

3

Sampling

How well a sample represents a population depends on the sample frame, the sample size, and the specific design of selection procedures. If probability sampling procedures are used, the precision of sample estimates can be calculated. This chapter describes various sampling procedures and their effects on the representativeness and precision of sample estimates. Two of the most common ways of sampling populations, area probability and random-digit-dialing samples, are described in some detail.

There are occasions when the goal of information gathering is not to generate statistics about a population but to describe a set of people in a more general way. Journalists, people developing products, political leaders, and others sometimes just want a sense of people's feelings without great concern about numerical precision. Researchers do pilot studies to measure the range of ideas or opinions that people have or the way that variables seem to hang together. For these purposes, people who are readily available (friends, coworkers) or people who volunteer (magazine survey respondents, people who call talk shows) may be useful. Not every effort to gather information requires a strict probability sample survey. For the majority of occasions when surveys are undertaken, however, the goal is to develop statistics about a population. This chapter is about sampling when the goal is to produce numbers that can be subjected appropriately to the variety of statistical techniques available to social scientists. Although many of the same general principles apply to any sampling problem, the chapter focuses on sampling people.

The way to evaluate a sample is not by the results, the characteristics of the sample, but by examining the process by which it was selected. There are three key aspects of sample selection:

1. The sample frame is the set of people that has a chance to be selected, given the sampling approach that is chosen. Statistically speaking, a sample only can be representative of the population included in the sample frame. One design issue is how well the sample frame corresponds to the population a researcher wants to describe.
2. Probability sampling procedures must be used to designate individual units for inclusion in a sample. Each person must have a known chance of selection set by the sampling procedure. If researcher discretion or respondent characteristics such as

respondent availability or initiative affect the chances of selection, there is no statistical basis for evaluating how well or how poorly the sample represents the population; commonly used approaches to calculating confidence intervals around sample estimates are not applicable.

3. The details of the sample design, its size and the specific procedures used for selecting units, will influence the precision of sample estimates, that is, how closely a sample is likely to approximate the characteristics of the whole population.

These details of the sampling process, along with the rate at which information actually is obtained from those selected, constitute the facts needed to evaluate a survey sample.

Response rates are discussed in Chapter 4, which also includes a brief discussion of quota sampling, a common modification of probability sampling that yields *nonprobability* samples. In this chapter, sampling frames and probability sampling procedures are discussed. Several of the most common practical strategies for sampling people are described. Interested readers will find much more information on sampling in Kish (1965), Sudman (1976), Kalton (1983), Groves (2004), Henry (1990), and Lohr (1998). Researchers planning to carry out a survey almost always would be well advised to obtain the help of a sampling statistician. This chapter, however, is intended to familiarize readers with the issues to which they should attend, and that they will likely encounter, when evaluating the sampling done for a survey.

THE SAMPLE FRAME

Any sample selection procedure will give some individuals a chance to be included in the sample while excluding others. Those people who have a chance of being included among those selected constitute the sample frame. The first step in evaluating the quality of a sample is to define the sample frame. Most sampling of people is done in one of three ways:

1. Sampling is done from a more or less complete list of individuals in the population to be studied.
2. Sampling is done from a set of people who go somewhere or do something that enables them to be sampled (e.g., patients who received medical care from a physician, people who entered a store in a particular time period, or people who attended a meeting). In these cases, there is not an advance list from which sampling occurs; the creation of the list and the process of sampling may occur simultaneously.
3. Addresses or housing units are sampled as a first stage of selecting a sample of people living in those housing units. Housing units can be sampled from lists of addresses, by sampling geographic areas and then sampling housing units located on those geographic areas, or by sampling telephone numbers that can be associated with housing units.

There are three characteristics of a sample frame that a researcher should evaluate:

1. *Comprehensiveness*, that is, how completely it covers the target population.
2. Whether or not a person's *probability of selection* can be calculated.
3. *Efficiency*, or the rate at which members of the target population can be found among those in the frame.

Comprehensiveness. From a statistical perspective, a sample can only be representative of the sample frame, that is, the population that actually had a chance to be selected. Most sampling approaches leave out at least a few people from the population the researcher wants to study. For example, household-based samples exclude people who live in group quarters, such as dormitories, prisons, and nursing homes, as well as those who are homeless. Available general lists, such as those of people with driver's licenses, registered voters, and homeowners, are even more exclusive. Although they cover large segments of some populations, they also omit major segments with distinctive characteristics. As a specific example, published telephone directories omit those without landline telephones, those who have requested that their numbers not be published, and those who have been assigned a telephone number since the most recent directory was published. In some central cities, such exclusions amount to more than 50% of all households. In such cities, a sample drawn from a telephone directory would be representative of only about half the population, and the half that is represented could easily be expected to differ in many ways from the half that is not.

A growing threat to telephone surveys is the increase of cell phone use. In the past, most telephone surveys depended on sampling telephone numbers that could be linked to households. Those households that are served only by cell phones are left out of such samples. In 2011, that was about a third of all U.S. households (Blumberg and Luke, 2011).

E-mail addresses provide another good example. There are some populations, such as those in business or school settings, that have virtually universal access to e-mail, and more or less complete lists of the addresses of these populations are likely to be available. On the other hand, as an approach to sampling households in the general population, sampling those with e-mail addresses leaves out many people and produces a sample that is very different from the population as a whole in many important ways. Moreover, there is not currently a way to create a good list of all or even most of those who have e-mail addresses.

Two recent innovations, spurred by the desire to conduct surveys via the Internet, deserve mentioning. First, large numbers of people have been recruited via the Internet to participate in surveys and other research studies. These people fill out initial baseline questionnaires covering a large number of characteristics. The answers to these questions can then be used to "create" a sample from the total pool of volunteers that roughly matches those of the whole population a researcher wants to study. When such a "sample" is surveyed, the results may or may not yield accurate information about the whole population. Obviously, no one is included in such a sample who does not use the Internet and is not interested in volunteering to be in the surveys. Often, the same

people participate in numerous surveys, thereby further raising questions about how well the respondents typify the general population (Couper, 2007).

In an effort to address some of those concerns, another approach is to sample households as a first step to recruiting a pool of potential respondents for Internet surveys. Those without access to computers may be given a computer to use and Internet service. Thus, the sample recruited for inclusion potentially is representative of all those living in households and the fact that there is an effort to actively recruit them, rather than passively relying on people to volunteer, may make the resulting sample less “self-selected.” Nonetheless, from a statistical perspective, statistics based on samples from that pool do not necessarily apply to the balance of the population. Rather, in both of the previous examples, those responding to a survey can only be said to be representative of the populations that volunteered or agreed to be on these lists (Couper, 2007). The extent to which they are like the rest of the population must be evaluated independently of the sampling process (Baker et al., 2010; Yeager et al., 2011).

A key part of evaluating any sampling scheme is determining the percentage of the population one wants to describe that has a chance of being selected and the extent to which those excluded are distinctive. Very often a researcher must make a choice between an easier or less expensive way of sampling a population that leaves out some people and a more expensive strategy that is also more comprehensive. If a researcher is considering sampling from a list, it is particularly important to evaluate the list to find out in detail how it was compiled, how and when additions and deletions are made, and the number and characteristics of people likely to be left off the list.

