara's is $75,000. Alan's sample is highly accurate, but Barbara's is only slightly less accurate for one-third the cost. Ricardo's sample is the cheapest, but it is not representative.

Within-Household Sampling. Once we sample a household or similar unit (e.g., family or dwelling unit) in cluster sampling, the question arises as to whom we should choose. A potential source of bias is introduced if the first person who answers the telephone, the door, or the mail is used in the sample. The first person who answers should be selected only if his or her answering is the result of a truly random process. This is rarely the case. Certain people are unlikely to be at home, and in some households one person (e.g., a husband) is more likely than another to answer the telephone or door. Researchers use within-household sampling to ensure that after a random household is chosen, the individual within the household is also selected randomly.

We can randomly select a person within a household in several ways.[9] The most common method is to use a selection table specifying whom you should pick (e.g., oldest male, youngest female) after determining the size and composition of the household (see Table 2). This removes any bias that might arise from choosing the first person to answer the door or telephone or from the interviewer's selection of the person who appears to be friendliest.

Probability Proportionate to Size (PPS). There are two ways we can draw cluster samples. The method just described is proportionate or unweighted

CHART 2 Illustration of Cluster Sampling

Goal: Draw a random sample of 240 people in Mapleville.

Step 1: Mapleville has 55 districts. Randomly select 6 districts.

1 2 3* 4 5 6 7 8 9 10 11 12 13 14 15* 16 17 18 19 20 21 22 23 24 25 26
27* 28 29 30 31* 32 33 34 35 36 37 38 39 40* 41 42 43 44 45 46 47 48
49 50 51 52 53 54* 55

* = Randomly selected.

Step 2: Divide the selected districts into blocks. Each district contains 20 blocks. Randomly select 4 blocks from the district.

Example of District 3 (selected in step 1):

1 2 3 4* 5 6 7 8 9 10* 11 12 13* 14 15 16 17* 18 19 20

* = Randomly selected.

Step 3: Divide blocks into households. Randomly select households.

Example of Block 4 of District 3 (selected in step 2):

Block 4 contains a mix of single-family homes, duplexes, and four-unit apartment buildings. It is bounded by Oak Street, River Road, South Avenue, and Greenview Drive. There are 45 households on the block. Randomly select 10 households from the 45.

1	#1 Oak Street	16	"	31*	"
2	#3 Oak Street	17*	#154 River Road	32*	"
3*	#5 Oak Street	18	#156 River Road	33	"
4	"	19*	#158 River Road	34	#156 Greenview Drive
5	"	20*	"	35*	"
6	"	21	#13 South Avenue	36	"
7	#7 Oak Street	22	"	37	"
8	"	23	#11 South Avenue	38	"
9*	#150 River Road	24	#9 South Avenue	39	#158 Greenview Drive
10*	"	25	#7 South Avenue	40	"
11	"	26	#5 South Avenue	41	"
12	"	27	#3 South Avenue	42	"
13	#152 River Road	28	#1 South Avenue	43	#160 Greenview Drive
14	"	29*	"	44	"
15	"	30	#152 Greenview Drive	45	"

* = Randomly selected.

Step 4: Select a respondent within each household.

Summary of cluster sampling:

1 person randomly selected per household
10 households randomly selected per block
4 blocks randomly selected per district
6 districts randomly selected in the city
1 x 10 x 4 x 6 = 240 people in sample

TABLE 2 Within-Household Sampling

Selecting individuals within sampled households. Number selected is the household chosen in Chart 2.

NUMBER	*LAST NAME*	*ADULTS (OVER AGE 18)*	*SELECTED RESPONDENT*
3	Able	1 male, 1 female	Female
9	Bharadwaj	2 females	Youngest female
10	DiPiazza	1 male, 2 females	Oldest female
17	Wucivic	2 males, 1 female	Youngest male
19	Cseri	2 females	Youngest female
20	Taylor	1 male, 3 females	Second oldest female
29	Velu	2 males, 2 females	Oldest male
31	Wong	1 male, 1 female	Female
32	Gray	1 male	Male
35	Mall-Krinke	1 male, 2 females	Oldest female

EXAMPLE SELECTION TABLE (ONLY ADULTS COUNTED)

MALES	*FEMALES*	*WHOM TO SELECT*	*MALES*	*FEMALES*	*WHOM TO SELECT*
1	0	Male	2	2	Oldest male
2	0	Oldest male	2	3	Youngest female
3	0	Youngest male	3	2	Second oldest male
4+	0	Second oldest male	3	3	Second oldest female
0	1	Female	3	4	Third oldest female
0	2	Youngest female	4	3	Second oldest male
0	3	Second oldest female	4	4	Third oldest male
0	4+	Oldest female	4	5+	Youngest female
1	1	Female	5+	4	Second oldest male
1	2	Oldest female	5+	5+	Fourth oldest female
1	3	Second oldest female			
2	1	Youngest male			
3	1	Second oldest male			

+ = or more

cluster sampling. It is proportionate because the size of each cluster (or number of elements at each stage) is the same. The more common situation is for the cluster groups to be of different sizes. When this is the case, we must adjust the probability for each stage in sampling.

The foregoing example with Alan, Barbara, and Ricardo illustrates the problem with unweighted cluster sampling. Barbara drew a simple random sample of 300 colleges from a list of all 3,000 colleges, but she made a mistake—unless every college has an identical number of students. Her method gave each college an equal chance of being selected—a 300/3,000, or 10 percent chance. But colleges have different numbers of students, so each student does not have an equal chance to end up in her sample.

