CHAPTER 3

Probability and Sampling

William R. Nugent

Probability theory concerns the relative frequency with which certain events occur. Probability is important in sampling because it is the vehicle that allows the researcher to use the information in a sample to make inferences about the population from which the sample was obtained. The purpose of this chapter is to give an overview of probability and sampling. The goals of this chapter are (a) to give the reader a brief introduction to probability and the basis of sampling in probability, (b) to cover basic sampling methods with a focus on probability sampling, and (c) to illustrate these sampling methods by use of numerical examples that the reader can replicate.

Probability

Probability refers to the likelihood that an event will occur. The probability of an event is given by a number between 0 and 1, with probabilities closer to 0 indicating that the event is less likely and those closer to 1 indicating that it is more likely. A probability of 0 means that the event never will occur, whereas a probability of 1 means that it is certain to occur. The probability of Event A, symbolized as p(A), is defined as the ratio of the number of "favorable" outcomes (i.e., the number of outcomes that count as the specific Event A) to the total number of outcomes:

$$p(A) = (\text{number of favorable outcomes})/(\text{total number of possible outcomes}) \quad [1]$$

For example, consider the probability of selecting an ace in a single draw from a deck of 52 cards. There are four favorable outcomes: the ace of hearts, the ace of clubs, the ace of diamonds, and the ace of spades. There are a total of 52 possible outcomes, p(ace) = 4/52.

Now, suppose that we have a set of N "things" (N indicates that the number of things in the set is some arbitrary number) such as persons, objects, or phenomena. Probability theory tells us that if we want to select n of the things (where $n < N$) from the set of N things, then there are

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \qquad [2]$$

different combinations of n things that we can draw from the set of N things. The symbol $\binom{N}{n}$ is read as "N on n," and $N! = N \times [N-1] \times [N-2] \times \ldots \times 2 \times 1$ (where $N!$ is read as "N factorial"). For example, $5! = 5 \times 4 \times 3 \times 2 \times 1$. The terms $n!$ and $(N-n)!$ are similarly defined (Ash, 1993).

For example, consider a poker hand of five cards. Since there are 52 cards in a normal deck of cards, there will be

$$\binom{52}{5} = \frac{52!}{5!(52-5)!} = 2598960$$

possible poker hands of five cards. If each of these possible poker hands has the same probability of being drawn by a card player, then the probability of any given hand is

$$\frac{1}{\binom{52}{5}} = \frac{1}{2598960} = .000000385$$

Thus, for example, the probability of drawing the ace, king, queen, jack, and 10 of hearts—a royal flush of hearts—is .000000385. This is the same probability as drawing, say, the 2 of clubs, the 7 of spades, the ace of diamonds, the 9 of clubs, and the 10 of hearts.

Probability and Sampling

Now consider 50 social work students in a research class. Suppose that 7 of these students are drawn at "random." This means that each student in the class has the same probability of being selected into the sample, 1/50. There are, therefore,

$$\binom{50}{7} = 99884400$$

possible samples of 7 students drawn at random from the class of 50. Since each student has the same probability of being selected, the probability of any particular sample of 7 students being selected is

$$\frac{1}{\binom{50}{7}} = \frac{1}{99884400} = .00000001$$

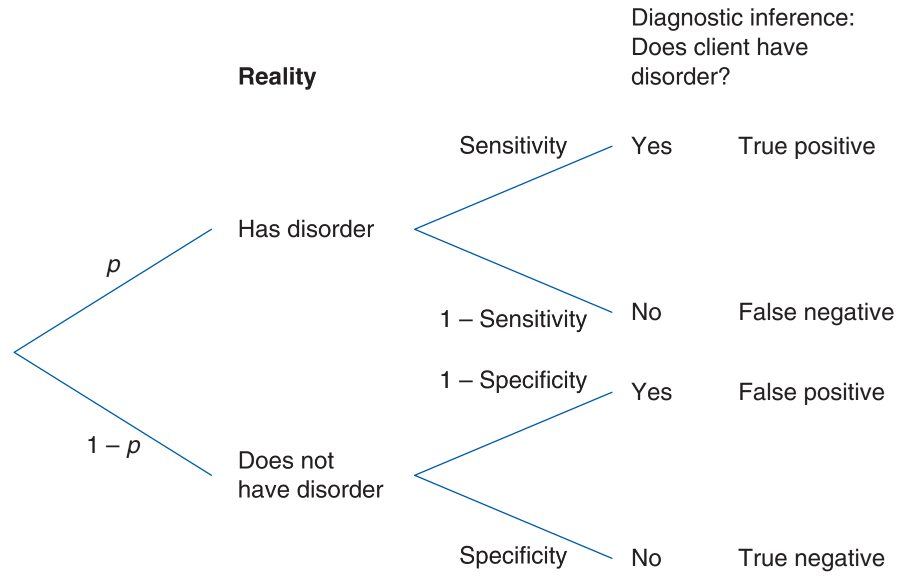
Any sample of size n drawn from a population in such a way that every possible sample of size n has the same probability of being selected is called a random sample (Ash, 1993; Scheaffer, Mendenhall, & Ott, 1996). Any sample that is selected in this manner is representative of the population in a probabilistic sense. Thus, probability theory allows the researcher to select samples that are representative of the population in this special way.