Probability of selection. Is it possible to calculate the probability of selection of each person sampled? A procedure that samples records of visits to a doctor over a year will give individuals who visit the doctor numerous times a higher chance of selection than those who see the doctor only once. It is not necessary that a sampling scheme give every member of the sampling frame the same chance of selection, as would be the case if each individual appeared once and only once on a list. It is essential, however, that the researcher be able to find out the probability of selection for each individual selected. This may be done at the time of sample selection by examination of the list. It also may be possible to find out the probability of selection at the time of data collection.

In the previous example of sampling patients by sampling doctor visits, if the researcher asks selected patients the number of visits to the physician they had in a year or if the researcher could have access to selected patients’ medical records, it would be possible to adjust the data at the time of analysis to take into account the different chances of selection. If it is not possible to know the probability of selection of each selected individual, however, it is not possible to estimate accurately the relationship between the sample statistics and the population from which it was drawn.

Any sample that is based on volunteering or self-selection will violate this standard. “Quota samples,” discussed near the end of Chapter 4, are another common example of using procedures for which the probability of selection cannot be calculated.

Efficiency. In some cases, sampling frames include units that are not members of the target population the researcher wants to sample. Assuming that eligible people can be identified at the point of data collection, being too comprehensive is not a problem. Hence, a perfectly appropriate way to sample people aged 65 or older living in households is to draw a sample of all households, find out if there are persons of the target age who are living in selected households, then exclude those households with no such residents. Random-digit dialing samples select telephone numbers (many of which are not in use) as a way of sampling housing units with telephones. The only question about such designs is whether or not they are cost effective.

Because the ability to generalize from a sample is limited by the sample frame, when reporting results the researcher must tell readers who was or was not given a chance to be selected and, to the extent that it is known, how those omitted were distinctive.

SELECTING A ONE-STAGE SAMPLE

Once a researcher has made a decision about a sample frame or approach to getting a sample, the next question is specifically how to select the individual units to be included. In the next few sections, the various ways that samplers typically draw samples are discussed.

Simple Random Sampling

Simple random sampling is, in a sense, the prototype of population sampling. The most basic ways of calculating statistics about samples assume that a simple random sample was drawn. Simple random sampling approximates drawing a sample out of a hat: Members of a population are selected one at a time, independent of one another and without replacement; once a unit is selected, it has no further chance to be selected.

Operationally, drawing a simple random sample requires a numbered list of the population. For simplicity, assume that each person in the population appears once and only once. If there were 8,500 people on a list, and the goal was to select a simple random sample of 100, the procedure would be straightforward. People on the list would be numbered from 1 to 8,500. Then a computer, a table of random numbers, or some other generator of random numbers would be used to produce 100 different numbers within the same range. The individuals corresponding to the 100 numbers chosen would constitute a simple random sample of that population of 8,500. If the list is in a computerized data file, randomizing the ordering of the list, then choosing the first 100 people on the reordered list, would produce an equivalent result.

Systematic Samples

Unless a list is short, has all units prenumbered, or is computerized so that it can be numbered easily, drawing a simple random sample as previously described can be laborious. In such situations, there is a way to use a variation called systematic

sampling that will have precision equivalent to a simple random sample and can be mechanically easier to create. Moreover, the benefits of stratification (discussed in the next section) can be accomplished more easily through systematic sampling.

When drawing a systematic sample from a list, the researcher first determines the number of entries on the list and the number of elements from the list that are to be selected. Dividing the latter by the former will produce a fraction. Thus, if there are 8,500 people on a list and a sample of 100 is required, $100/8,500$ of the list (i.e., 1 out of every 85 persons) is to be included in the sample. In order to select a systematic sample, a start point is designated by choosing a random number that falls within the sampling interval, in this example, any number from 1 to 85. The randomized start ensures that it is a chance selection process. Starting with the person in the randomly selected position, the researcher proceeds to take every 85th person on the list.

Most statistics books warn against systematic samples if a list is ordered by some characteristic, or has a recurring pattern, that will differentially affect the sample depending on the random start. As an extreme example, if members of a male-female couples club were listed with the male partner always listed first, any even number interval would produce a systematic sample that consisted of only one gender even though the club as a whole is evenly divided by gender. It definitely is important to examine a potential sample frame from the perspective of whether or not there is any reason to think that the sample resulting from one random start will be systematically different from those resulting from other starts in ways that will affect the survey results. In practice, most lists or sample frames do not pose any problems for systematic sampling. When they do, by either reordering the lists or adjusting the selection intervals, it almost always is possible to design a systematic sampling strategy that is at least equivalent to a simple random sample.

Stratified Samples

When a simple random sample is drawn, each new selection is independent, unaffected by any selections that came before. As a result of this process, any of the characteristics of the sample may, by chance, differ somewhat from the population from which it is drawn. Generally, little is known about the characteristics of individual population members before data collection. It is not uncommon, however, for at least a few characteristics of a population to be identifiable at the time of sampling. When that is the case, there is the possibility of structuring the sampling process to reduce the normal sampling variation, thereby producing a sample that is more likely to look like the total population than a simple random sample. The process by which this is done is called stratification.

For example, suppose one had a list of college students. The list is arranged alphabetically. Members of different classes are mixed throughout the list. If the list identifies the particular class to which a student belongs, it would be possible to rearrange the list to put freshmen first, then sophomores, then juniors, and finally seniors, with all classes grouped together. If the sampling design calls for selecting a sample of 1 in 10 of the members on the list, the rearrangement would ensure that exactly 1/10 of the freshmen were selected, 1/10 of the sophomores, and so forth. On the other hand, if either a simple random sample or a systematic sample was selected from the original alphabetical list, the proportion of the sample in the freshman year would be subject to

normal sampling variability and could be slightly higher or lower than was the case for the population. Stratifying in advance ensures that the sample will have exactly the same proportions in each class as the whole population.

Consider the task of estimating the average age of the student body. The class in which a student is a member almost certainly is correlated with age. Although there still will be some variability in sample estimates because of the sampling procedure, structuring the representation of classes in the sampling frame also will constrain the extent to which the average age of the sample will differ by chance from the population as a whole.

Almost all samples of populations of geographic areas are stratified by some regional variable so that they will be distributed in the same way as the population as a whole. National samples typically are stratified by region of the country and also by urban, suburban, and rural locations. Stratification only increases the precision of estimates of variables that are related to the stratification variables. Because some degree of stratification is relatively simple to accomplish, however, and because it never hurts the precision of sample estimates (as long as the probability of selection is the same across all strata), it usually is a desirable feature of a sample design.

Different Probabilities of Selection

Sometimes stratification is used as a first step to vary the rates of selection of various population subgroups. When probabilities of selection are constant across strata, a group that constitutes 10% of a population will constitute about 10% of a selected sample. If a researcher wanted a sample of at least 100 from a population subgroup that constituted 10% of the population, a simple random sampling approach would require an overall sample of 1,000. Moreover, if the researcher decided to increase the sample size of that subgroup to 150, this would entail taking an additional 500 sample members into the sample, bringing the total to 1,500, so that 10% of the sample would equal 150.

Obviously, there are occasions when increasing a sample in this way is not very cost effective. In the latter example, if the researcher is satisfied with the size of the samples of other groups, the design adds 450 unwanted interviews to add 50 interviews that are wanted. In some cases, therefore, an appropriate design is to select some subgroup at a higher rate than the rest of the population.