Barbara listed every college and sampled from the list. A large university with 40,000 students and a small college with 400 students had an equal chance of being selected. But if she chose the large university, the chance of a given student at that college being selected was 5 in 40,000 (5/40,000 = 0.0125 percent), whereas a student at the small college had a 5 in 400 (5/400 = 1.25 percent) chance of being selected. The small-college student was 100 times more likely to be in her sample. The total probability of a student from the large university being selected was 0.125 percent (10×0.0125) while it was 12.5 percent (10×1.25) for the small-college student. Barbara violated a principle of random sampling: that each element has an equal chance to be selected into the sample.

If Barbara uses **probability proportionate to size (PPS)** and samples correctly, then each final sampling element or student will have an equal probability of being selected. She does this by adjusting the chances of selecting a college in the first stage of sampling. She must give large colleges with more students a greater chance of being selected and small colleges a smaller chance. She adjusts the probability of selecting a college on the basis of the proportion of all students in the population who attend it. Thus, a college with 40,000 students will be 100 times more likely to be selected than one with 400 students. (See Example Box 5, Probability Proportionate to Size (PPS) Sampling.)

Random-Digit Dialing. **Random-digit dialing (RDD)** is a sampling technique used in research projects in which the general public is interviewed by telephone.[10] It does not use the published telephone directory as the sampling frame. Using a telephone directory as the sampling frame misses three kinds of people: those without telephones, those who have recently moved, and those with unlisted numbers. Those without phones (e.g., the poor, the uneducated, and transients) are missed in any telephone interview study, but 95 percent of people in advanced industrialized nations have a telephone. Several types of people have unlisted numbers: those who want to avoid collection agencies; those who are very wealthy; and those who want to have privacy and to avoid obscene calls, salespeople, and prank calls. In some urban areas in the United States, the percentage of unlisted numbers is 50 percent. In addition, people change their residences, so annual directories have numbers for people who have moved away and do not list those who have recently moved into an area.

If we use RDD, we randomly select telephone numbers, thereby avoiding the problems of telephone directories. The population is telephone numbers, not people with telephones. RDD is not difficult, but it takes time and can frustrate the person doing the calling.

Here is how RDD works in the United States. Telephone numbers have three parts: a three-digit area code, a three-digit exchange number or central office code, and a four-digit number. For example, the area code for Madison, Wisconsin, is 608, and there are many exchanges within the area code (e.g., 221, 993, 767, 455), but not all of the 999 possible three-digit exchanges (from 001 to 999) are active. Likewise, not all of the 9,999 possible four-digit numbers in an exchange (from 0000 to 9999) are being used. Some numbers are reserved for future expansion, are disconnected, or are temporarily withdrawn after someone moves. Thus, a possible U.S. telephone number consists of an active area code, an active exchange number, and a four-digit number in an exchange.

In RDD, a researcher identifies active area codes and exchanges and then randomly selects four-digit numbers. A problem is that the researcher can select any number in an exchange. This means that some selected numbers are out of service, disconnected, pay phones, or numbers for businesses; only some numbers are what the researcher wants: working residential phone numbers. Until the researcher calls, it is not possible to know whether the number is a working residential number. This means spending much time reaching numbers that are disconnected, are for businesses, and so forth. Research organizations often use

Probability proportionate to size (PPS) An adjustment made in cluster sampling when each cluster does not have the same number of sampling elements.

Random-digit dialing (RDD) A method of randomly selecting cases for telephone interviews that uses all possible telephone numbers as a sampling frame.

EXAMPLE BOX 5

Probability Proportionate to Size (PPS) Sampling

Henry wants to conduct one-hour, in-person interviews with people living in the city of Riverdale, which is spread out over a large area. Henry wants to reduce his travel time and expenses, so he uses a *cluster sampling design*. The last census reported that the city had about 490,000 people. Henry can interview only about 220 people, or about 0.05 percent of the city population. He first gathers maps from the city tax office and fire department, and retrieves census information on city blocks. He learns that there are 2,182 city blocks. At first, he thinks he can randomly select 10 percent of the blocks (i.e., 218), go to a block and count housing units, and then locate one person to interview in each housing unit (house, apartment, etc.), but the blocks are of unequal geographic and population size. He studies the population density of the blocks and estimates the number of people in each, and then develops a five-part classification based on the average size of a block as in the following chart.

Block Type	*Number of Clusters*	*Average Number People per Block*
Very high density	20	2,000
High density	200	800
Medium density	800	300
Low density	1,000	50
Semirural	162	10

Henry realizes that randomly selecting city blocks without adjustment will not give each person an equal chance of being selected. For example, 1 very high-density block has the same number of people as 40 low-density blocks. Henry adjusts proportionately to the block size. The easiest way to do this is to convert all city blocks to equal-size units based on the smallest cluster, or the semirural city blocks. For example, there are 2,000/10 or 200 times more people in a high-density block than a semirural block, so Henry increases the odds of selecting such a block to make its probability 200 times higher than a semirural block. Essentially, Henry creates adjusted cluster units of 10 persons each (because that is how many there are in the semirural blocks) and substitutes them for city blocks in the first stage of sampling. The 162 semirural blocks are unchanged, but after adjustment, he has 20 X 200 = 4,000 units for the very high density blocks, 200 X 80 = 16,000 units for the high-density blocks, and so forth, for a total of 49,162 such units. Henry now numbers each block, using the adjusted cluster units, with many blocks getting multiple numbers. For example, he assigns numbers 1 to 200 to the first very high density block, and so forth, as follows:

1 Very high density block #1
2 Very high density block #1
3 Very high density block #1

. . . and so forth

3,999 Very high density block #20
4,000 Very high density block #20
4,001 High-density block #1
4,002 High-density block #2

