

6

Making Research Decisions: Sampling

Chapter Contents

- Sampling Basics
 - Generalizing From a Sample to a Population
 - Sampling Methods
 - Nonprobability Sampling Methods
 - Probability Sampling Methods
 - How Big Should a Sample Be?
 - Calculating the Appropriate Sample Size
 - Sample Size Formula
 - Error Calculations
 - Issues and Assumptions
 - Final Thoughts
-

Sampling is a powerful tool and a critical part of communication campaign research because it has a direct relationship to the generalizability of research results. Practitioners use sampling in partnership with the research methods they select to help them solve complex problems, monitor their internal and external environments, and engage in sophisticated campaign planning and evaluation. Sampling helps practitioners get accurate information quickly at a relatively low cost. It provides them with a cost-effective way to collect information from a relatively small number of target audience members, called a *sample*, and draw conclusions about an entire target audience. These processes are based on principles of statistical sampling and inference.

Although it sounds complex, sampling really is simple. If we want to know whether our spaghetti sauce needs more garlic, we usually taste a small sample. We do not need to eat all of the sauce to determine whether more garlic is needed (and by the way, the sauce almost always needs more garlic). Researchers sample people in the same way. It is not necessary to contact all members of a target audience to understand their opinions, attitudes, and behaviors. Instead, practitioners can learn this information from a properly selected sample of target audience members with a high degree of confidence that they are accurate. The purpose of this chapter is to explain basic aspects of sampling including both probability and nonprobability sampling methods and sample size calculations, in a simple, easy-to-understand manner. Math phobes, note that we use only a relatively small amount of math in this chapter. Instead of manipulating numbers, we want readers to develop a conceptual understanding of the principles of sampling and statistical inference.

SAMPLING BASICS

Even though sampling methods are relatively easy to understand and use, understanding a few basic terms and concepts makes it easier to understand sampling practices. Although these definitions are not terribly interesting, they make it a lot easier to understand principles of sampling and inference. At a basic level, readers should understand the difference between a population and a sample. A *population* or *universe* consists of all the members of a group or an entire collection of objects. In public relations, a population most commonly refers to all the people in a target audience or public. When researchers conduct a *census*, they collect information from all members of a population to measure their attitudes, opinions, behaviors, and other characteristics. These measurements, called *parameters*, are the true values of a population's members; parameters are a characteristic or property of a population. In theory, parameters contain no error because they are the result of information collected from every population member. Often, parameters are expressed in summary form. If our research reveals that 59% of voters in King County, Washington, support a property tax initiative, for example, this characteristic is a parameter of the population of all voters in King County.

Research professionals and social scientists often find it difficult or impossible to conduct a census because they are expensive and time consuming. More important, a census is unnecessary in most situations. By collecting information from a carefully selected subset, or *sample*, of population members, researchers can draw conclusions about the entire population, often with a high degree of accuracy. This is why sampling is such a powerful part of communication campaign research.

A *sample* is a subset of a population or universe. When researchers conduct a survey using a sample, they use the resulting data to produce sample statistics. *Sample statistics* describe the characteristics of the sample in the same way that population parameters describe the characteristics of a population. Statistics result from the *observed scores* of sample members instead of from the true scores of all population members, and they necessarily contain some error because of this. The amount of error contained in sample statistics, however, usually is small enough that researchers can estimate, or infer, the attitudes, behaviors, and other characteristics of a population from sample statistics, often with a high degree of confidence.

If you find all of this confusing, read this section again slowly and it will become more clear, although no more exciting (we suggest taking two aspirin first). This topic and chapter improve in terms of their ease of understanding, but it is important for readers to have a basic understanding of sampling terminology and concepts before we discuss other aspects of sampling. It also becomes more clear as we move into discussions of sample representation, sampling techniques, and sample size calculations.

GENERALIZING FROM A SAMPLE TO A POPULATION

Researchers normally collect data to make generalizations. During a state gubernatorial election in Michigan, for example, a political campaign manager may survey a sample of registered voters to determine the opinions of all registered voters in the state. In this case, the campaign manager wants to generalize the results of the survey from a relatively small sample (perhaps consisting of no more than several hundred people) to all registered voters in the state. This process of generalization, when researchers draw conclusions about a population based on information collected from a sample, is called *inference*. Researchers generalize findings from samples to populations on a regular basis. How can researchers generalize in this way and be confident they are right? A sample must accurately represent the population from which it is drawn to allow investigators to make valid inferences about the population based on sample statistics.

An often-used example from the annals of survey research helps make the point. In 1920, editors of the *Literary Digest* conducted a poll to see whether they could predict the winner of the presidential election between Warren Harding and James Cox. Editors gathered names and addresses from telephone directories and automobile registration lists and sent postcards to people in six states. Based on the postcards they received, the *Literary Digest* correctly predicted that Harding would win the election. *Literary Digest* editors repeated this same general process over the next several elections and correctly predicted presidential election winners in 1920, 1924, 1928, and 1932.

Literacy Digest editors again conducted a poll to predict the winner of the 1936 election. This project was their most ambitious yet. This time, they sent ballots to 10 million people whose names they drew from telephone directories and automobile registration lists, as before. More than 2 million ballots were returned. Based on the results of its survey, the editors predicted that Republican challenger Alfred Landon would receive 57% of the popular vote in a stunning upset over Democratic incumbent Franklin Roosevelt. Roosevelt was reelected, however, by the largest margin in history to date. He received approximately 61% of the popular vote and captured 523 electoral votes to Landon's 8. What went wrong?

Simply put, the sample was unrepresentative. *Literacy Digest* editors drew the sample from telephone directories and automobile registration lists, both of which were biased to upper income groups. At that time, less than 40% of American households had telephones and only 55% of Americans owned automobiles. The omission of the poor from the sample was particularly significant because they voted overwhelmingly for Roosevelt, whereas the wealthy voted primarily for Landon (Freedman, Pisani, & Purves, 1978). Not only was the sample unrepresentative, but the survey method and low response rate (24%) contributed to biased results.

This often-used example illustrates a key point about the importance of sample representativeness. The results of research based on samples that are not representative do not allow researchers to validly generalize, or project, research results. It is unwise for investigators to make inferences about a population based on information gathered from a sample when the sample does not adequately represent a population. It is a simple, but important, concept to understand.