As an example, suppose that a researcher wished to compare male and female students, with a minimum of 200 male respondents, at a particular college where only 20% of the students are male. Thus a sample of 500 students would include 100 male students. If male students could be identified in advance, however, one could select male students at twice the rate at which female students were selected. In this way, rather than adding 500 interviews to increase the sample by 100 males, an additional 100 interviews over the basic sample of 500 would produce a total of about 200 interviews with males. Thus, when making male-female comparisons, one would have the precision provided by samples of 200 male respondents and 400 female respondents. To combine these samples, the researcher would have to give male respondents a weight of half that given to females to compensate for the fact that they were sampled at twice the rate of the rest of the population. (See Chapter 10 for more details about weighting.)

Even if individual members of a subgroup of interest cannot be identified with certainty in advance of sampling, sometimes the basic approach previously outlined can be applied. For instance, it is most unusual to have a list of housing units that identifies the race of occupants in advance of contact. It is not uncommon, however, for Asian families to be more concentrated in some neighborhood areas than others. In that instance, a researcher may be able to sample households in areas that are predominantly Asian at a higher than average rate to increase the number of Asian respondents. Again, when any group is given a chance of selection different from other members of the population, appropriate compensatory weighting is required to generate accurate population statistics for the combined or total sample.

A third approach is to adjust the chance of selection based on information gathered after making contact with potential respondents. Going back to the college student survey, if student gender could not be ascertained in advance, the researchers could select an initial sample of 1,000 students, have interviewers ascertain the gender of each student, then have them conduct a complete interview with all selected male students (200) but only half of the female students they identified (400). The result would be exactly the same as with the approach previously described.

Finally, one other technical reason for using different probabilities of selection by stratum should be mentioned. If what is being measured is much more variable in one group than in another, it may help the precision of the resulting overall estimate to oversample the group with the high level of variability. Groves (2004) provides a good description of the rationale and how to assess the efficiency of such designs.

MULTISTAGE SAMPLING

When there is no adequate list of the individuals in a population and no way to get at the population directly, multistage sampling provides a useful approach.

In the absence of a direct sampling source, a strategy is needed for linking population members to some kind of grouping that can be sampled. These groupings can be sampled as a first stage. Lists then are made of individual members of selected groups, with possibly a further selection from the created list at the second (or later) stage of sampling. In sampling terminology, the groupings in the last stage of a sample design are usually referred to as “clusters.” The following section illustrates the general strategy for multistage sampling by describing its use in three of the most common types of situations in which lists of all individuals in the target population are not available.

Sampling Students From Schools

If one wanted to draw a sample of all students enrolled in the public schools of a particular city, it would not be surprising to find that there was not a single complete list of such individuals. There is, however, a sample frame that enables one to get at and include all the students in the desired population: namely, the list of all the public schools in that city. Because every individual in the study population can be attached to one and only one of those units, a perfectly acceptable sample of students can be

selected using a two-stage strategy: first selecting schools (i.e., the clusters) and then selecting students from within those schools.

Assume the following data:

There are 20,000 students in a city with 40 schools

Desired sample = 2,000 = 1/10 of students

Four different designs or approaches to sampling are presented in Figure 3.1. Each would yield a probability sample of 2,000 students.

Figure 3.1 Four Designs for Selecting Students From Schools

	Probability of Selection at Stage 1 (schools)	×	Probability of Selection at Stage 2 (students in selected schools)	=	Overall Probability of Selection
(a) Select all schools, list all students, and select 1/10 students in each school	1/1	×	1/10	=	1/10
(b) Select 1/2 the schools, then select 1/5 of all students in them	1/2	×	1/5	=	1/10
(c) Select 1/5 of the schools, then select 1/2 of all students in them	1/5	×	1/2	=	1/10
(d) Select 1/10 schools, then collect information about all students in them	1/10	×	1/1	=	1/10

SOURCE: Fowler, 2008.

The four approaches listed all yield samples of 2,000; all give each student in the city an equal (1 in 10) chance of selection. The difference is that from top to bottom, the designs are increasingly less expensive; lists have to be collected from fewer schools, and fewer schools need to be visited. At the same time, the precision of each sample is likely to decline as fewer schools are sampled and more students are sampled per school. The effect of this and other multistage designs on the precision of sample estimates is discussed in more detail in a later section of this chapter.

Area Probability Sampling

Area probability sampling is one of the most generally useful multistage strategies because of its wide applicability. It can be used to sample any population that can be defined geographically, for example, the people living in a neighborhood, a city, a state, or a country. The basic approach is to divide the total target land area into exhaustive, mutually exclusive subareas with identifiable boundaries. These subareas are the clusters. A sample of subareas is drawn. A list then is made of housing units in selected subareas, and a sample of listed units is drawn. As a final stage, all people in selected housing units may be included in the sample, or they may be listed and sampled as well.

This approach will work for jungles, deserts, sparsely populated rural areas, or downtown areas in central cities. The specific steps to drawing such a sample can be very complicated. The basic principles, however, can be illustrated by describing how one could sample the population of a city using city blocks as the primary subarea units to be selected at the first stage of sampling. Assume the following data:

A city consists of 400 blocks

20,000 housing units are located on these blocks

Desired sample = 2,000 housing units = 1/10 of all housing units

Given this information, a sample of households could be selected using a strategy parallel to the previous selection of students. In the first stage of sampling, blocks (i.e., the clusters) are selected. During the second stage, all housing units on selected blocks are listed and a sample is selected from the lists. Two approaches to selecting housing units are presented in Figure 3.2.

Parallel to the school example, the first approach, involving more blocks, is more expensive than the second; it also is likely to produce more precise sample estimates for a sample of a given size.

Figure 3.2 Two Designs for a Two-Stage Sample of Housing Units

	Probability of Selection at Stage 1 (blocks)	×	Probability of Selection at Stage 2 (housing units in selected blocks)	=	Overall Probability of Selection
(a) Select 80 blocks (1/5), then take 1/2 of units on those blocks	1/5	×	1/2	=	1/10
(b) Select 40 blocks (1/10), then take all units on those blocks	1/10	×	1/1	=	1/10

SOURCE: Fowler, 2008.

None of the previous sample schemes takes into account the size of the Stage 1 groupings (i.e., the size of the blocks or schools). Big schools and big blocks are selected at the same rates as small ones. If a fixed fraction of each selected group is to be taken at the last stage, there will be more interviews taken from selected big schools or big blocks than from small ones; the size of the samples (cluster sizes) taken at Stage 2 will be very divergent.

If there is information available about the size of the Stage 1 groups, it is usually good to use it. Sample designs tend to provide more precise estimates if the number of units selected at the final step of selection is approximately equal in all clusters. Other advantages of such designs are that sampling errors are easier to calculate and the total size of the sample is more predictable. To produce equal-sized clusters, Stage 1 units should be sampled proportionate to their size.

The following example shows how blocks could be sampled proportionate to their size as the first stage of an area probability approach to sampling housing units (apartments or single family houses). The same approach could be applied to the previous school example, treating schools in a way analogous to blocks in the following process.