. . . and so forth

49,160 Semirural block #160
49,161 Semirural block #161
49,162 Semirural block #162

Henry still wants to interview about 220 people and wants to select one person from each adjusted cluster unit. He uses simple random sample methods to select 220 of the 49,162 adjusted cluster units. He can then convert the cluster units back to city blocks. For example, if Henry randomly selected numbers 25 and 184, both are in very high density block #1, telling him to select two people from that block. If he randomly picked the number 49,161, he selects one person in semirural block #161. Henry now goes to each selected block, identifies all housing units in that block, and randomly selects among housing units. Of course, Henry may use within-household sampling after he selects a housing unit.

computers to select random digits and dial the phone automatically. This speeds the process, but a human must still listen and find out whether the number is a working residential one (see Expansion Box 3, Random Digit Dialing.)

The sampling element in RDD is the phone number, not the person or the household. Several families or individuals can share the same phone number, and in other situations, each person may have a separate phone number. This means that after a working residential phone is reached, a second stage of sampling, within household sampling, is necessary to select the person to be interviewed.

Example Box 6, (Example Sample, the 2006 General Social Survey) illustrates how the many sampling terms and ideas can be used together in a specific real-life situation.

EXPANSION BOX 3

Random-Digit Dialing (RDD)

During the past decade, participation in RDD surveys has declined. This is due to factors such as new call-screening technologies, heightened privacy concerns due to increased telemarketing calls, a proliferation of nonhousehold telephone numbers, and increased cell telephone users (most RDD samples include only landline numbers). When they compared a new technique, address-based sampling (ABS), to RDD for the U.S. adult population, Link et al. (2008) estimated that RDD sampling frames may be missing 15–19 percent of the population. Although the alternative was superior to RDD in some respects, ABS had other limitations including overrepresentation of English-speaking non-Hispanics and more educated persons than RDD. One issue in RDD sampling involves reaching someone by phone. A researcher might call a phone number dozens of times that is never answered. Does the nonanswer mean an eligible person is not answering or that the number is not really connected with a person? A study (Kennedy, Keeter, and Dimock, 2008) of this issue estimates that about half (47 percent) of unanswered calls in which there are six call-back attempts have an eligible person who is not being reached.

Decision Regarding Sample Size

New social researchers often ask, "How large does my sample have to be?" The best answer is, "It depends." It depends on population characteristics, the type of data analysis to be employed, and the degree of confidence in sample accuracy needed for research purposes. As noted, a large sample size alone does not guarantee a representative sample. A large sample without random sampling or with a poor sampling frame creates a less representative sample than a smaller one that has careful random sampling and an excellent sampling frame.

We can address the question of sample size in two ways. One method is to make assumptions about the population and use statistical equations about random sampling processes. The calculation of sample size by this method requires a statistical discussion that goes beyond the level of this text.[11] We must make assumptions about the degree of confidence (or number of errors) that is acceptable and the degree of variation in the population. In general, the more diverse a population, the more precise is the statistical analysis, the more variables will be examined simultaneously, and the greater confidence is required in sample accuracy (e.g., it makes a difference in critical health outcomes, huge financial loss, or the freedom or incarceration of innocent people), the larger the required sample size. The flip side is that samples from homogeneous populations with simple data analysis of one or a few variables that are used for low-risk decisions can be equally effective when they are smaller.

A second method to decide a sample size is a rule of thumb, a conventional or commonly accepted amount. We use rules of thumb because we rarely have the information required by the statistical estimation method. Also, these rules give sample sizes close to those of the statistical method. Rules of thumb are based on past experience with samples that have met the requirements of the statistical method.

A major principle of sample size is that the smaller the population, the larger the sampling ratio has to be for a sample that has a high probability of yielding the same results as the entire population. Larger populations permit smaller sampling ratios for equally good samples because as the population

EXAMPLE BOX 6

Example Sample, the 2006 General Social Survey

Sampling has many terms for the different types of samples. A complex sample illustrates how researchers use them. We can look at the 2006 sample for the best-known national U.S. survey in sociology, the General Social Survey (GSS). It has been conducted since 1972. Its sampling has been updated several times over the years based on the most sophisticated social science sampling techniques to produce a representative population within practical cost limits. The *population* consists of all resident adults (18 years of age or older) in the United States for the *universe* of all Americans. The *target population* consists of all English- or Spanish-speaking mentally competent adults who live in households but excludes people living in institutional settings. The researchers used a complex multistage area probability sample to the block or segment level. At the block level, they used *quota sampling* with quotas based on gender, age, and employment status. They selected equal numbers of men and women as well as persons over and under 35 years of age.

The sample design combined a *cluster sample* and a *stratified sample*. U.S. territory was divided into standard metropolitan statistical areas (SMSAs, a U.S. Census Bureau classification) and nonmetropolitan counties. The SMSAs and counties were stratified by region, age, and race before selection. Researchers adjusted clusters using *probability proportionate to size (PPS)* based on the number of housing units in each county or SMSA.

The sampling design had three basic stages. *Stage 1*: Randomly select a "primary sampling unit" (a U.S. census tract, a part of a SMSA, or a county) from among the stratified "primary sampling units." Researchers also classified units by whether there were stable mailing addresses in a geographic area or others. *Stage 2*: Randomly select smaller geographic units (e.g., a census tract, parts of a county), and *Stage 3*: Randomly select housing units on blocks or similar geographic units. As a final stage, researchers used the household as the sampling element and randomly selected households from the addresses in the block. After selecting an address, an interviewer contacted the household and chose an eligible respondent from it. The interviewer looked at a quota selection table for possible respondents and interviewed a type of respondent (e.g., second oldest) based on the table. Interviewers used computer-assisted personal interviewing (CAPI).