In fact, George Gallup (of Gallup poll notoriety) understood the concept well. In July 1936, he predicted in print that the *Literacy Digest* poll would project Landon as the landslide winner and that the poll would be incorrect. He made these predictions months before the *Literacy Digest* poll was conducted. He also predicted that Roosevelt would win reelection and perhaps receive as much as 54% of the popular vote. Gallup's predictions were correct, even though his numbers concerning the election were off. How could Gallup be sure of his predictions? The primary basis of his explanation was that the *Literacy Digest* reached only middle- and upper-class individuals who were much more likely to vote Republican. In other words, he understood that the *Literacy Digest* sample was not representative (Converse, 1987). As an additional note, for those who believe that a larger sample always is better, here is evidence to the contrary. When researchers use *nonprobability* sampling methods, sample size has no scientifically verifiable effect on the representativeness of a sample. Sample size makes no difference because the sample simply is not representative of the population. A large, unrepresentative sample is as unrepresentative as a small, unrepresentative sample. In fact, had editors

used a probability sampling method along with an appropriate survey method, a sample size of less than 1% of the more than 2 million voters who responded to the *Digest* poll almost certainly would have produced a highly accurate prediction for both the *Literary Digest* editors and George Gallup.

SAMPLING METHODS

Sampling is the means by which researchers select people or elements in a population to represent the entire population. Researchers use a *sampling frame*—a list of the members of a population—to produce a sample, using one of several methods to determine who will be included in the sample. Each person or object in the sample is a *sampling element* or *unit*. When practitioners study target audience members, the sampling frame typically consists of a list of members of a target audience, whereas the sampling unit is an individual person. All the sampling units together compose the sample. If a nonprofit organization wants to examine the perceptions and opinions of its donors, for example, the sampling frame might be a mailing list of donors' names and addresses, whereas the sampling unit would be the individual names and addresses selected from the list as part of the sample. When researchers generate a sample, they select sampling units from the sampling frame.

When researchers draw a sample, their goal is to accurately represent a population. This allows them to make inferences about the population based on information they collect from the sample. There are two types of samples: *probability* and *nonprobability*. Researchers select probability samples in a *random* way so that each member of a population has an equal chance, or probability, of being included in a sample. When researchers draw a nonprobability sample, an individual's chance of being included in a sample is not known. There is no way to determine the probability that any population member will be included in the sample because a *non-random* selection process is used. Some population members may have no chance of being included in a sample, whereas other population members may have multiple chances of being included in a sample.

When researchers select probability, or random, samples, they normally can make accurate inferences about the population under study based on information from the sample. That is, probability samples tend to produce results that are highly generalizable from a sample to a population. When researchers select samples in any way other than probability-based, random sampling, they cannot be sure that a sample accurately represents the population from which it was drawn. In this case, they have no basis for validly making inferences about a population from the sample. Even though a nonprobability sample may perfectly represent a population, investigators cannot scientifically demonstrate its level of representativeness.

For this reason, the results of research that use nonprobability samples are low in generalizability (external validity).

Why use nonprobability sampling if the research results it produces are not representative? In some cases, researchers may use a nonprobability sample because it is quick and easy to generate. At other times, the cost of generating a probability-based sample may be too high, so researchers use a less expensive nonprobability sample instead. The use of a nonprobability sample is not automatically a problem or even necessarily a concern. It is a significant limitation, however, in practitioners' application of research results to campaign planning and problem solving. Research managers often use nonprobability samples in exploratory research or other small-scale studies, perhaps as a precursor to a major study. In addition, some commonly used research methods, such as focus groups or mall intercept surveys, rely exclusively on nonprobability sampling.

The lack of generalizability should serve as a warning to communication campaign managers. Do not assume research results based on nonprobability samples are accurate. When practitioners want to explore a problem or potential solution in an informal fashion, get input on an idea, or obtain limited feedback from members of a target audience, a nonprobability sample normally is an acceptable choice. As editors of the *Literary Digest* discovered, however, nonprobability samples have limitations and should not serve as the sole basis by which practitioners seek to understand audiences and develop programs.

NONPROBABILITY SAMPLING METHODS

There are several methods of generating nonprobability samples. No matter how random the selection process appears in each of these sampling methods, researchers do not select sample members in a random manner. This means that population members have an unequal chance of being selected as part of a sample when investigators use these sampling methods. The most common types of nonprobability sampling are incidental (also called *convenience*) sampling, quota sampling, dimensional sampling, purposive (judgmental) sampling, volunteer sampling, and snowball sampling.

Incidental, or Convenience, Sampling

Researchers select incidental, or convenience, samples by using whoever is convenient as a sample element. A public opinion survey in which interviewers stop and survey those who walk by and are willing to participate constitutes such a sample. Mall intercepts generally rely on convenience samples because their sample consists of shoppers who happen to walk by and are willing to complete a questionnaire. Like all nonprobability

samples, incidental samples do not generally provide accurate estimates of the attributes of a target population. There simply is no way for researchers to determine the degree to which research results from a convenience sample are representative of a population. Like all nonprobability sampling methods, incidental samples are most appropriate when research is exploratory, precise statistics concerning a population are not required, or the target population is impossible to accurately define or locate (Johnson & Reynolds, 2005).

Quota Sampling

When researchers use this sampling method, they often are interested in the subgroups that exist in a population and draw their sample so that it contains the same proportion of subgroups. Investigators fill the quotas nonrandomly, typically using sample members who are convenient to fill subgroup quotas. In practice, research staff members typically base quotas on a small number of population characteristics such as respondents' age, sex, educational level, type of employment, or race or ethnicity. An interviewer conducting a survey on a college campus, for example, might be assigned to interview a certain number of freshmen, sophomores, juniors, and seniors. The interviewer might select the sample nonrandomly by standing in front of the university library and asking people to complete a survey. Interviewers would stop surveying members of individual population subgroups as they filled each quota.

Dimensional Sampling

This method is similar to quota sampling in that researchers select study participants nonrandomly according to predetermined quota, but project managers extend sample quotas to include a variety of population attributes. Generally, interviewers ensure that they include a minimum number of individuals for various combinations of criteria. Extending the college survey example, interviewers might nonrandomly select participants to meet additional criteria, or dimensions. Interviewers might have to interview a minimum number of males and females, traditional and non-traditional students, or married and unmarried students, for example, in addition to the class quota. Interviewers could use a seemingly endless number of potential attributes to stratify a sample.

No matter how many attributes research team members use when selecting a sample, they select both quota and dimensional sample members using nonprobability selection methods. The result is that researchers cannot determine whether their participants fully represent the similarities and differences that exist among subgroups in the population. Ultimately, there is no scientific way to determine whether a nonprobability sample is

representative and no scientific evidence to suggest quota sampling is more representative than other nonprobability sampling methods. Researchers correct the nonprobability selection weakness of quota sampling and dimensional sampling when they use *stratified* sampling, which we address shortly.