1. Decide how many housing units are to be selected at the last stage of sampling—the average cluster size. Let us choose 10, for example.
2. Make an estimate of the number of housing units in each Stage 1 unit (block).
3. Order the blocks so that geographically adjacent or otherwise similar blocks are contiguous. This effectively stratifies the sampling to improve the samples, as previously discussed.
4. Create an estimated cumulative count across all blocks of housing units. A table like Figure 3.3 will result.

Figure 3.3 Example of Selecting Housing Units From Blocks

Block Number	Estimated Housing Units	Cumulative Housing Units	Hits (Random Start = 70; Interval = 100 HUs)
1	43	43	–
2	87	130	70
3	99	229	170
4	27	256	–
5	15	271	270

SOURCE: Fowler, 2008.

Determine the interval between clusters. If we want to select 1 in 10 housing units and a cluster of about 10 on each selected block, we need an interval of 100 housing units between clusters. Put another way, instead of taking 1 house at an interval of every 10 houses, we take 10 houses at an interval of every 100 houses; the rate is the same, but the pattern is “clustered.”

After first choosing a random number from 1 to 100 (the interval in the example) as a starting point, we proceed systematically through the cumulative count, designating the primary units (or blocks) hit in this first stage of selection. In the example, the random start chosen (70) missed block 1 (though 43 times in 100 it would have been hit); the 70th housing unit was in block 2; the 170th housing unit was in block 3; and the 270th housing unit was located in block 5.

A list then is made of the housing units on the selected blocks (2, 3, and 5). The next step is to select housing units from those lists. If we were sure the estimates of the sizes of blocks were accurate, we could simply select 10 housing units from each selected block, using either simple random or systematic sampling; a systematic sample would usually be best because it would distribute the chosen units around the block.

It is common for estimates of the size of Stage 1 units such as blocks to be somewhat in error. We can correct for such errors by calculating the rate at which housing units are to be selected from blocks as:

$$\text{Rate of HU selection on block} = \frac{\text{Ave. cluster size}}{\text{Estimated HUs on Block}} = \frac{10}{87} = \frac{1}{8.7} \quad (\text{On Block 2})$$

In our example, we would take 1 per 8.7 housing units on Block 2, 1 per 9.9 housing units on Block 3, and 1 per 1.5 housing units on Block 5. If a block is bigger than expected (e.g., because of new construction), more than 10 housing units will be drawn; if it is smaller than expected (e.g., because of demolition), fewer than 10 housing units will be drawn. If it is exactly what we expected (e.g., 87 housing units on block 2), we take 10 housing units (87/8.7 = 10). In this way, the procedure is self-correcting for

errors in initial estimates of block size, while maintaining the same chance of selection for housing units on all blocks. No matter the estimated or actual size of the block, the chance of any housing unit being selected is 1 in 10.

The area probability sample approach can be used to sample any geographically defined population. Although the steps are more complicated as the area gets bigger, the approach is the same. The key steps to remember are the following:

- All areas must be given some chance of selection. Combine areas where no units are expected with adjacent areas to ensure a chance of selection; new construction may have occurred or estimates may be wrong.
- The probability of selecting a block (or other land area) times the probability of selecting a housing unit from a selected block should be constant across all blocks.

Using Address Lists

The area probability approaches previously outlined usually entail creating special-purpose lists of housing units associated with selected land areas. There are, however, lists of addresses that are available that can be used to draw samples of addresses in the United States. The most comprehensive list is developed by the United States Postal Service, which compiles and regularly updates lists of all the addresses to which it delivers mail. These lists are available to researchers through commercial vendors.

These lists can be used to create the frame for the second (or later) stage of an area probability sample or to form the first stage frame for sampling a town or other defined area. In the past, rural mailing addresses were not necessarily housing units. When mail is delivered to a post office box, obviously it cannot be used to sample a housing unit. However, a high percentage of rural housing units now have street names and addresses, so this is a much smaller problem than it used to be.

In some cities, some mailing addresses in multi-unit structures do not identify housing units very clearly. So an interviewer might have trouble knowing exactly which unit in a three-family structure had been selected, if the address did not include something like “2nd floor.” However, while these issues can be significant challenges in some neighborhoods, they affect a relatively small percentage of all housing units. Moreover, a great appeal of such address-based sampling is that the samples can be used for data collection in at least three different modes: in-person interviewers, mail, and by telephone for the roughly 60% of housing units that can be matched using reverse directories to a landline number. Households can also be contacted by mail and asked to go to a website to respond to a survey online (Messer and Dillman, 2011; Millar and Dillman, 2011; Brick, Williams and Montaquila, 2011). Thus, address-based samples greatly enhance the potential for multimode data collection strategies (see Chapter 5).

Random-Digit Dialing

Random-digit dialing (RDD) provides an alternative way to draw a sample of housing units in order to sample the people in those households. Suppose the 20,000 housing

units in the previous example are covered by six telephone exchanges. One could draw a probability sample of 10% of the housing units that have telephones as follows:

1. There is a total of 60,000 possible telephone numbers in those 6 exchanges (10,000 per exchange). Select 6,000 of those numbers (i.e., 10%), drawing 1,000 randomly generated, four-digit numbers per exchange.
2. Dial all 6,000 numbers. Not all the numbers will be household numbers; in fact, many of the numbers will not be working, will be disconnected or temporarily not in service, or will be businesses. Because 10% of all possible telephone numbers that could serve the area have been called, about 10% of all the households with landline telephones in that area will be reached by calling the sample of numbers.

This is the basic random-digit-dialing approach to sampling. The obvious disadvantage of the approach as described is the large number of unfruitful calls. Nationally, fewer than 25% of possible numbers are associated with residential housing units; the rate is about 30% in urban areas and about 10% in rural areas. Waksberg (1978) developed a method of taking advantage of the fact that telephone numbers are assigned in groups. Each group of 100 telephone numbers is defined by a three-digit area code, a three-digit exchange, and two additional numbers (area code-123-45XX). By carrying out an initial screening of numbers by calling one random number in a sample of groups, then calling additional random numbers only within the groups of numbers where a residential number was found, the rate of hitting housing units can be raised to more than 50%. In this design, the groups of 100 telephone numbers are the clusters.

Most survey organizations now use a list-assisted approach to RDD. With the advancement of computer technology, companies can compile computerized versions of telephone listings. These computerized phone books are updated every 3 months. Once all these books are in a computer file, a search can yield all clusters (area code-123-45XX) that have at least one published residential telephone number. These companies can then produce a sample frame of all possible telephone numbers in clusters that have at least one published residential telephone number. Sampling can now be carried out using this sample frame.

This approach has two distinct advantages. The first is that the initial screening of telephone numbers required by the Waksberg method is no longer needed. The construction of the sample frame has already accomplished this. The second advantage is that the sample selected using this frame is no longer clustered. By using all clusters that contain residential telephone numbers as a sample frame, a simple or systematic random sample of telephone numbers can be drawn. This approach to RDD is more cost effective and efficient than its predecessors were.

The accumulation of lists of individuals and their characteristics has made possible some other efficiencies for telephone surveys. One comparatively simple advance is that reverse telephone directories can be used to tie addresses to some telephone numbers. One of the downsides of RDD is that households do not receive advance notice that an interviewer will be calling. Lists make it possible to sort selected numbers into groups (or strata) based on whether or not there is a known residential address

associated with a number. Those for whom there is a known address can be sent an advance letter.