In the 2006 sample, researchers first identified 9,535 possible household addresses or locations. However, this number dropped to 7,987 after they eliminated vacant addresses and ones where no one who spoke either English or Spanish lived. After taking into account people who refused to participate, were too ill, were ineligible, or did not finish an interview (23.3%), the final sample included 4,510 persons (for details, see http://publicdata.norc.org:41000/gss/Documents/Codebook/A.pdf)

size grows, the returns in accuracy for sample size decrease.

In practical terms, this means for small populations (under 500), we need a large sampling ratio (about 30 percent) or 150 people, while for large populations (over 150,000), we can obtain equally good accuracy with a smaller sampling ratio (1 percent), and samples of about 1,500 can be equally accurate, all things being the same. Notice that the population of 150,000 is 30 times larger but the sample is just 10 times larger. Turning to very large populations (more than 10 million), we can achieve accuracy with tiny sampling ratios (0.025 percent), or samples of about 2,500. The size of the population ceases to be relevant once the sampling ratio is very small, and samples of about 2,500 are as accurate for populations of 200 million as for 10 million. These are approximate sizes, and practical limitations (e.g., cost) also play a role.

A related principle is that for small samples, a small increase in sample size produces a big gain in accuracy. Equal increases in sample size produce an increase in accuracy more for small than for large samples. For example, an increase in sample size from 50 to 100 reduces errors from 7.1 percent to 2.1 percent, but an increase from 1,000 to 2,000

TABLE 3 Sample Size of a Random Sample for Different Populations with a 99 Percent Confidence Level

POPULATION SIZE	SAMPLE SIZE	% POPULATION IN SAMPLE
200	171	85.5%
500	352	70.4%
1,000	543	54.3%
2,000	745	37.2%
5,000	960	19.2%
10,000	1,061	10.6%
20,000	1,121	5.6%
50,000	1,160	2.3%
100,000	1,173	1.2%

decreases errors from only 1.6 percent to 1.1 percent.[12] (See Table 3.)

Notice that our plans for data analysis influence the required sample size. If we want to analyze many small subgroups within the population, we need a larger sample. Let us say we want to see how elderly Black females living in cities compare with other subgroups (elderly males, females of other ages and races, and so forth). We will need a large sample because the subgroup is a small proportion (e.g., 10 percent) of the entire sample. A rule of thumb is to have about 50 cases for each subgroup we wish to analyze. If we want to analyze a group that is only 10 percent of our sample, then we will need a sample 10 times 50 (500 cases) in the sample for the subgroup analysis. You may ask how you would know that the subgroup of interest is only 10 percent of the sample until you gather sample data? This is a legitimate question. We often must use various other sources of information (e.g., past studies, official statistics about people in an area), then make an estimate, and then plan our sample size requirements from the estimate.

Making Inferences. The reason we draw probability samples is to make inferences from the sample to the population. In fact, a subfield of statistical data analysis is called *inferential statistics*. We directly observe data in the sample but are not interested in a sample alone. If we had a sample of 300 from 10,000 students on a college campus, we are less interested in the 300 students than in using information from them to infer to the population of 10,000 students. Thus, a gap exists between what we concretely have (variables measured in sample data) and what is of real interest (population parameters) (see Figure 4).

We can express the logic of measurement in terms of a gap between abstract constructs and concrete indicators. Measures of concrete, observable data are approximations for abstract constructs. We use the approximations to estimate what is of real interest (i.e., constructs and causal laws). Conceptualization and operationalization bridge the gap in measurement just as the use of sampling frames, the sampling process, and inference bridge the gap in sampling.

We can integrate the logic of sampling with the logic of measurement by directly observing measures of constructs and empirical relationships in samples (see Figure 4). We infer or generalize from what we observe empirically in samples to the abstract causal laws and parameters in the population. Likewise, there is an analogy between the logic of sampling and the logic of measurement for validity. In measurement, we want valid indicators of constructs: that is, concrete observable indicators that accurately represent unseen abstract constructs. In sampling, we want samples that have little sampling error: that is, concrete collections of cases that accurately represent unseen and abstract populations. A valid measure deviates little from the construct it represents. A good sample has little sampling error, and it permits estimates that deviate little from population parameters.

We want to reduce sampling errors. For equally good sampling frames and precise random selection processes, the sampling error is based on two factors: the sample size and the population diversity. Everything else being equal, the larger the sample size, the smaller the sampling error. Likewise, populations with a great deal of homogeneity will have smaller sampling errors. We can think of it this way: if we had a choice between