Purposive Sampling

In purposive, or *judgmental*, sampling, researchers select sample members because they meet the special needs of the study based on the interviewer's judgment. A researcher's goal when using purposive sampling typically is to examine a specially selected population that is unusually diverse or particularly limited in some way, rather than to study a larger, more uniform population (Johnson & Reynolds, 2005). If a product manufacturer wants to open a new plant in another country, for example, company management needs to learn the concerns of local business, government, and labor leaders. In this case, the sample is relatively small and diverse, and interviewers may simply select sample members using their own discretion to determine which respondents fit into the sample and are "typical" or "representative." This creates situations in which sample-selection decisions may vary widely among interviewers. Even if the definition of the population is reasonably clear, the procedures researchers use when drawing a sample may vary greatly among interviewers, limiting the comparability of sample members (Warwick & Lininger, 1975). These nonrandom selection procedures limit the generalizability of research results based on purposive samples, as is the case with all nonprobability sampling methods.

Volunteer Sampling

When media organizations ask viewers to call in or e-mail their opinions, they are using a volunteer, or self-selected, sample. Instant phone-in polls have become a common way for the media to determine and report so-called public opinion, for example, in an attempt to attract and keep the interests of viewers and listeners. There are numerous potential sources of bias when research is based on a volunteer sample. First, sample representation is hindered because only the people who are exposed to the survey have an opportunity to participate. All other potential respondents are unaware of the poll. Second, those who feel strongly about the topic of a poll may view the survey as an opportunity to vote for their viewpoint. Such individuals may respond more than once and/or encourage other like-minded individuals to respond in the same way. The result is that volunteer samples are not representative, and research results based on volunteer samples are highly untrustworthy. Organizations that use

volunteer samples should use them strictly for their entertainment value, not their scientific value.

Snowball Sampling

When researchers use snowball sampling, they collect data from a limited number of population members and then ask these individuals to identify other members of the population who might be willing to participate in the study. The sample continues to grow as new research participants direct interviewers to additional sample prospects. The sample snowballs, starting from a small number of people and growing larger as each new participant suggests other potential participants.

Researchers may have no choice but to have to rely on snowball sampling when they can locate only a few members of a population. If a social welfare organization wanted to learn about the particular difficulties of migrant workers, for example, it might start by interviewing those migrant workers it could locate. After each interview was concluded, interviewers could ask participants to identify other workers who might be willing to participate in the study. Interviewers hope the sample would grow to a desirable size through this process. Research results based on such a sample, however, have little or no generalizability, no matter how large the sample grows. A snowball sample relies on nonrandom methods of selection, and there is no way to scientifically determine the degree to which it represents the population from which it is drawn because of this. As with all projects based on nonprobability samples, managers need to interpret and generalize research findings resulting from snowball samples carefully.

PROBABILITY SAMPLING METHODS

Researchers generate probability samples using a random selection process so that each member of a population has an equal chance, or probability, of being included in a sample. The use of probability sampling normally allows investigators to make accurate inferences about a population based on information collected from a sample. Investigators' inferences, or conclusions, about the population are not perfectly accurate even when they use probability sampling. Researchers calculate estimates of the population parameter within a given range of possible values at a specific level of probability. The result of this process is that research findings based on probability samples normally are highly representative. That is, they possess a high degree of generalizability, or external validity. The most common type of probability sample is simple random sampling. Common variations of simple random sampling include systematic sampling, stratified sampling, and cluster sampling.

Simple Random Sampling

Researchers must ensure that each member of a population has an equal chance of being included in a sample and select each sample element independently to produce a random sample. Simple random sampling is the most basic method of random sampling, and investigators use it to ensure that the sample they produce is representative of the population. Although true representation never is guaranteed unless a census is taken, the use of a random-selection process significantly reduces the chances of subgroup overrepresentation or underrepresentation, which helps eliminate sample bias. Researchers then can estimate, or infer, population parameters based on sample statistics. Although these inferences are not perfect—they have some error—investigators use statistical procedures to understand this error as noted previously.

From a practical standpoint, the primary requirement for simple random sampling is that researchers clearly and unambiguously identify each member of a population through the use of a comprehensive sampling frame. This allows the direct, independent, and random selection of sample elements, typically through a list in which each element is identified (Warwick & Lininger, 1975). The most common methods of simple random sampling use a list of population members for a sample frame. Research staff members might number each element on the list sequentially, for example, and select the sample by using a table of random numbers or a computer program that produced random numbers. Each number selected by researchers would correspond with a member of the sampling frame. The result is a random sample that normally is highly representative of its population.

If the Public Relations Society of America (PRSA) wanted to survey its members to determine their level of satisfaction with its programs and services, a project manager could take a membership list and assign a number to each PRSA member sequentially. The manager would create the sample by randomly generating numbers assigned to specific PRSA members. Staff members would produce more numbers until they generated an appropriate sample size. If properly conducted, this random process would produce a probability sample of PRSA members who have a high likelihood of accurately representing the attitudes and opinions of all the members of PRSA.

Systematic Random Sampling

Researchers use an unbiased system to select sample members from a list when they use systematic random sampling. This system allows them to generate a probability-based sample that normally is highly representative of the population from which it was drawn, without some of the inconveniences associated with simple random sampling. Those who use simple

random sampling often find the process long and unnecessarily tedious, especially when a population is large. When researchers use systematic random sampling, they develop an uncomplicated system using the total sample size and the size of the population to help them draw a probability-based sample relatively easily.

First, research team members determine the final number of completed interviews they need for a study. Researchers often need to generate a total sample that is several times larger than their targeted number of completed interviews because of the number of sample elements who are difficult to contact or who refuse to participate in a survey. Once researchers determine the total sample size, they determine a *sampling interval* by dividing the number of elements in the sampling frame (this is the total population) by the desired total sample size. The result is a number (n) that researchers use to generate a sample by selecting every n^{th} element from a sampling frame. Researchers must select the first sample element randomly from the frame to produce a probability sample, so they randomly select the first element from within the sampling interval. They complete the sample-selection process by selecting every n^{th} element from the sampling frame and the result is a systematic random sample.

An example helps to clarify systematic random sampling. If corporate personnel managers want to survey their classified staff as part of a program to improve employee relations, their first step is to determine the final number of completed interviews they want for the study. We discuss sample size calculations later in this chapter, but for this example, let's say that after some careful thinking and a little fun with math, administrators determine they want a total of approximately 400 completed interviews from the approximately 6,000 employees who work as full- or part-time classified staff. After some additional calculations (explained in chapter 12), researchers determine that an original total sample size of 850 classified staff members would produce about 400 completed surveys from participants, as shown in Figure 6.1. The projects' directors decide to use a mailing list of classified staff members as a sampling frame because it contains the names and addresses of all classified staff members and has no duplicate listings. They divide the sampling frame (6,000) by the original sample size (850) to determine the sampling interval (approximately 7). Project managers must select the first sample element randomly, so they use a table of random numbers to produce the first number between 1 and 7. If project managers drew the number 5, they would draw the sample by selecting the fifth name on the list and selecting every seventh name after that. Thus, researchers would draw name 5, name 12, name 19, name 26, and so on. By using the sampling interval, researchers produce a systematic random sample.