More elaborately, if there are lists of people who have known characteristics that are targeted for a survey—an age group, those living in a particular geographic area, people who gave to a particular charity—a stratum can be made of telephone numbers likely to connect to households that are being targeted. Numbers in the other strata or for which information is not available may be sampled at lower rates, thereby giving all households a known chance of selection but increasing the efficiency of the data collection by concentrating more effort on households likely to yield eligible respondents. Note that if the probabilities of selection are not the same for all respondents, weighting must be used at the analysis stage, as described in Chapter 10.

RDD became perhaps the main way of doing general population surveys in the late 1970s when most households in the United States had telephone service of the type now called landlines. The growing use of individual cell phones has posed a growing problem for RDD sampling. In 2011, about one third of housing units in the United States did not have household (landline) telephone service (Blumberg and Luke, 2011). It is possible to sample from both landline and cell phone lists and services, but the complexity of sampling, data collection, and postsurvey weighting are greatly increased if cell phone numbers are included in the sample frames. To give one example of the complexity: RDD sampling uses area codes to target populations in defined geographic areas. However, cell phone numbers are much less tied to where people actually live. A survey based on cell phone area codes will reach some people who live outside the targeted geographic area and, worse, will omit those who live in the area but whose cell phones have distant area codes.

Cell phone samples are drawn from banks of exchanges from which cell phone numbers have been assigned, just like the early versions of landline number sampling. This means many numbers in a cell phone number sample are not working. Because cell phones belong to individuals, there are also a lot of cell phone numbers that are not linked to an adult 18 or older, a common standard for eligibility for a survey. Because many people who have cell phones also have landlines in their homes, analysis of data coming from a combined cell phone and landline sample have to be adjusted to take into account multiple chances of selection that some people have. Because cell phones travel with people, interviewers are likely to call people at times or in places where an interview is not possible. Finally, another challenge of doing surveys on cell phones is that many people have to pay for the time they spend on their cell phones. See Brick, Dipko, Presser, Tucker, and Yuan (2006) and Lavrakas, Shuttles, Steeh, and Fienberg, (2007) for a fuller discussion of the challenges of using cell phone samples.

Like any particular sampling approach, RDD is not the best design for all surveys. Additional pros and cons will be discussed in Chapter 5. The introduction of RDD as one sampling option made a major contribution to expanding survey research capabilities in the last 20 years of the 20th century. With the growth of cell phones and response rate challenges faced by all those doing telephone interviewing (discussed in Chapter 4), the extent to which RDD sampling will continue to be a staple of survey research is not clear at this point.

Respondent Selection

Both area probability samples and RDD designate a sample of housing units. There is then the further question of who in the household should be interviewed. The best decision depends on what kind of information is being gathered. In some studies, the information is being gathered about the household and about all the people in the household. If the information is commonly known and easy to report, perhaps any adult who is home can answer the questions. If the information is more specialized, the researcher may want to interview the household member who is most knowledgeable. For example, in the National Health Interview Survey, the person who “knows the most about the health of the family” is to be the respondent for questions that cover all family members.

There are, however, many things that an individual can report only for himself or herself. Researchers almost universally feel that no individual can report feelings, opinions, or knowledge for some other person. There are also many behaviors or experiences (e.g., what people eat or drink, what they have bought, what they have seen, or what they have been told) that usually can only be reported accurately by self-reporters.

When a study includes variables for which only self-reporting is appropriate, the sampling process must go beyond selecting households to sampling specific individuals within those households. One approach is to interview every eligible person in a household. (So there is no sampling at that stage.) Because of homogeneity within households, however, as well as concerns about one respondent influencing a later respondent’s answers, it is more common to designate a single respondent per household. Obviously, taking the person who happens to answer the phone or the door would be a nonprobabilistic and potentially biased way of selecting individuals; interviewer discretion, respondent discretion, and availability (which is related to working status, lifestyle, and age) would all affect who turned out to be the respondent. The key principle of probability sampling is that selection is carried out through some chance or random procedure that designates specific people. The procedure for generating a probability selection of respondents within households involves three steps:

1. Ascertain how many people living in a household are eligible to be respondents (e.g., how many are 18 or older).
2. Number these in a consistent way in all households (e.g., order by decreasing age).
3. Have a procedure that objectively designates one person to be the respondent.

Kish (1949) created a detailed procedure for designating respondents using a set of randomized tables that still is used today. When interviewing is computer assisted, it is easy to have the computer select one of the eligible household members. The critical features of the procedure are that no discretion be involved and that all eligible people in selected households have a known (and nonzero) probability of selection.

One of the concerns about respondent selection procedures is that the initial interaction on first contacting someone is critical to enlisting cooperation. If the respondent selection procedure is too cumbersome or feels intrusive, it may adversely affect the

rate of response. Thus, there have been various efforts to find streamlined ways to sample adults in selected households. Groves and Lyberg (1988) review several strategies for simplifying respondent selection procedures.

One popular method is the “last birthday” method. The household contact is asked to identify the adult who last had a birthday, and that person is the designated respondent. In principle, this should be an unbiased way to select a respondent. In practice, it depends on the initial contact having information about all household members’ birthdays and actually following the selection rules as they are supposed to.

Another relatively new approach keys selection to the person the interviewer first talks with. First, the number of eligible people in the household is determined. If there are two or more eligible, a randomized algorithm chooses either the initial informant at the appropriate rate or chooses among the “other” eligible adults (if there is more than one) (Rizzo, Brick, & Park, 2004).

However the respondent is chosen, when only one person is interviewed in a household, a differential rate of selection is introduced. If an adult lives in a one-adult household, he or she obviously will be the respondent if the household is selected. In contrast, an adult living in a three-adult household only will be the respondent one third of the time. Whenever an identifiable group is selected at a different rate from others, weights are needed so that oversampled people are not overrepresented in the sample statistics. In the example earlier in this chapter, when male students were selected at twice the rate of female students, their responses were weighted by one-half so that their weighted proportion of the sample would be the same as in the population. The same general approach applies when one respondent is chosen from households with varying numbers of eligible people.

The simplest way to adjust for the effect of selecting one respondent per household is to weight each response by the number of eligible people in that household. Hence, if there are three adults, the weight is three; if there are two eligible adults, the weight is two; and if there is only one eligible adult, the weight is one. If a weighting scheme is correct, the probability of selection times the weight is the same for all respondents. (See Chapter 10.)

DRAWING SAMPLES FROM TWO OR MORE SAMPLE FRAMES

All of the discussion to this point has been about deriving a population from a single frame, or with a single approach. However, sometimes the best sampling design involves using two different frames to sample a population, then combining the results. One reason for using multiple frames is to improve the coverage. The other common reason for dual-frame designs is to improve the efficiency of data collection.

Improving Coverage

We previously noted that using a sample of landline telephone numbers to sample housing units has become seriously compromised by the fact that nearly a third of

housing units in the United States no longer are served by landlines. Most of those households include one or more people who have mobile or cell phones. Neither a landline nor a cellphone-based sampling approach covers all households. By sampling from both frames, researchers can come closer to covering the whole population than with either approach alone.

There are some serious challenges to combining results from using two sample frames. The primary challenge is tied to the third standard for sampling frames, the calculation of the probabilities of selection. In order to combine data from two samples of the same population, it is necessary to be able to calculate the probability of each respondent's being selected to be in the sample.