Other Applications of Probability

There are other very useful applications of probability for social workers. One important one concerns the making of a diagnosis, such as a medical diagnosis or a diagnosis in mental health, such as that made using the text-revised fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV-TR;* American Psychiatric Association, 2000). Consider the tree diagram in Figure 3.1 (Ash, 1993). The branch to the left divides a population of persons into two groups: those who have some disease or disorder (upper branch) and those who do not (lower branch). This branching represents reality. The people in the two groups either actually have the disorder or actually do not. The letter p by the left upper branch represents the prevalence of the disease or disorder in the population (Nugent, 2004, 2005). The prevalence is the proportion (or percentage) of persons in the population who have the disease or disorder. The $1 - p$ (one minus p) by the lower left branch is, in a sense, the "prevalence" of "not having the disorder" in the population.

Now people do not walk around with LED readouts on their foreheads telling whether they have the disease or disorder. Rather, this is hidden from a practitioner, who must use some assessment and diagnostic procedure to infer whether the person has the disease or disorder. The branches to the right in Figure 3.1 show the possible resulting inferences that can be made by a practitioner about whether a person has the disease or disorder. The upper right branches, which represent the inferential results for persons who in reality have the disease or disorder, show two inferential outcomes: (1) A person is inferred to have the disease or disorder, and in reality he or she does indeed have the disease or disorder. This is called a *true positive.* (2) A person is inferred to *not* have the disease or disorder when in fact he or she does. This is called a *false negative.* The term *sensitivity* next to the upper branch on the right of this figure refers to the *true-positive rate,* which is the proportion of persons who have the disorder who are accurately inferred as having the disorder on the basis of the results of the assessment and diagnostic procedure. The sensitivity can be thought of as an indicator of how well the assessment and diagnostic procedure functions with persons who have the disease or disorder. The term 1 – sensitivity (one minus sensitivity) by the lower branch in the upper right of Figure 3.1 indicates the *false-negative rate,* which is the proportion of persons who have the disease or disorder who are incorrectly inferred as *not* having the disease or disorder.

Figure 3.1
Two-Stage Model of Diagnostic Inference Making



The two lower branches in the right side of Figure 3.1 show the inferential outcomes for persons who actually do not have the disease or disorder. The lower branch in this portion of the figure shows the inferential results for a person who does not have the disease or disorder and is judged as not having the disease or disorder, an outcome called a *true negative.* The term *specificity* by this branch refers to the *true-negative rate,* which is the proportion of persons who do not have the disease or disorder who are correctly inferred as not having the disease or disorder. The specificity is another index describing how well the assessment and diagnostic procedure functions with persons who do not have the disease or disorder.

The upper branch in this portion of the figure shows the result that persons who do not have the disorder are inferred as having it. This outcome is called a *false positive,* and the 1 – specificity (one minus specificity) refers to the *false-positive rate,* which is the proportion of persons who do not have the disease or disorder who are incorrectly inferred as having the disease or disorder on the basis of the results of the assessment and diagnostic procedure.

These concepts can be used to define an important notion: the *positive predictive value, or the PPV.* The positive predictive value is the probability that a person *who has been diagnosed as having the disease or disorder actually has the disease or disorder.* This can be thought of as an index describing the accuracy of a diagnostic inference based on an assessment and diagnostic procedure. The PPV is a number that can range from 0 to 1 and is in fact a conditional probability; it is the probability of one outcome—in this case, that a person actually has the disease or disorder—given that another event has occurred—in this case, that the person has been diagnosed as having the disease or disorder on the basis of the results of the assessment and diagnostic procedure. The closer the PPV is to 1, the more likely it is that the person who has been diagnosed as having the disease or disorder actually has the disease or disorder, and the closer the PPV is to 0, the less likely it is that the person who has been diagnosed as having the disease or disorder actually has the disease or disorder.