Systematic random samples and simple random samples are not exactly the same; however, systematic samples closely approximate simple random samples to produce a probability sample that normally is highly representative. In terms of bias, the greatest danger researchers face when

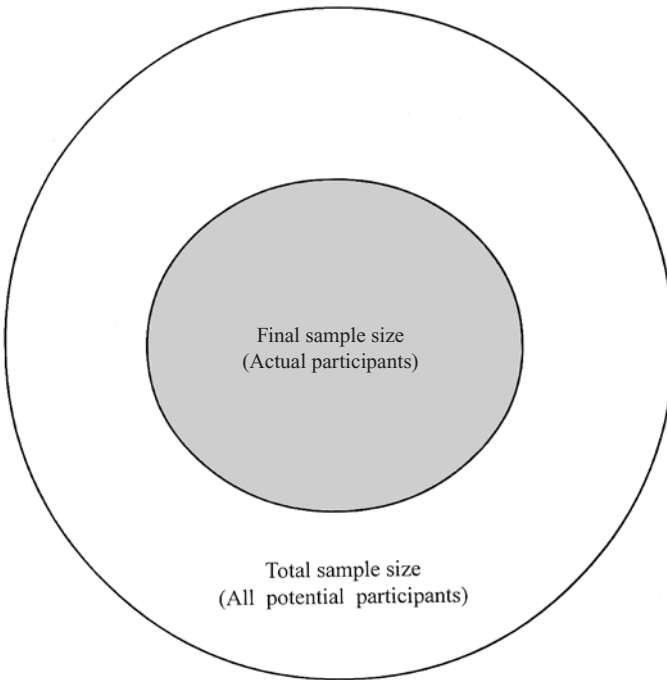


FIG. 6.1. The relationship of final sample size to total sample size. Communication managers need to generate a total original sample that is several times larger than required for their final sample because of the number of people who cannot be contacted or refuse to participate in a survey.

they use systematic sampling is periodicity. *Periodicity* refers to bias that occurs when a sampling list has a cyclical repetition of some population characteristic that coincides with a sampling interval. If this occurs, the sample elements selected are not generalizable to the population from which they were drawn. Researchers should be careful to inspect population lists before sampling to make sure there are no obvious signs of periodicity. When researchers are careful, the potential for bias in systematic sampling normally is small. Ultimately, researchers use systematic random sampling more than simple random sampling because of its simplicity and usefulness in complex sampling situations (Sudman, 1976).

Stratified Sampling

Researchers divide a population into different subgroups, or *strata*, when they engage in stratified sampling, similar to quota sampling. The key difference between the methods is that investigators use a random, probability-based method to select sample elements when they engage in stratified sampling, whereas they use a nonrandom, nonprobability-based

method to select sample elements when they engage in quota sampling. Researchers have two primary reasons for stratification: to control the representativeness of the sample and to use different probability-based selection procedures in different strata (Warwick & Lininger, 1975). Researchers also may use stratified sampling when they are primarily interested in the key similarities and differences among members of strata, when the prior information they have for individual strata is different, or when they want to improve sampling efficiency when research costs differ by strata (Sudman, 1976).

Researchers use two types of stratified sampling: proportional and disproportional. When they use *proportional* sampling, project managers draw sample members from each stratum in proportion to their existence in the population. The resulting sample proportionally represents individual strata as they exist in a population. Researchers use *disproportionate* sampling to help ensure that the overall sample accurately produces results that represent the opinions, attitudes, and behaviors of a significant stratum within the population. Project managers may use disproportionate sampling when strata are too small to be accurately represented in a sample selected through other means. In this case, research staff may find it necessary to weight the data to obtain unbiased estimates of the total population. This may be necessary, for example, when researchers' use of other probability sampling methods would underrepresent the opinions, attitudes, and behaviors of minority members of a population. When researchers use either proportional or disproportional stratified sampling to their advantage, they can produce highly representative, probability-based samples.

Cluster Sampling

Researchers select sample elements using groups rather than individuals when they use cluster sampling. The sample frame consists of clusters rather than individuals, and each cluster serves as a sample element. The clusters researchers use for sampling commonly are preexisting natural groups or administrative groups of the population. These may include geographical designations such as neighborhoods, cities, counties, or zip code areas, for example, or other common groupings such as universities, hospitals, or schools.

Researchers often use cluster sampling to make data collection more efficient (Sudman, 1976). If a metropolitan school district wanted to learn about the attitudes and experiences of its students, it could send interviewers to meet one on one with individually selected student sample members. This process, however, would be expensive and increase the time needed to complete the project. If researchers used schools as sample clusters and randomly selected from among them, the project would require less time and travel, which would increase the efficiency of data collection.

Investigators also use cluster sampling when a comprehensive list of sample elements is not available. If an organization wanted to sample city residents as part of a community relations program, project managers would likely have trouble locating a complete list of all community residents, and the process would be costly and time consuming. If researchers wanted to use cluster sampling, they could create a sampling frame by using city blocks as clusters. After research staff identified and labeled each block, they could randomly select an appropriate number of blocks. Next, researchers would randomly sample dwelling units within each block. Finally, interviewers would randomly sample people living in each dwelling unit and collect data. Researchers call this sampling process *multistage sampling* because sampling takes place in different stages; they select city blocks in stage one, followed by dwelling units in stage two and individual people in stage three.

Researchers' primary concern when using cluster sampling is the potential for increased error relative to other probability-based sampling methods. When investigators use cluster sampling, standard error may increase if sample members' attitudes, behaviors, and other characteristics generally are the same, or homogeneous, within each cluster. In this instance, samples selected from within homogeneous clusters will not reflect the diversity of attitudes and behaviors that exist in the larger population. Project managers can help counter this problem by selecting a high number of small clusters and selecting a relatively low number of sample elements from within each cluster (Johnson & Reynolds, 2005).

Cluster samples, along with systematic and stratified samples, are acceptable alternatives to simple random sampling. In each case, population elements have an equal chance of being included in a sample. Ultimately, researchers' choice of a sampling method often depends on the time and money available for a project, the population being sampled, the subject under investigation, and the availability of a comprehensive list of target population members.

HOW BIG SHOULD A SAMPLE BE?

One of the first questions clients, managers, and others involved in a research project typically ask is "What is the best sample size for a project?" Unfortunately, as is the case so often in life and particularly in survey research, the answer is a firm "it depends." In fact, the methods researchers use to determine the appropriate sample size for a study can be relatively complicated and even controversial. Research professionals often use different formulas to calculate sample size—in some cases based on different assumptions about population characteristics—and may suggest conflicting sample sizes as a result. Several common misperceptions exist concerning sample size calculations including the following:

Myth 1: Bigger samples are better. The *Literary Digest* case demonstrates the fallacy concerning bigger sample sizes. When researchers use probability sampling methods, a mathematically calculated sample size based on an appropriate formula nearly always produces trustworthy results with known ranges of error. Researchers can use simple math to verify this information. When researchers use non-probability sampling methods, there is no scientific way to determine how well a sample represents a population or how much error survey results contain. Remember, a large unrepresentative sample is no more representative than a small unrepresentative sample. In addition, a representative sample that is unnecessarily large is a waste of resources. A sample's size should be the result of a researcher's purposeful decision-making process, not a number that researchers stumble upon as they try to generate the largest sample possible.