For each person interviewed based on living in a household with landline service, the probability of selecting that household is multiplied by the number of eligible people living in that household to get the probability of selecting that person via landline sampling. In addition, it must be ascertained whether or not that respondent uses a cell phone via which he or she could have been interviewed via the cellphone sampling.

For each person interviewed based on cell phone sampling, the probability of selecting that particular cell phone to call has to be calculated. In addition, it must be ascertained whether or not that person could have been selected to be interviewed via the landline sampling by asking if he or she lives in a housing unit with landline service and, if so, how many eligible people live there.

The resulting sample will include the following classes of respondents:

1. Interviewed by landline, could only be reached by landline.
2. Interviewed by landline, could have been reached by cell phone as well.
3. Interviewed by cell phone, could only have been reached by cell phone.
4. Interviewed by cell phone, could have been reached by landline as well.

Each of these groups of respondents had a different probability of selection, and there will be variation in the probabilities of selection within groups as well. Once all these probabilities have been calculated, weights can be applied to adjust for the different probabilities of selection, as discussed further in Chapter 10. At that point, the data from both samples can be combined and meaningful analysis can begin.

Improving Sampling Efficiency

Suppose a survey of Irish Americans living in a city was desired. It was estimated that about 10% of the adults in the city considered themselves to be Irish-American. One approach would be to draw a sample of households and conduct a brief screening interview to identify Irish Americans, then conduct an interview when one was found. However, such an approach would entail screening thousands of households to identify, say, 1,000 eligible respondents. The local Irish American Society has compiled a list of people in town who belong to Irish organizations, have signed up for Irish events, and/or who contributed to Irish charities. The estimate is that perhaps half the Irish

Americans in the city are on the list, although some of the people on the list might not meet the criteria for being Irish American. The researchers would not want to draw the entire sample from the list, as those on the list are, by definition, more active in the Irish community than those not on the list. However, if the researchers selected 500 people from the list, they then would only have to screen half as many households to complete their sample target of 1,000. The efficiency and the cost of the survey would be considerably reduced.

To analyze the results, as previously shown, respondents would be broken into three groups:

1. Selected from the list; all of the people in the community on the list had a chance to be selected in the household sample.
2. Identified and selected from the household sample, but name also appeared on the list.
3. Identified and selected from the household sample, but name did not appear on the list.

The probability of selection would be calculated for each respondent, and weights would be applied such that the probability of selection times the weight was the same for each person from whom data were collected (as described in Chapter 10 in more detail). The result would be an unbiased sample of the Irish Americans living in the city.

MAKING ESTIMATES FROM SAMPLES AND SAMPLING ERRORS

The sampling strategies previously presented were chosen because they are among the most commonly used and they illustrate the major sampling design options. A probability sampling scheme eventually will designate a specific set of households or individuals without researcher or respondent discretion. The basic tools available to the researcher are simple random and systematic sampling, which are modified by stratification, unequal rates of selection, and clustering. The choice of a sampling strategy rests in part on feasibility and costs; it also involves the precision of sample estimates. A major reason for using probability sampling methods is to permit use of a variety of statistical tools to estimate the precision of sample estimates. In this section, the calculation of such estimates and how they are affected by features of the sample design are discussed.

Researchers usually have no interest in the characteristics of a sample per se. The reason for collecting data about a sample is to reach conclusions about an entire population. The statistical and design issues in this chapter are considered in the context of how much confidence one can have that the characteristics of a sample accurately describe the population as a whole.

As described in Chapter 2, a way to think about sampling error is to think of the distribution of means one might get if many samples were drawn from the same population with the same procedure. Although some sources of error in surveys are biasing and produce systematically distorted figures, sampling error is a random (and hence not a systematically biasing) result of sampling. When probability procedures are used to

select a sample, it is possible to calculate how much sample estimates will vary by chance because of sampling.

If an infinite number of samples are drawn, the sample estimates of descriptive statistics (e.g., means) will form a normal distribution around the true population value. The larger the size of the sample and the less the variance of what is being measured, the more tightly the sample estimates will bunch around the true population value, and the more accurate a sample-based estimate usually will be. This variation around the true value, stemming from the fact that by chance samples may differ from the population as a whole, is called “sampling error.” Estimating the limits of the confidence one can have in a sample estimate, given normal chance sampling variability, is one important part of evaluating figures derived from surveys.

The design of sample selection (specifically, whether it involves stratification, clustering, or unequal probabilities of selection) affects the estimates of sampling error for a sample of a given size. The usual approach to describing sampling errors, however, is to calculate what they would be for a simple random sample, and then to calculate the effects of deviations from a simple random sampling design. Hence, the calculation of sampling errors for simple random samples is described first.

Sampling Errors for Simple Random Samples

This is not a textbook on sampling statistics. Estimating the amount of error one can expect from a particular sample design, however, is a basic part of the survey design process. Moreover, researchers routinely provide readers with guidelines regarding error attributable to sampling, guidelines that both the knowledgeable reader and the user of survey research data should know and understand. To this end, a sense of how sampling error is calculated is a necessary part of understanding the total survey process.

Although the same logic applies to all statistics calculated from a sample, the most common sample survey estimates are means or averages. The statistic most often used to describe sampling error is called the *standard error* (of a mean). *It is the standard deviation of the distribution of sample estimates of means that would be formed if an infinite number of samples of a given size were drawn.* When the value of a standard error has been estimated, one can say that 67% of the means of samples of a given size and design will fall within the range of ± 1 standard error of the true population mean; 95% of such samples will fall within the range of ± 2 standard errors. The latter figure (± 2 standard errors) often is reported as the “confidence interval” around a sample estimate.

The estimation of the standard error of a mean is calculated from the variance and the size of the sample from which it was estimated:

$$SE = \sqrt{\frac{\text{Var}}{n}}$$

SE = standard error of a mean

Var = the variance (the sum of the squared deviations from the sample mean over n)

n = size of the sample on which an estimate such as a mean is based

The most common kind of mean calculated from a sample survey is probably the percentage of a sample that has a certain characteristic or gives a certain response. It may be useful to show how a percentage is the mean of a two-value distribution.

A mean is an average. It is calculated as the sum of the values divided by the number of cases: $\Sigma x/n$. Now suppose there are only two values, 0 (no) and 1 (yes). There are 50 cases in a sample; 20 say “yes” when asked if they are married, and the rest say “no.” If there are 20 “yes” and 30 “no” responses, calculate the mean as

$$\frac{\Sigma x}{n} = \frac{(20 \times 1 + 30 \times 0)}{50} = \frac{20}{50} = 0.40$$

A percentage statement, such as 40% of respondents are married, is just a statement about the mean of a 1/0 distribution; the mean is .40. The calculation of standard errors of percentages is facilitated by the fact that the variance of a percentage can be calculated readily as $p \times (1 - p)$, where p = percentage having a characteristic (e.g., the 40% married in the previous example) and $(1 - p)$ is the percentage who lack the characteristic (e.g., the 60% not married).