The PPV can be computed from the terms in Figure 3.1 by the formula

$$\text{PPV} = \frac{\text{prevalence} \times \text{sensitivity}}{(\text{prevalence} \times \text{sensitivity}) + [(1 - \text{prevalence}) \times (1 - \text{sensitivity})]} \qquad [3]$$

All of the terms in this formula are shown in Figure 3.1. The prevalence of the disease or disorder in the population of interest, as well as the sensitivity and specificity values associated with the assessment and diagnostic procedure, must be obtained from research results that can be found in journal articles, book chapters, technical reports, scientific Web sites, and other sources (Nugent, 2004, 2005).

Now let's illustrate the notion of the PPV with an example from mental health. One of the most commonly used diagnostic systems is the *Diagnostic and Statistical Manual of Mental Disorders,* the most recent version of which is the *DSM-IV-TR* (American Psychiatric Association, 2000). There is a structured interview protocol, the Diagnostic Interview Schedule (DIS; Robins, Helzer, Croughan, & Ratcliff, 1981), designed to make reliable *DSM* diagnoses. Murphy, Monson, Laird, Sobol, and Leighton (2000) recently conducted research on the DIS and estimated the sensitivity of this interview protocol for detecting major depressive disorder (MDD) to be .55 and the specificity to be .90. Other research has suggested that the prevalence of MDD among adolescents is about .034 (i.e., 3.4%; Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993).

Now assume that the prevalence of MDD among adolescents seeking services is higher than in the general population of adolescents, a plausible presumption. Let's assume for sake of illustration that the prevalence of MDD is actually three times higher among adolescents actually seeking services or being brought in for services by parents or other authority figures than the prevalence of depression in the general adolescent population. Thus, the prevalence is $3 \times .034 = .102$, or 10.2%. If the DIS is used to make inferences about whether an adolescent has MDD, the PPV of the resulting diagnosis is

$$\text{PPV} = \frac{.102 \times .55}{(.102 \times .55) + [(1 - .102)(1 - .90)]} = .38$$

This number can be interpreted as telling the practitioner that a given adolescent who has been diagnosed as having MDD using the DIS has a probability of only .38 of actually having MDD. It can also be interpreted saying that only 38% of adolescents diagnosed as having MDD on the basis of the DIS will actually have this disorder.

The PPV and related indices are very important for practitioners to understand and be able to use and interpret. The reader is referred to Nugent (2004, 2005) for more in-depth discussion. There are also several Web sites at the end of this chapter that the reader can go to for more about these concepts.

Sampling

The term *sampling* refers to the methods that researchers use to select the groups of persons, objects, or phenomena that they actually observe. The very large set of persons, objects, or phenomena about which researchers wish to learn is called the *population,* and the individual persons, objects, or phenomena are referred to as the population *elements.* The group of persons, objects, or phenomena that they select from the population and observe is referred to as the *sample.* Most of the time, researchers wish to use the sample to make inferences about the population. *Sampling, then, concerns the methods used to obtain the samples of persons, objects, or phenomena from the population about which we wish to make inferences.*

Probability Samples

A very important type of sample is called a *probability sample.* Probability sampling is done in such a manner that we can compute probabilities for specific samples being obtained. A basic principle of probability sampling is that a sample will be representative of a population *if each member of a population has the same probability of appearing in the sample* (Rubin & Babbie, 1997). Probability samples have at least two advantages. First, they are unbiased and representative of the population in a probabilistic sense. Second, we can estimate the amount of error involved in using the statistics we get from the sample as estimates of the values we would obtain if we observed the entire population. These population values are referred to as the *population parameters.* The more commonly used probability sampling methods are the *random sampling* methods, consisting of simple random sampling, stratified random sampling, and systematic sampling, among others.

Simple Random Sampling

A *simple random sample* of size n is defined as a sample obtained in such a manner that *every possible sample of size* n *has the same probability of being selected.* This sample is unbiased in that no population element or sample of size n has a greater or lesser probability of being selected than does any other element or sample of size n. A simple random sample of size n is obtained in the following manner. First, a list that identifies every element (e.g., a list of names) in the population is created. This exhaustive list is called the *sampling frame.* Each population element is given a numeric identifier, and then a random number table or computer is used to generate a list of n random numbers. These random numbers then are used to select the population elements that will be in the sample (Scheaffer et al., 1996).

Consider the population of scores shown in Table 3.1. Suppose that the numbers in Table 3.1 represent the attitudes of the persons constituting the population toward a proposed social policy (1 = *completely opposed,* 2 = *strongly opposed,* 3 = *moderately opposed,* 4 = *a little opposed,* 5 = *neither opposed nor in favor,* 6 = *a little in favor,* 7 = *moderately in favor,* 8 = *strongly in favor,* 9 = *completely in favor*). There are 900 scores in this population (for now, ignore the stratification). The mean of this population of scores is 5.000, and the variance is 6.667. Let us imagine that we want a simple random sample of $n = 15$ scores from this population. We would assign a three-digit numeric identifier to each score in Table 3.1. Because there are 900 scores in this population, we will need numeric identifiers that are three digits long. Thus, we would label the score in the first row of the first column in Table 3.1 as 001, the score in the second row of the first column as 002, and so on down the first column to the final score, which would be labeled 025. The score in the first row of the second column would be labeled 026, the score in the second row of the second column would be labeled 027, and so on. This numeric labeling would continue through the final (36th) column. The score in the first row of the final column would be labeled 876, the score in the second row of the final column would be labeled 877, and so on to the score in the final row of this column, which is labeled 900.