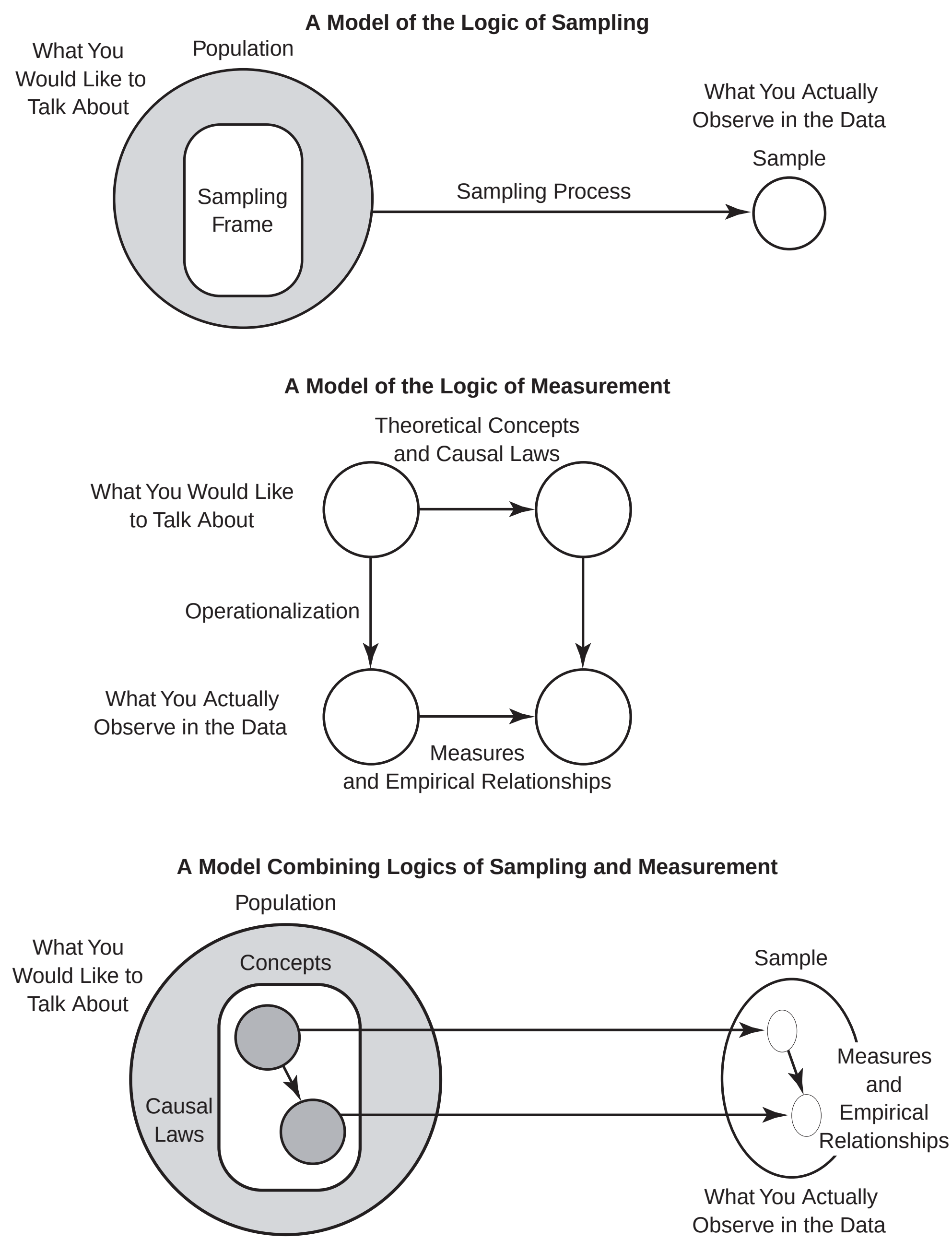


FIGURE 4 **Model of the Logic of Sampling and of Measurement**

sampling/picking 10 or 50 marbles out of a jar of 1000 red and white marbles to determine the number of red marbles, it would be better to pick 50. Likewise, if there are ten colors of marbles in a jar, we are less able to predict accurately the number of red marbles than if there were only two colors of marbles.

Sampling error is related to confidence intervals. If two samples are identical except one is much larger, the larger one will have a smaller sampling error and narrower confidence intervals. Likewise, if two samples are identical except that the cases in one are more similar to each other, the one with greater homogeneity will have a smaller sampling error and narrower confidence intervals. A narrow confidence interval means that we are able to estimate more precisely the population parameter for a given level of confidence.

Here is an example: You want to estimate the annual income of bricklayers. You have two samples. Sample 1 gives a confidence interval of $30,000 to $36,000 around the estimated population parameter of $33,000 for an 80 percent level of confidence. However, you want a 95 percent level of confidence. Now the range is $25,000 to $45,000. A sample that has a smaller sampling error (because it is much larger) might give the $30,000 to $36,000 range for a 95 percent confidence level.

Strategies When the Goal Differs from Creating a Representative Sample

In qualitative research, the purpose of research may not require having a representative sample from a huge number of cases. Instead, a nonprobability sample often better fits the purposes of a study. In nonprobability samples, you do not have to determine the sample size in advance and have limited knowledge about the larger group or population from which the sample is taken. Unlike a probability sample that required a preplanned approach based on mathematical theory, nonprobability sampling often gradually selects cases with the specific content of a case determining whether it is chosen. Table 4 shows a variety of nonprobability sampling techniques.

TABLE 4 Types of Nonprobability Samples

TYPE OF SAMPLE	PRINCIPLE
Convenience	Get any cases in any manner that is convenient.
Quota	Get a preset number of cases in each of several predetermined categories that will reflect the diversity of the population, using haphazard methods.
Purposive	Get all possible cases that fit particular criteria, using various methods.
Snowball	Get cases using referrals from one or a few cases, then referrals from those cases, and so forth.
Deviant case	Get cases that substantially differ from the dominant pattern (a special type of purposive sample).
Sequential	Get cases until there is no additional information or new characteristics (often used with other sampling methods).
Theoretical	Get cases that will help reveal features that are theoretically important about a particular setting/topic.
Adaptive	Get cases based on multiple stages, such as snowball followed by purposive. This sample is used for hidden populations.