We have already seen that the standard error of a mean is as follows:

$$SE = \sqrt{\frac{\text{Var}}{n}}$$

Because $p(1 - p)$ is the variance of a percentage,

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

is the standard error of a percentage. In the previous example, with 40% of a sample of 50 people being married, the standard error of that estimate would be as follows:

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.40 \times 0.60}{50}} = \sqrt{\frac{0.24}{50}} = 0.07$$

Thus we would estimate that the probability is .67 (i.e., ± 1 standard error from the sample mean) that the true population figure (the percentage of the whole population that is married) is between .33 and .47 ($.40 \pm .07$). We are 95% confident that the true population figure lies within two standard errors of our sample mean, that is, between .26 and .54 ($.40 \pm .14$).

Table 3.1 is a generalized table of sampling errors for samples of various sizes and for various percentages, provided that samples were selected as simple random samples. Each number in the table represents two standard errors of a percentage. Given knowledge (or an estimate) of the percentage of a sample that gives a particular answer, the table gives 95% confidence intervals for various sample sizes. In the previous example, with 50 cases yielding a sample estimate of 40% married, the table

Table 3.1 Confidence Ranges for Variability Attributable to Sampling*

Sample Size	Percentage of Sample With Characteristic				
	5/95	10/90	20/80	30/70	50/50
35	7	10	14	15	17
50	6	8	11	13	14
75	5	7	9	11	12
100	4	6	8	9	10
200	3	4	6	6	7
300	3	3	5	5	6
500	2	3	4	4	4
1,000	1	2	3	3	3
1,500	1	2	2	2	2

SOURCE: Fowler, 2008.

NOTE: Chances are 95 in 100 that the real population figure lies in the range defined by the number indicated in table, given the percentage of sample reporting the characteristic and the number of sample cases on which the percentage is based.

*This table describes variability attributable to sampling. Errors resulting from nonresponse or reporting errors are not reflected in this table. In addition, this table assumes a simple random sample. Estimates may be subject to more variability than this table indicates because of the sample design or the influence of interviewers on the answers they obtained; stratification might reduce the sampling errors below those indicated here.

reports a confidence interval near .14, as we calculated. If a sample of about 100 cases produced an estimate that 20% were married, the table says we can be 95% sure that the true figure is 20% \pm 8 percentage points (i.e., 12% to 28%).

Several points about the table are worth noting. First, it can be seen that increasingly large samples always reduce sampling errors. Second, it also can be seen that adding a given number of cases to a sample reduces sampling error a great deal more when the sample is small than when it is comparatively large. For example, adding 50 cases to a sample of 50 produces a quite noticeable reduction in sampling error. Adding 50 cases to a sample of 500, however, produces a virtually unnoticeable improvement in the overall precision of sample estimates.

Third, it can be seen that the absolute size of the sampling error is greatest around percentages of .5 and decreases as the percentage of a sample having a characteristic approaches either zero or 100%. We have seen that standard errors are related directly to variances. The variance $p(1 - p)$ is smaller as the percentages get further from .5. When $p = 0.5$, $(0.5 \times 0.5) = 0.25$. When $p = 0.2$, $(0.2 \times 0.8) = 0.16$.

Fourth, Table 3.1 and the equations on which it is based apply to samples drawn with simple random sampling procedures. Most samples of general populations are not simple random samples. The extent to which the particular sample design will affect calculations of sampling error varies from design to design and for different variables in the same survey. More often than not, Table 3.1 will constitute an underestimate of the sampling error for a general population sample.

Finally, it should be emphasized that the variability reflected in Table 3.1 describes potential for error that comes from the fact of sampling rather than collecting information about every individual in a population. The calculations do not include estimates of error from any other aspects of the survey process.

Effects of Other Sample Design Features

The preceding discussion describes the calculation of sampling errors for simple random samples. Estimates of sampling errors will be affected by different sampling procedures. Systematic sampling should produce sampling errors equivalent to simple random samples if there is no stratification. Stratified samples can produce sampling errors that are lower than those associated with simple random samples of the same size for variables that differ (on average) by stratum, if rates of selection are constant across strata.

Unequal rates of selection (selecting subgroups in the population at different rates) are designed to increase the precision of estimates for oversampled subgroups, thus

- (a) they generally will produce sampling errors for the whole sample that are higher than those associated with simple random samples of the same size, for variables that differ by stratum, except
- (b) when oversampling is targeted at strata that have higher than average variances for some variable, the overall sampling errors for those variables will be lower than for a simple random sample of the same size.

Clustering will tend to produce sampling errors that are higher than those associated with simple random samples of the same size for variables that are more homogeneous within clusters than in the population as a whole. Also, the larger the size of the cluster at the last stage, the larger the impact on sampling errors will usually be.

It often is not easy to anticipate the effects of design features on the precision of estimates. Design effects differ from study to study and for different variables in the same survey. To illustrate, suppose every house on various selected blocks was the same with respect to type of construction and whether or not it was occupied by the owner. Once one respondent on a block reports he is a home owner, the additional interviews on that block would yield absolutely no new information about the rate of home ownership in the population as a whole. For that reason, whether the researcher took one interview per block or 20 interviews per block, the reliability of that estimate would be exactly the same, basically proportionate to the number of blocks from which any interviews at all were taken. At the other extreme, the height of adults is likely to

vary as much within a block as it does throughout a city. If the respondents on a block are as heterogeneous as the population as a whole, clustering does not decrease the precision of estimates of height from a sample of a given size. Thus, one has to look at the nature of the clusters or strata and what estimates are to be made to evaluate the likely effect of clustering on sampling errors.

The effects of the sample design on sampling errors often are unappreciated. It is not uncommon to see reports of confidence intervals that assume simple random sampling when the design was clustered. It also is not a simple matter to anticipate the size of design effects beforehand. As noted, the effects of the sample design on sampling errors are different for every variable; their calculation is particularly complicated when a sample design has several deviations from simple random sampling, such as both clustering and stratification. Because the ability to calculate sampling errors is one of the principal strengths of the survey method, it is important that a statistician be involved in a survey with a complex sample design to ensure that sampling errors are calculated and reported appropriately. When surveys involve clustering, stratification or varying probabilities of selection, analysis modules, which exist in most standard analysis programs, to properly calculate standard errors for complex sample designs should be used (see Chapter 10).

Finally, the appropriateness of any sample design feature can be evaluated only in the context of the overall survey objectives. Clustered designs are likely to save money both in sampling (listing) and in data collection. Moreover, it is common to find many variables for which clustering does not inflate the sampling errors very much. Oversampling one or more groups often is a cost-effective design. As with most issues discussed in this book, the important point is for a researcher to be aware of the potential costs and benefits of the options and to weigh them in the context of all the design options and the main purposes of the survey.

HOW BIG SHOULD A SAMPLE BE?

Of the many issues involved in sample design, one of the most common questions posed to a survey methodologist is how big a survey sample should be. Before providing an approach to answering this question, perhaps it is appropriate to discuss three common but inappropriate ways of answering it.

One common misconception is that the adequacy of a sample depends heavily on the fraction of the population included in that sample—that somehow 1%, or 5%, or some other percentage of a population will make a sample credible. The estimates of sampling errors previously discussed do not take into account the fraction of a population included in a sample. The sampling error estimates from the preceding equations and from Table 3.1 can be reduced by multiplying them by the value $(1 - f)$, where f = the fraction of the population included in a sample.