Myth 2: As a rule of thumb, researchers should sample a fixed percentage of a population to produce an acceptable sample size. It is not uncommon for those uninitiated in survey research methods to suggest using a fixed percentage of the population to determine sample size. If researchers sampled 10% of a 50,000-person population, for example, they would generate a sample of 5,000 participants. Once again, probability-based sampling methods allow the use of mathematical formulas to calculate sample sizes that normally produce highly trustworthy results with known ranges of error. Arbitrarily sampling a certain percentage of the population is unnecessary and results in an arbitrary sample size. Such a practice is as illogical as if you ate a certain percentage of the food in your refrigerator because you were hungry. Just as the amount of food you eat should be based on your body's needs (with notable exceptions for mocha almond fudge ice cream and chocolate in any form), so should a study's sample size be based on the requirements of a research project instead of on an arbitrary percentage of a population.

Myth 3: Researchers should base sample sizes on industry standards or "typical" sample sizes used in other research projects. In reality, there is little that is standard about a research project. Although researchers may use familiar formulas to determine the sample size for a study, they should not use these formulas without careful consideration. A project's unique needs, the individual characteristics of a population and its resulting sample, and other issues greatly affect sampling decisions. Researchers serve the needs of clients and organizations best when they make thoughtful sample-size decisions, based on the unique requirements of individual research projects.

Having said all of this, we must engage in a small amount of backpedaling. Many communication students (and some campaign practitioners for that matter) have varying degrees of math phobia and short attention spans when it comes to highly technical or theoretical information. It is necessary to understand a little theory and use a small amount of basic algebra to calculate sample sizes. With this in mind, in this text we use simple sample-size calculation formulas and avoid math when possible. We also try to provide a basic conceptual understanding of these formulas and the concepts they use. In short, we take some shortcuts to make these topics as accessible as possible. The result is that we do not follow our own advice in some instances. Please keep in mind that there is more to this sometimes complicated topic than we discuss in this text. If you find these basic concepts and formulas easy to understand, or if you will be involved in research on a regular basis, you should read more about additional aspects of sampling and sample-size calculations so that you are fully informed.

For the math-challenged among us, we offer our encouragement. Read the next section slowly, draw pictures of the concepts if it helps you understand them, and try the math out yourself. Put in a little effort and you should emerge with a clear understanding of the topic. To explain sample-size calculations, first we provide a conceptual understanding of sample-calculation concepts and processes. Then, we do some basic sample-size calculations, based on the concepts we have explained. Finally, we calculate the amount of error that exists in survey data once researchers have completed a survey.

CALCULATING THE APPROPRIATE SAMPLE SIZE

Anyone can determine with precision the optimal size for a sample, provided they understand a few key concepts based on probability theory and a bell-shaped curve. These concepts include sample distribution and standard deviation, confidence level, confidence interval, and variance. Once you grasp these concepts, it is easy to understand the basis for sample-size calculations; the rest is simply a matter of applying the formulas.

Sample Distribution and Standard Deviation

Sample distribution and standard deviation are the first and, in some ways, most complex concepts to understand. A *sample distribution* is a grouping or arrangement of a characteristic that researchers measure for each sample member, and it reflects the frequency with which researchers assign sample characteristics to each point on a measurement scale (Williams, 1992). Almost any characteristic that researchers can measure has a sampling distribution, but in survey research investigators typically study sample members' opinions, attitudes, behaviors, and related characteristics. If we

were to chart a sampling distribution, the result would be shaped like a bell, provided the sampling distribution was normal. It would be tall in the middle where the average of the sampling distribution is located because most people would be near the average. There would be fewer people toward either edge, or tails, of the bell because fewer people would have characteristics or behaviors so far above or below the average.

If we were practitioners at a university health facility, for example, we might conduct a survey to better understand smoking behavior among students. We could ask a randomly selected student sample to fill out a questionnaire that contained attitudinal and behavioral questions, including a question about the number of cigarettes participants had smoked in the previous 7 days. Participants' responses likely would vary greatly. Many students would have smoked no cigarettes in the previous 7 days, whereas other students would have smoked a high number of cigarettes. When we compute students' responses to our smoking question, we could use the information to generate a sample distribution. If our research revealed the average number of cigarettes smoked in the past week by participants was 3.5, this number would be placed under the middle of the curve at its tallest point and most participants would be near the average, or mean, in the large part of the bell-shaped distribution. Our sample's smoking distribution would get smaller at its tails because fewer participants would smoke in numbers that were far above or below average. Figure 6.2 contains a normally distributed, bell-shaped curve for the smoking example.

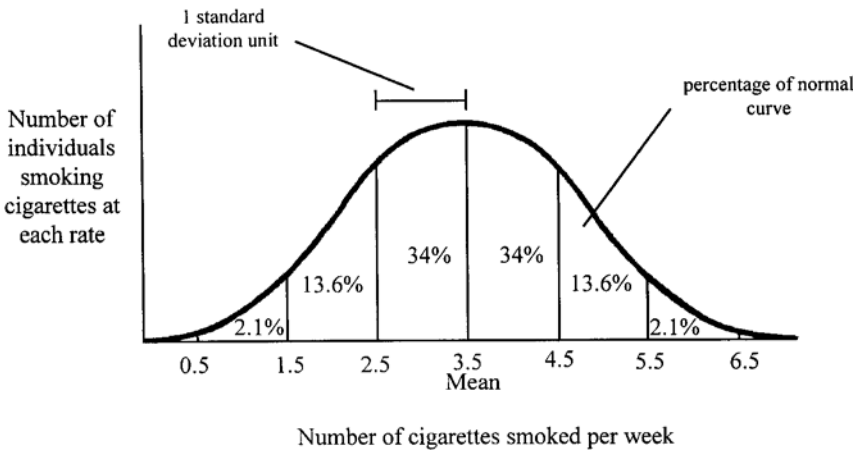


FIG. 6.2. Smoking distribution example. The number in each portion of the curve shows the percentage of the sample that corresponds to each segment. For example, 34% of this sample smokes between 2.5 and 3.5 cigarettes per week. The percentages added together equal more than 99% of a normal distribution. The segments of the curve are divided according to standard deviations from the mean.