Purposive or Judgmental Sampling

Purposive sampling (also known as *judgmental sampling*) is a valuable sampling type for special situations. It is used in exploratory research or in field research.[12] It uses the judgment of an expert in

Purposive sampling A nonrandom sample in which the researcher uses a wide range of methods to locate all possible cases of a highly specific and difficult-to-reach population.

selecting cases, or it selects cases with a specific purpose in mind. It is inappropriate if the goal is to have a representative sample or to pick the "average" or the "typical" case. In purposive sampling, cases selected rarely represent the entire population.

Purposive sampling is appropriate to select unique cases that are especially informative. For example, we want to use content analysis to study magazines to find cultural themes. We can use three specific popular women's magazines to study because they are trend setting. In the study *Promises I Can Keep* that opened this chapter, the researchers selected eight neighborhoods using purposive sampling. We often use purposive sampling to select members of a difficult-to-reach, specialized population, such as prostitutes. It is impossible to list all prostitutes and sample randomly from the list. Instead, to locate persons who are prostitutes, a researcher will use local knowledge (e.g., locations where prostitutes solicit, social groups with whom prostitutes associate) and local experts (e.g., police who work on vice units, other prostitutes) to locate possible prostitutes for inclusion in the research project. A researcher will use many different methods to identify the cases because the goal is to locate as many cases as possible.

We also use purposive sampling to identify particular types of cases for in-depth investigation to gain a deeper understanding of types (see Example Box 7, Purposive Sampling).

Snowball Sampling

We are often interested in an interconnected network of people or organizations.[13] The network could be scientists around the world investigating the same problem, the elites of a medium-size city, members of an organized crime family, persons who sit on the boards of directors of major banks and corporations, or people on a college campus who

EXAMPLE BOX 7

Purposive Sampling

In her study *Inside Organized Racism,* Kathleen Blee (2002) used purposive sampling to study women who belong to racist hate organizations. The purpose of her study was to learn why and how women became actively involved in racist hate organizations (e.g., neo-Nazi, Ku Klux Klan). She wanted "to create a broadly based, national sample of women racist group members" (p. 198). A probability sample was not possible because no list of all organizations exists, and the organizations keep membership lists secret.

Blee avoided using snowball sampling because she wanted to interview women who were not connected to one another. To sample women for the study, she began by studying the communication (videotapes, books, newsletters, magazines, flyers, Web sites) "distributed by every self-proclaimed racist, anti-Semitic, white supremacist, Christian Identity, neo-Nazi, white power skinhead, and white separatist organization in the United States for a one-year period" (p. 198). She also obtained lists from antiracist organizations that monitor racist groups and examined the archives at the libraries of Tulane University and the University of Kansas for right-wing extremism. She identified more than one hundred active organizations. From these, she found those that had women members or activists and narrowed the list to thirty racist organizations. She then tried to locate women who belonged to organizations that differed in ideological emphasis and organizational form in fifteen different states in four major regions of the United States.

In a type of cluster sampling, she first located organizations and then women active in them. To find women to interview, she used personal contacts and referrals from informed persons: "parole officers, correctional officials, newspaper reporters and journalists, other racist activists and former activists, federal and state task forces on gangs, attorneys, and other researchers" (p. 200). She eventually located thirty-four women aged 16 to 90 years of age and conducted two 6-hour life history interviews with each.

Source: Excerpt from page 198 of *Inside Organized Racism*: *Women in the Hate Movement,* by Kathleen M. Blee. © 2002 by the Regents of the University of California. Published by the University of California Press.

have had sexual relations with each other. The crucial feature is that each person or unit is connected with another through a direct or indirect linkage. This does not mean that each person directly knows, interacts with, or is influenced by every other person in the network. Rather, taken as a whole, with direct and indirect links, most people are within an interconnected web of linkages.

For example, Sally and Tim do not know each other directly, but each has a good friend, Susan, so they have an indirect connection. All three are part of the same friendship network. Researchers represent such a network by drawing a *sociogram*, a diagram of circles connected with lines. The circles represent each person or case, and the lines represent friendship or other linkages (see Figure 5).

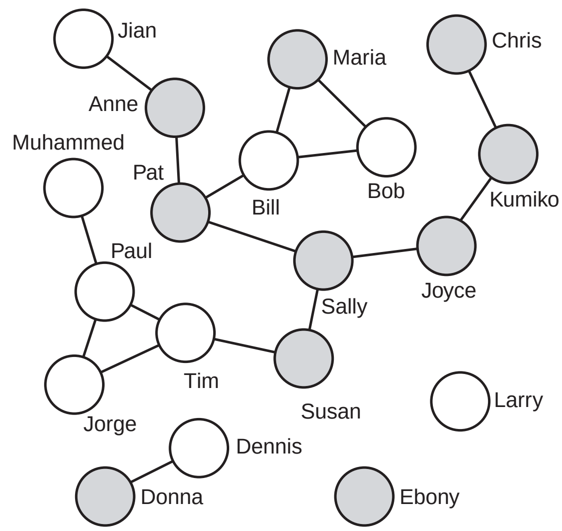
Snowball sampling (also called *network*, *chain referral*, *reputational*, and *respondent-driven sampling*) is a method for sampling (or selecting) the cases in a network. The method uses an analogy to a snowball, which begins small but becomes larger as we roll it on wet snow and it picks up additional snow. Snowball sampling is a multistage technique. It begins with one or a few people or cases and spreads out based on links to the initial cases.

For example, we want to study friendship networks among the teenagers in our community. We might start with three teenagers who do not know each other. We ask each teen to name four close friends. Next we go to each set of four friends and ask each person to name four close friends. This continues to the next round of four people and repeats again. Before long, a large number of people have been identified. Each person in the sample is directly or indirectly tied to the original teenagers, and several people may have named the same person. The process stops, either because no new names are given, indicating a closed network, or because the network is so large that it is at the limit of what can be studied. The sample includes those named by at least one other person in the network as being a close friend.