When one is sampling 10% or more of a population, this adjustment can have a discernible effect on sampling error estimates. The vast majority of survey samples, however, involve very small fractions of populations. In such instances, small increments

in the fraction of the population included in a sample will have no effect on the ability of a researcher to generalize from a sample to a population.

The converse of this principle also should be noted. The size of the population from which a sample of a particular size is drawn has virtually no impact on how well that sample is likely to describe the population. A sample of 150 people will describe a population of 15,000 or 15 million with virtually the same degree of accuracy, assuming that all other aspects of the sample design and sampling procedures are the same. Compared to the total sample size and other design features, such as clustering, the impact of the fraction of a population sampled on sampling errors is typically trivial. It is most unusual for the fraction of the target population included in a sample to be an important consideration when deciding on a sample size.

A second inappropriate approach to deciding on sample size is somewhat easier to understand. Some people have been exposed to so-called standard survey studies, and from these they have derived a typical or appropriate sample size. Thus some people will say that good national survey samples generally are 1,500, or that good community samples are 500. Of course, it is reasonable to look at what other competent researchers have considered to be adequate sample sizes of a particular population. The sample size decision, however, like most other design decisions, must be made on a case-by-case basis, with the researchers considering the variety of goals to be achieved by a particular study and taking into account numerous other aspects of the research design.

A third wrong approach to deciding on sample size is the most important one to address, for it can be found in many statistical textbooks. The approach goes like this: A researcher should decide how much margin of error he or she can tolerate or how much precision is required of estimates. Once one knows the need for precision, one simply uses a table such as Table 3.1, or appropriate variations thereon, to calculate the sample size needed to achieve the desired level of precision.

In some theoretical sense, there is nothing wrong with this approach. In practice, however, it provides little help to most researchers trying to design real studies. First, it is unusual to base a sample size decision on the need for precision of a single estimate. Most survey studies are designed to make numerous estimates, and the needed precision for these estimates is likely to vary.

In addition, it is unusual for a researcher to be able to specify a desired level of precision in more than the most general way. It is only the exception, rather than the common situation, when a specific acceptable margin for error can be specified in advance. Even in the latter case, the previous approach implies that sampling error is the only or main source of error in a survey estimate. When a required level of precision from a sample survey is specified, it generally ignores the fact that there will be error from sources other than sampling. In such cases, the calculation of precision based on sampling error alone is an unrealistic oversimplification. Moreover, given fixed resources, increasing the sample size may even decrease precision by reducing resources devoted to response rates, question design, or the quality of data collection.

Estimates of sampling error, which are related to sample size, do play a role in analyses of how big a sample should be. This role, however, is complicated.

The first prerequisite for determining a sample size is an analysis plan. The key component of that analysis plan usually is not an estimate of confidence intervals for the overall sample, but rather an outline of the subgroups within the total population for which separate estimates are required, together with some estimates of the fraction of the population that will fall into those subgroups. Typically, the design process moves quickly to identifying the smaller groups within the population for which figures are needed. The researcher then estimates how large a sample will be required to provide a minimally adequate sample of these small subgroups. Most sample size decisions do not focus on estimates for the total population; rather, they are concentrated on the minimum sample sizes that can be tolerated for the smallest subgroups of importance.

The process then turns to Table 3.1, not at the high end but at the low end of the sample size continuum. Are 50 observations adequate? If one studies Table 3.1, it can be seen that precision increases rather steadily up to sample sizes of 150 to 200. After that point, there is a much more modest gain to increasing sample size.

Like most decisions relating to research design, there is seldom a definitive answer about how large a sample should be for any given study. There are many ways to increase the reliability of survey estimates. Increasing sample size is one of them. Even if one cannot say that there is a single right answer, however, it can be said that there are three approaches to deciding on sample size that are inadequate. Specifying a fraction of the population to be included in the sample is never the right way to decide on a sample size. Sampling errors primarily depend on sample size, not on the proportion of the population in a sample. Saying that a particular sample size is the usual or typical approach to studying a population also is virtually always the wrong approach. An analysis plan that addresses the study's goals is the critical first step. Finally, it is very rare that calculating a desired confidence interval for one variable for an entire population is the determining calculation in how big a sample should be.

SAMPLING ERROR AS A COMPONENT OF TOTAL SURVEY ERROR

The sampling process can affect the quality of survey estimates in three different ways:

- If the sample frame excludes some people whom we want to describe, sample estimates will be biased to the extent that those omitted differ from those included.
- If the sampling process is not probabilistic, the relationship between the sample and those sampled is problematic. One can argue for the credibility of a sample on grounds other than the sampling process; however, there is no statistical basis for saying a sample is representative of the sampled population unless the sampling process gives each person selected a known probability of selection.
- The size and design of a probability sample, together with the distribution of what is being estimated, determine the size of the sampling errors, that is, the chance variations that occur because of collecting data about only a sample of a population.

Often sampling errors are presented in ways that imply they are the only source of unreliability in survey estimates. For surveys that use large samples, other sources of error are likely to be more important. A main theme of this book is that nonsampling errors warrant as much attention as sampling errors. Also, it is not uncommon to see sampling errors reported that assume simple random sampling procedures when the sample design involved clusters, or even when it was not a probability sample at all. In these ways, ironically, estimates of sampling errors can mislead readers about the precision or accuracy of sample estimates.

Sampling and analyzing data from a sample can be fairly straightforward if a good list is used as a sampling frame, if a simple random or systematic sampling scheme is used, and if all respondents are selected at the same rate. With such a design, Table 3.1 and the equations on which it is based will provide good estimates of sampling errors. Even with such straightforward designs, however, researchers need to consider all sources of error, including the sample frame, nonresponse, and response errors (all discussed in subsequent chapters) when evaluating the precision of survey estimates. Moreover, when there are doubts about the best way to sample, or when there are deviations from simple random sampling, it is virtually essential to involve a sampling specialist both to design an appropriate sampling plan and to analyze results properly from a complex sample design.

EXERCISES

1. In order to grasp the meaning of sampling error, repeated systematic samples of the same size (with different random starts) can be drawn from the same list (e.g., a telephone directory). The proportions of those samples having some characteristic (e.g., a business listing) taken together will form a distribution. That distribution will have a standard deviation that is about one half the entry in Table 3.1 for samples of the sizes drawn. It is also valuable to calculate several of the entries in Table 3.1 (i.e., for various sample sizes and proportions) to help understand how the numbers were derived.
2. What percentage of adults in the United States would you estimate:
 - a. Have driver's licenses?
 - b. Have listed telephone numbers?
 - c. Are registered to vote?
 - d. Have a personal e-mail address (not through their work)?
 - e. Do not have a landline telephone?
3. What are some likely differences between those who would be in those sample frames and those who would not?
4. Give three examples where drawing a two-stage sample would probably be a good idea (or might be the only way to get a good sample).
5. Compared with simple random samples, do the following tend to increase, decrease, or have no effect on sampling errors?
 - a. Clustering
 - b. Stratifying
 - c. Using a systematic sampling approach

FURTHER READINGS

- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed., chaps. 3–4). New York: John Wiley.
- Kalton, G. (1983). *Introduction to survey sampling*. Beverly Hills, CA: Sage.
- Kish, L. (1965). *Survey sampling*. New York: John Wiley.
- Lohr, S. L. (1998). *Sampling design and analysis*. New York: Brooks/Cole.