As we planned our campaign, we could make inferences about the population (all students at our university) based on the responses of our sample. Error occurs when researchers take measurements from a sample and use them to make inferences about a population because there are differences between a sample distribution and a population distribution. We could not determine the exact average number of cigarettes smoked weekly by students at our university, for example, unless we conducted a census by interviewing every student. We did not conduct a census in this example and because of this, the responses of our sample would not exactly represent the true responses of the population. In our smoking survey, our sample mean for cigarettes smoked in the past 7 days might be 3.5, whereas the true value for the population might be 3.8. The difference between the opinions and behaviors of the sample and the opinions and behaviors of the population is *error*.

As researchers, we must understand this error, so we use a tool to measure it called *standard deviation*. Standard deviation is a standardized measure of dispersion (or variation) around a mean. Basically, a standard deviation is a standardized unit of measurement that researchers use to measure distances from a sampling distribution's midpoint to its outer limits (don't get lost here). Think of standard deviation as a simple unit of measurement. Researchers use standard deviation to measure distance from the mean in a bell-shaped curve in the same way a carpenter uses inches to measure the length of a board, as Figure 6.3 shows.

Researchers use standard deviation for various purposes. If the publishers of a book survey 10 people and ask them to read and rate it using a

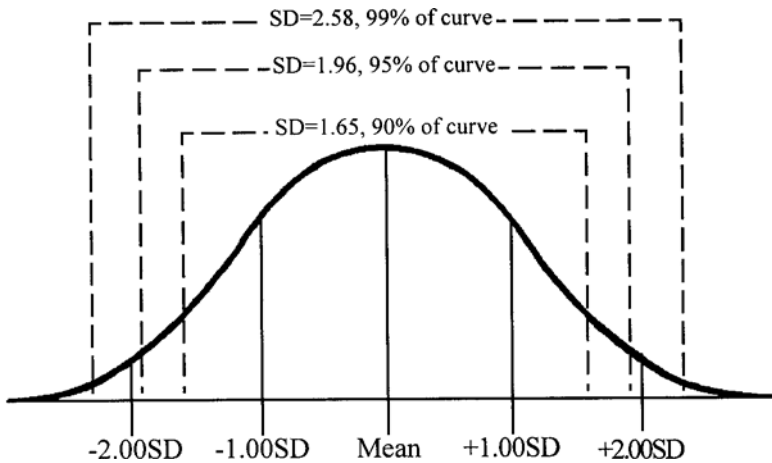


FIG. 6.3. Normally distributed bell-shaped curve. Along the normal distribution, 1.65 standard deviations (SD) measure 90% of the curve; 1.96 standard deviations measure 95% of the curve; and 2.58 standard deviations measure 99% of the curve.

scale of 0 to 10, for example, the text might receive an average rating of 5. If all 10 people who read the book actually rated the text as a 5, the average rating is highly accurate and there is no standard deviation. If 5 people rate the text as a 10, however, and 5 people rate the text as a 0, the mean rating still is 5. This time, however, the average rating is not very accurate. No one, in fact, actually gave the text a 5 rating. The standard deviation would be relatively large because there is a lot of dispersion among the scores. Although the means are the same in each case, they actually are different, and standard deviation helps us measure and understand this. Using our smoking survey example, if every participant in our smoking survey said they smoked 3.5 cigarettes in the past 7 days, our mean would be highly accurate and we would have no deviation from the mean. When we ask sample members about their smoking habits, however, we will undoubtedly receive different responses, and we can use the mean and standard deviation to understand these responses.

How do standard deviation and sample distribution help us when we calculate sample size? A standard deviation gives researchers a basis for estimating the probability of correspondence between the normally distributed, bell-shaped curve of a perfect population distribution and a probability-based sample distribution that always contains some error. Researchers call standard deviation measurements *standard* because they associate with, or measure, specific areas under a normal curve. One standard deviation measures about 68% of a normally distributed curve; two standard deviations measure a little more than 95% of a normally distributed curve; and three standard deviations measure more than 99% of a normally distributed curve. Research professionals use standard deviations to determine the confidence level associated with a sample, as we demonstrate later in this chapter.

Confidence Level

A *confidence level* is the degree of certainty researchers can have when they draw inferences about a population based on data from a sample. Basically, it is the level of probability researchers have that they can accurately generalize a characteristic they find in a sample to a population. In essence, the confidence level answers the question, "How confident are we that our sample is representative of the population?" A confidence level of 90% means researchers are 90% confident that the sample accurately represents the population. In the same way, a confidence level of 95% means researchers are 95% confident that the inferences they draw about the population from the sample are accurate.

This raises an important question: Are researchers really 90% or 95% confident about the representativeness of the sample, or are they simply guessing, perhaps based on their experience? In fact, researchers' claims

of a confidence level are accurate because the confidence level is based on standard deviations. Remember, a standard deviation allows researchers to estimate probability between a normally distributed population curve and a less-than-perfect sampling distribution because standard deviation measurements associate with specific areas under the curve. A standard deviation of 1.65 measures 90% of a normally distributed curve, a standard deviation of 1.96 measures 95% of a normally distributed curve, and a standard deviation of 2.58 measures 99% of a normally distributed curve (remember these numbers because we will use them again shortly).

This means that when researchers calculate sample size, they select standard deviations associated with specific areas under a normally distributed curve to provide the desired confidence level. When investigators use 1.65 in the sample-size formula, they calculate a sample size that provides a 90% confidence level; when they use 1.96 in the formula, they calculate a sample size that provides a 95% confidence level; when they use 2.58 in the formula, they calculate a sample size that provides a 99% confidence level.

Most often, researchers use 1.96 standard deviations to calculate sample size, resulting in a 95% confidence level. A confidence level of 95% means our sample statistics will more-or-less accurately represent the true parameter of a population 95% of the time. Here is another way to think about this: If we conducted a survey of the same population 100 times, our sample responses would be accurate in 95 of the 100 surveys we conducted. The 95% confidence level is a standard convention of social science, but researchers can use other confidence levels. In particular, if researchers desire an exceptionally high degree of confidence when making inferences about a population based on data from a sample, they may choose a higher confidence level. Rarely do researchers use a lower confidence level.

Confidence Interval

A *confidence interval* is a range or margin of error that researchers permit when making inferences from a sample to a population. As noted, the inferences researchers make about a population based on sample data are not completely accurate. Unless investigators conduct a census, the observed values they collect from a sample (statistics) will not provide completely accurate information concerning a population's true values (parameters).

The population parameter falls somewhere within the range of the confidence interval, although researchers never are exactly sure where the parameter is located unless they conduct a census. The confidence interval usually is stated as a positive-to-negative range, such as $\pm 3\%$ error or $\pm 5\%$ error. A confidence interval of $\pm 3\%$ has a total error margin of 6%, whereas a confidence interval of $\pm 5\%$ has a total error margin of 10%. If 57% of registered voters in California express support for a citizens' initiative in a

survey with a $\pm 5\%$ confidence interval, for example, the true population value may be as high as 62% (+5%) or as low as 52% (-5%).