Deviant Case Sampling

We use **deviant case sampling** (also called *extreme case sampling*) when we are interested in cases that differ from the dominant pattern, mainstream, or predominant characteristics of other cases. Similar to purposive sampling, we use a variety of techniques to locate cases with specific characteristics. The goal is to locate a collection of unusual, different, or peculiar cases that are not representative of the whole. We select cases because they are unusual. We can sometimes learn more about social life by considering cases that fall outside the general pattern or including what is beyond the main flow of events.

For example, we want to study high school dropouts. Let us say that previous research suggested that a majority of dropouts come from low-income,



Note: Shading indicates various skin tones.

FIGURE 5 **Sociogram of Friendship Relations**

Snowball sampling A nonrandom sample in which the researcher begins with one case and then, based on information about interrelationships from that case, identifies other cases and repeats the process again and again.

Deviant case sampling A nonrandom sample, especially used by qualitative researchers, in which a researcher selects unusual or nonconforming cases purposely as a way to provide increased insight into social processes or a setting.

single-parent families and tend to be racial minorities. The family environment is one in which parents and/or siblings have low education or are themselves dropouts. In addition, many dropouts engage in illegal behavior. We might seek dropouts who are members of the majority racial group, who have no record of illegal activities, and who are from stable two-parent, upper-middle-income families. By looking at atypical dropouts we might learn more about the reasons for dropping out.

Sequential Sampling

Sequential sampling is also similar to purposive sampling. We use purposive sampling to try to locate as many relevant cases as possible. Sequential sampling differs because we continue to gather cases until the amount of new information ends or a certain diversity of cases is reached. The principle is to gather cases until we reach a saturation point. In economic terms, information is gathered until the marginal utility, or incremental benefit for additional cases, levels off or drops significantly. It requires that we continuously evaluate all collected cases. For example, we locate and plan in-depth interviews of sixty widows over 70 years of age who have been living without a spouse for 10 or more years. Depending on our purposes, getting an additional twenty widows whose life experiences, social backgrounds, and worldviews differ little from the first sixty may be unnecessary.

Sequential sampling A nonrandom sample in which a researcher tries to find as many relevant cases as possible until time, financial resources, or his or her energy is exhausted or until there is no new information or diversity from the cases.

Theoretical sampling A nonrandom sample in which the researcher selects specific times, locations, or events to observe in order to develop a social theory or evaluate theoretical ideas.

Hidden population A population of people who engage in clandestine, socially disapproved of, or concealed activities and who are difficult to locate and study.

Adaptive sampling A nonprobability sampling technique used for hidden populations in which several approaches to identify and recruit, including a snowball or referral method, may be used.

Theoretical Sampling

In **theoretical sampling**, what we sample (e.g., people, situations, events, time periods) comes from grounded theory. A growing theoretical interest guides the selection of sample cases. The researcher selects cases based on new insights that the sample could provide. For example, a field researcher could be observing a site and a group of people during weekdays. Theoretically, the researcher may question whether the people act the same at other times or aspects of the site change. He or she could then sample other time periods (e.g., nights and weekends) to have a fuller picture and learn whether important conditions are the same.

Adaptive Sampling and Hidden Populations

In contrast to sampling the general population or visible and accessible people, sampling **hidden populations** (i.e., people who engage in clandestine or concealed activities) is a recurrent issue in the studies of deviant or stigmatized behavior (such as victims of sexual violence, illegal drug users). This method illustrates the creative application of sampling principles, mixing qualitative and quantitative styles of research and combining probability with nonprobability techniques.

Adaptive sampling is a design that adjusts based on early observations.[15] For example, we ask illegal drug users to refer other drug users as in snowball sampling. However, we adjust the way that we trace through the network based on our research topic. We might identify a geographic area, divide it into sections randomly, and then select participants in that area through strategies such as random-digit dialing or by posting recruitment fliers. Once we identify members of the targeted hidden population, we use them in a snowball technique to find others. AIDS researchers or studies of illegal drug users that have sampled "hidden populations" are instructive, often relying on modified snowball techniques. (See Example Box 8, Hidden Populations).

EXAMPLE BOX 8

Hidden Populations

Three studies of hidden populations illustrate the difficulties of sampling. Martin and Dean (1993) sampled gay men from New York City. The men had to live in the city, be over age 18, not be diagnosed as having AIDS, and engage in sex with other men. The authors began with a purposive sample using five diverse sources to recruit 291 respondents. They first contacted 150 New York City organizations with predominately homosexual or bisexual members. They next screened these to 90 organizations that had men appropriate for the study. From the 90, the researchers drew a stratified random sample of 52 organizations by membership size. They randomly selected five members from each of the organizations. Reports of Martin and Dean's study appeared in local news sources. This brought calls from forty-one unsolicited volunteers. They also found thirty-two men as referrals from respondents who had participated in a small pilot study, seventy-two men from an annual New York City Gay Pride Parade, and fifteen eligible men whom they contacted at a New York City clinic and asked to participate. They next used snowball sampling by asking each of the 291 men to give a recruitment packet to three gay male friends. Each friend who agreed to participate was also asked to give packets to three friends. This continued until it had gone five levels out from the initial 291 men. Eventually, Dean recruited 746 men into the study. The researchers checked their sample against two random samples of gay men in San Francisco, a random-digit dialing sample of 500, and a cluster sample of 823 using San Francisco census tracts. Their sample paralleled those from San Francisco on race, age, and the percent being "out of the closet."