What is an acceptable confidence interval for survey results? As is often the case in survey research, the answer depends on various factors. Many applied communication and market research surveys have a $\pm 5\%$ confidence interval, but there is nothing critical about this range of error. Researchers commonly choose smaller confidence levels when they want to reduce the margin of error and increase the precision of the inferences they draw concerning a population. When media organizations poll the public to predict election outcomes, for example, they often use a smaller confidence interval, such as $\pm 3\%$. Ultimately, researchers should make decisions about confidence intervals based on the necessities and challenges of individual research projects.

It may surprise you to learn that the confidence level and the confidence interval do not have to add to 100%. Those new to research often assume the confidence level and confidence interval must add to 100% because researchers often conduct surveys with a $\pm 5\%$ error at a 95% confidence level. It is incidental that these numbers add up to 100. It is legitimate to conduct a survey with a $\pm 3\%$ margin of error at a 95% confidence level, for example, or a survey with a $\pm 2.5\%$ margin of error at a 99% confidence level. In addition, many researchers use a 95% confidence level as a standard and only make adjustments to the confidence interval when calculating sample size. As we noted previously, researchers should make decisions concerning confidence levels and confidence intervals based on the requirements of individual research projects.

Variance

Simply put, *variance* is dispersion. When researchers calculate sample size, it helps them to understand how the characteristic or variable they are examining is dispersed throughout a population. If we want to understand the use of public transportation in our community as a way to reduce traffic and pollution, for example, it would be useful to know the percentage of community members who actually use public transportation. In short, we want to know how public transportation use is dispersed throughout our community as a characteristic of the population.

For research purposes, it is useful to consider variance as a simple percentage. Community members who use public transportation, for example, fit into one category that makes up a certain percentage of the population. Community members who do not use public transportation do not belong in this category and make up the remaining percentage of the population. Together, the percentages add up to 100%. Researchers can examine the dispersion of most variables this way because a population can be divided into two categories on the basis of almost any characteristic. This

includes, for example, students who smoke cigarettes and students who do not smoke cigarettes, community residents who live in a certain neighborhood and residents who live in other neighborhoods, workers who are employed and workers who are unemployed, and people who drink coffee and people who do not.

Any time researchers examine a variable or characteristic, they want to know its dispersion within a population because they can use this information to help them calculate sample size. Population members who have a characteristic or variable fit into a single category, and researchers use this to distinguish them from the rest of the population. In the formula we examine shortly, the percentage of a population that belongs to a category is expressed as a decimal. The remaining percentage of the population (that does not belong to the category) also is expressed as a decimal and subtracted from 1. Together, these two numbers add to 1.0 or 100% of the population.

Despite the importance of variance, researchers often set aside variance percentages when they calculate sample size because percentages only reflect the dispersion of a single characteristic or variable in a population. Researchers commonly examine multiple variables in a single survey, each with a different percentage of dispersion. Each variable would require a different sample size, which is impractical and unnecessary. Researchers address this problem by using the largest measure of variance available to calculate sample size because, at a minimum, it provides an acceptable measure of dispersion for all variables. To use the largest measure of variance, researchers use .5 (or 50%) as the percentage of a population that belongs to a category. Researchers also use .5 as the percentage for the rest of the population because $1 - .5 = .5$ and these percentages add up to 1.0, or 100%, of the population. Although it is not necessary for researchers to use .5 and $1 - .5$ in every sample-size calculation, this practice is regularly required by the necessities of a multifaceted research project, and so we use it in all of our sample-size calculations.

SAMPLE-SIZE FORMULA

Now that you understand standard deviation, confidence level, confidence interval, and variance, you are ready to calculate sample size. Researchers commonly use the following formula—or formulas that are similar but more complicated—to calculate sample size:

$$n = \left(\frac{cl}{ci} \right)^2 (v)(1 - v)$$

where

n (number) = the number of completed interviews or what we call the final sample size

cl (confidence level) = the standard deviation associated with a specific area under a normal curve and corresponding to the desired confidence level (by definition, 90% confidence level = 1.65; 95% confidence level = 1.96; and 99% confidence level = 2.58)

ci (confidence interval) = the margin of error expressed as a decimal ($\pm 3\%$ error would be expressed as .03; $\pm 5\%$ error would be expressed as .05; $\pm 10\%$ error would be expressed as .10)

v (variance) = the variance or distribution of a variable in a population, expressed as a percentage in decimal form. For our purposes, variance is always .5. Note also that $1 - v$ is the percentage of a population that has no variable distribution; $1 - v$ always is .5 when v is .5, as we have recommended.

Here is a basic sample-size calculation using this formula. We calculate the sample size at a 95% confidence level and a $\pm 5\%$ margin of error, or confidence interval:

$$n = \left(\frac{1.96}{.05} \right)^2 (.5) (.5) = 384$$

Based on this formula, we need a final sample size of 384 people—or 384 completed interviews—to produce findings with a $\pm 5\%$ margin of error at a 95% confidence level.

What if we want less error (a smaller confidence interval), meaning more trust in the precision of our survey results? It is easy to adjust the formula to fit the demands of any research situation. In the following calculations, for example, we determine sample sizes based on different confidence levels. We calculate each sample size with a $\pm 5\%$ confidence interval, but with different confidence levels to show how different confidence levels affect sample size. To change confidence levels, we use standard deviations that correspond to different areas under a normally distributed, bell-shaped curve. Recall that the standard deviation for a 90% confidence level is 1.65; the standard deviation for a 95% confidence level is 1.96, and the standard deviation for a 99% confidence level is 2.58. Notice that we increase sample size as we increase the confidence level. The only difference in each calculation is the level of confidence researchers have when they make inferences from a sample to a population. Here is the final sample size—or number of completed interviews—needed for a survey with a 90% confidence level and a $\pm 5\%$ margin of error:

$$n = \left(\frac{1.65}{.05} \right)^2 (.5) (.5) = 272$$

Here is the final sample size (number of completed interviews) for a survey with a 95% confidence level and a $\pm 5\%$ margin of error:

$$n = \left(\frac{1.96}{.05} \right)^2 (.5)(.5) = 384$$

Finally, here is the sample size (number of completed interviews) for a survey with a 99% confidence level with a $\pm 5\%$ margin of error:

$$n = \left(\frac{2.58}{.05} \right)^2 (.5)(.5) = 666$$

How do changes in the margin of error, or confidence interval, affect final sample size (the completed number of interviews) researchers need? In the following calculations, we determine sample sizes with the same level of confidence but differing margins of error. Each sample size is calculated at a 95% confidence level. Here is the final sample size for a survey with a $\pm 10\%$ margin of error at a 95% confidence level:

$$n = \left(\frac{1.96}{.10} \right)^2 (.5)(.5) = 96$$

Here is the final sample size for a survey with a $\pm 5\%$ margin of error at a 95% confidence level:

$$n = \left(\frac{1.96}{.05} \right)^2 (.5)(.5) = 384$$

Finally, here is the final sample size for a survey with a $\pm 3\%$ margin of error at a 95% confidence level:

$$n = \left(\frac{1.96}{.03} \right)^2 (.5)(.5) = 1,067$$

In each case, we reduced the margin of error while maintaining a consistent level of confidence.

ERROR CALCULATIONS

The same information you learned to calculate sample size also will help you calculate the margin of error for a survey, once you have collected data. In most cases, the number of completed interviews—or what we call the *final sample size*—is not 384 or 1,060 completed interviews, even if this is researchers' targeted sample size. Researchers aiming for a specific

sample size typically collect additional interviews for various reasons. Research staff may have to throw out some interviews, for example, because of problems with data collection, such as a survey that is only partially complete. At other times, researchers may collect a larger sample size so they have a stronger basis from which to make sample subgroup comparisons. Regardless of the reason, researchers can use standard deviation, confidence level, and variance to calculate the margin of error that exists in a survey's results based on its final sample size. Here is the formula:

$$e = cl \sqrt{\frac{(v)(1 - v)}{n}}(100)$$

where

e (error) = the final margin of error for the completed survey based on sample size

cl (confidence level) = the standard deviation associated with a specific area under a normal curve and corresponding to the desired confidence level (as before, 90% confidence level = 1.65; 95% confidence level = 1.96; and 99% confidence level = 2.58)

v (variance) = the variance or distribution of a variable in a population, expressed as a percentage in decimal form. As before, variance always is .5, and 1 - *v* is the percentage of a population that has no variable distribution; 1 - *v* always is .5, when *v* is .5 as we have recommended;

n (number) = the number of completed interviews or what we call the final sample size.

Here is the margin of error for a survey in which the final sample size, or number of completed interviews, is 485. The calculation is made based on a 95% confidence level:

$$1.96 \sqrt{\frac{(.5)(.5)}{485}}(100) = 4.45$$

In this example, the margin of error for this survey is ±4.45% based on 485 completed interviews.

How do changes in the confidence level affect the margin of error, or sampling interval, for survey results? In the following calculations, we determine margins of error for survey results using the same final sample size, or completed number of interviews, at different levels of confidence. We calculate each margin of error using a final sample size of 575. Here is the margin of error at a 90% confidence level:

$$1.65 \sqrt{\frac{(.5)(.5)}{575}}(100) = 3.44$$

Here is the margin of error at a 95% confidence level:

$$1.96\sqrt{\frac{(.5)(.5)}{575}}(100) = 4.09$$

Here is the margin of error at a 99% confidence level:

$$2.58\sqrt{\frac{(.5)(.5)}{575}}(100) = 5.38$$

These calculations reveal the trade-off between confidence level and the margin of error, or confidence interval, for a survey. If researchers want to increase their level of confidence or certainty as they make inferences from sample data to a population, they must be willing to accept a larger range of error in their survey's results. If researchers desire a smaller range of error, they must be willing to accept a lower confidence level when they make inferences.

ISSUES AND ASSUMPTIONS

The formulas we have presented require various assumptions and raise some important issues. We have addressed many of these issues and assumptions in the preceding sections, but note that you may need to alter these formulas or disregard them completely, as the assumptions on which we have based these formulas change. One of the primary assumptions of all sample-size formulas, for example, concerns researchers' use of probability sampling methods. When researchers use nonprobability sampling methods, no sample-size formula will produce an accurate result because it is impossible for researchers to determine the representativeness of the sample.

One issue we have not addressed concerns the need to correct the formula according to a population's size. Researchers sometimes use sample-size formulas that contain something called *finite population correction*. Finite population correction is an adjustment factor that is part of a sample-size formula. Table 6.1 contains population-corrected final sample sizes for probability-based survey results with a $\pm 5\%$ margin of error at a 95% confidence level. The appropriate sample size for a population of 1 million people is 384, the same sample size we calculated for a survey with a 95% confidence level and a $\pm 5\%$ margin of error.

Is it necessary for researchers to correct for population size? Generally, most researchers have little need for population correction unless the size of the population is small and the sample is more than 5% of the total population (Czaja & Blair, 1996). In most sample surveys, population correction makes little difference in sample-size calculations and researchers simply

TABLE 6.1
Population-Corrected Sample Sizes

<i>Population Size (N)</i>	<i>Sample Size (n)</i>	<i>Population Size (N)</i>	<i>Sample Size (n)</i>	<i>Population Size (N)</i>	<i>Sample Size (n)</i>
5	5	650	242	2,500	333
10	10	700	248	3,000	341
15	15	750	254	3,500	346
20	19	800	260	4,000	351
25	24	850	265	4,500	354
50	44	900	269	5,000	357
75	63	950	274	6,000	361
100	80	1,000	278	7,000	364
150	108	1,100	285	8,000	367
200	132	1,200	291	9,000	368
250	152	1,300	297	10,000	370
300	169	1,400	302	15,000	375
350	183	1,500	306	20,000	377
400	196	1,600	310	25,000	378
450	207	1,700	313	50,000	381
500	217	1,800	317	75,000	382
550	226	1,900	320	100,000	383
600	234	2,000	322	275,000+	384

Note: Figures reported are for probability-based survey results with a $\pm 5\%$ margin of error at a 95% confidence level. Calculations are based on Cochran's (1977) formula for finite population correction. Further information is available in Kish (1965). According to this formula, even populations more than 1 million require a sample size of 384.

exclude a population correction factor because it is unnecessary. In fact, researchers would generally use the same sample size for a survey of registered voters in Chicago, a survey of registered voters in Illinois, or a survey of all registered voters in the entire United States! Although there are important exceptions, once a population reaches a certain size, sample sizes generally remain consistent. For this reason, and to keep our sample calculations simple, the sample size formula we presented does not include population correction.

FINAL THOUGHTS

Sampling is a powerful tool that helps practitioners obtain accurate information at a reasonable cost. Researchers' selection of a proper sampling method is as important as their selection of a proper research method to the success of a study. Even the most carefully planned and executed study will produce untrustworthy results if research managers use an improper sampling method. Although sampling can be complex, it is in readers' own best interest to learn all they can about the sample-selection procedures used

in a study. Few people would buy an automobile without first inspecting the vehicle they are purchasing, yet a surprising number of practitioners make research “purchases” without ever inspecting one of the most critical elements of their research project, the sampling procedures used in a study.

As demonstrated in this chapter, it is not necessary for practitioners to become sampling experts to understand many important issues related to sampling selection processes. It is necessary, however, for practitioners to understand basic distinctions in sampling methods and to work with researchers to ensure that the sample used in a study has the greatest chance of accurately representing the attitudes, opinions, and behaviors of the population from which it is drawn.