Heckathorn (1997, 2002) studied active drug injectors in two small Connecticut cities and the surrounding area. As of July 1996, medical personnel had diagnosed 390 AIDS cases in the towns; about half of the cases involved drug injection. The sampling was purposive in that each sampled element had to meet certain criteria. Heckathorn also used a modified snowball sampling with a "dual reward system." He gave each person who completed an interview a monetary reward and a second monetary reward for recruiting a new respondent. The first person was asked not to identify the new person to the researcher, a practice sometimes referred to as *masking* (i.e., protecting friends). This avoids the "snitching" issue and "war on drugs" stigma, especially strong in the U.S. context. This modified snowball sampling is like sequential sampling in that after a period of time, fewer and fewer new recruits are found until the researcher comes to saturation or an equilibrium.

Wang et al. (2006) used a respondent-driven sampling method to recruit 249 illicit drug users in three rural Ohio counties to examine substance abuse and health care needs. To be eligible for the sample, participants had to be over 18 years of age, not be in drug abuse treatment, and not have used cocaine or methamphetamines in the past month. After locating an eligible participant, the researchers paid him or her $50 dollars to participate. The participant could earn an additional $10 by recruiting eligible peers. In a snowball process, each subsequent participant was also asked to make referrals. The authors identified nineteen people to start. Only a little more than half (eleven of the nineteen) referred peers for the study who were eligible and participated. Over roughly 18 months, the researchers were able to identify 249 participants for their study. They compared the study sample with characteristics of estimates of the illegal drug–using population and found that the racial composition of the originally identified participants (White) led to overrepresentation of that racial category. Otherwise, it appeared that the method was able to draw a reasonable sample of the hidden population.

CONCLUSION

This chapter discussed probability and nonprobability sampling (see Summary Review Box 1, Types of Samples). A key point is that a sampling strategy should match in a specific study's purpose. In general, probability sampling is preferred for a representative sample; it allows for using statistical tests in data analysis. In addition to simple random sampling, the chapter referred to other probability samples: systematic, stratified, RDD, and cluster sampling. The

SUMMARY REVIEW BOX 1

Types of Samples

EIGHT TYPES OF NONPROBABILITY SAMPLES

Type of Sample	*Principle*
Adaptive	Get a few cases using knowledge of likely locations of a hidden population, use random techniques or recruit, and then use a snowball sample to expand from a few cases.
Convenience	Get any cases in any manner that is convenient.
Deviant case	Get cases that substantially differ from the dominant pattern (a special type of purposive sample).
Purposive	Get all possible cases that fit particular criteria using various methods.
Quota	Using haphazard methods, get a preset number of cases in each of several predetermined categories that will reflect the diversity of the population.
Sequential	Get cases until there is no additional information or new characteristics (often used with other sampling methods).
Snowball	Get cases using referrals from one or a few cases, then referrals from those cases, and so forth.
Theoretical	Get cases that will help reveal features that are theoretically important about a particular setting/topic.

FOUR TYPES OF PROBABILITY SAMPLES

Type of Sample	*Technique*
Cluster	Create a sampling frame for large cluster units, draw a random sample of the cluster units, create a sampling frame for cases within each selected cluster unit, then draw a random sample of cases, and so forth.
Simple random	Create a sampling frame for all cases and then select cases using a purely random process (e.g., random-number table or computer program).
Stratified	Create a sampling frame for each of several categories of cases, draw a random sample from each category, and then combine the several samples.
Systematic	Create a sampling frame, calculate the sampling interval 1/k, choose a random starting place, and then take every 1/k case.

discussions of sampling error, the central limit theorem, and sample size indicated that probability sampling produces most accurate sampling when the goal is creating a representative sample.

The chapter also discussed several types of nonprobability samples: convenience, deviant case quota, sequential, snowball, and theoretical. Except for convenience, these types are best suited for studies in which the purpose is other than creating a sample that is highly representative of a population.

Before you move on, it may be useful to restate a fundamental principle of all social research: Do not compartmentalize the steps of the research process; rather, learn to see the interconnections among the steps. Research design, measurement, sampling, and specific research techniques are interdependent. In practice, we need to think about data collection as we design research and develop measures. Likewise, sampling issues influence research design, measurement, and data collection strategies. As you will see, good social research depends on simultaneously controlling quality at several different steps: research design, conceptualization, measurement, sampling, and data collection and handling. Making serious errors at any one stage could make an entire research project worthless.

KEY TERMS

adaptive sampling
central limit theorem
cluster sampling
confidence intervals
convenience sampling
deviant case sampling
hidden populations
parameter
population
probability proportionate to size (PPS)
purposive sampling
quota sampling
random-digit dialing (RDD)
random-number table
random sample
sample
sampling distribution
sampling element
sampling error
sampling frame
sampling interval
sampling ratio
sequential sampling
simple random sample
snowball sampling
statistic
stratified sampling
systematic sampling
target population
theoretical sampling

REVIEW QUESTIONS

1. When is purposive sampling used?
2. When is the snowball sampling technique appropriate?
3. What is a sampling frame and why is it important?
4. Which sampling method is best when the population has several groups and a researcher wants to ensure that each group is in the sample?
5. How can researchers determine a sampling interval from a sampling ratio?
6. When should a researcher consider using probability proportionate to size?
7. What is the population in random-digit dialing? Does this type avoid sampling frame problems? Explain.
8. How do researchers decide how large a sample to use?
9. How are the logic of sampling and the logic of measurement related?
10. When is random-digit dialing used, and what are its advantages and disadvantages?