We then would use a random number table, such as that found in Scheaffer et al. (1996) or Rubin and Babbie (1997), and select 15 different random numbers that are three digits long. Then, we would find the 15 scores in Table 3.1 with the numeric identifiers that match the random numbers. The 15 elements selected in this manner would constitute our simple random sample.

Once we have obtained our simple random sample, we would compute a sample statistic that we wish to use as an estimate of a population parameter. One such sample statistic is the sample mean, $\bar{y}$, which serves as our estimate of the population parameter μ (i.e., the population mean),

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad [4]$$

where y_i is the score for the ith person, and n is the sample size. We also compute the *error bound* (or sampling error) associated with the use of the sample statistic as an estimate of the population parameter. For the mean, we would compute the error bound, B,

$$B = \pm 2\sqrt{\frac{s^2}{n}\left(\frac{N-n}{N}\right)} \qquad [5]$$

where s^2 is the sample estimate of the population variance,

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad [6]$$

Table 3.1 Population of Scores

Stratum 1	Stratum 2	Stratum 3
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999
111122223333	444455556666	777788889999

and where N is the number of elements in the population (Scheaffer et al., 1996). The term $(N - n)/N$ in Equation (5) and in later equations is called the finite population correction factor (FPC). Usually, if the FPC is equal to or greater than about .95, then it is dropped from the equation for the error bound. We can use the error bound to construct an approximate 95% confidence interval for the population mean by adding and subtracting B from the sample mean:

$$\bar{Y} \pm B \qquad [7]$$

About 95% of all confidence intervals created using this method will encompass the true population mean (Scheaffer et al., 1996).

Let us obtain a simple random sample of $n = 15$ scores from the population in Table 3.1. The random number table in Scheaffer et al. (1996) is used to obtain the following list of 15 random numbers: 104, 223, 241, 421, 375, 779, 895, 854, 289, 635, 094, 103, 071, 510, and 023. These 15 random numbers then are used to select the following scores from Table 3.1: 2, 3, 3, 5, 4, 8, 9, 9, 3, 7, 1, 2, 1, 6, and 1. The estimate of the population mean based on this simple random sample is

$$\hat{\mu} = [(2 + 3 + 3 + 5 + 4 + 8 + 9 + 9 + 3 + 7 + 1 + 2 + 1 + 6 + 1)/15] = 64/15 = 4.3$$

The sample estimate of the population variance is 8.3524. For this simple random sample, $(N - n)/N = (900 - 15)/900 = .98 > .95$, so the FPC can be dropped, and the error bound is

$$B = 2\sqrt{\frac{8.3524}{15}} = 1.5$$

and so the approximate 95% confidence interval for the population mean is

$$4.3 \pm 1.5$$

or 2.8 to 5.8.

A special case of the above is when the variable to be estimated is a proportion, such as the proportion of persons who are likely to be voting for a particular candidate in an election. In this case, the error bound will be

$$B = \pm 2\sqrt{\frac{p(1-p)}{n-1}\left(\frac{N-n}{N}\right)} \qquad [8]$$

where p is the sample proportion who are likely to be voting for the candidate of interest. For example, suppose that in a simple random sample of 200 voters, 35%, or .35, claim to be going to vote for Candidate X. The sample estimate of the population proportion of voters who will be voting for Candidate X is .35. Since the total population size is very large in this case (such as the population of a particular state), the FPC will be dropped, and the error bound will be

$$B = \pm 2\sqrt{\frac{.35(1-.35)}{200-1}} = \pm .067,$$

so the approximate 95% confidence interval is $.35 \pm .067$ or, in percentage terms, 28.3% to 41.7%.

Stratified Random Sampling

A second form of random sampling is called *stratified random sampling.* This method allows us to use information that we already have about the population. For example, suppose that it is known that the population of persons whose attitude scores are shown in Table 3.1 can be separated into three strata: the 300 persons in Stratum 1 who possess low levels of some variable y (e.g., level of education), the 300 persons in Stratum 2 who possess moderate levels of y, and the 300 persons in Stratum 3 who possess high levels of y. A stratified random sample of size n would be obtained from the population in Table 3.1 by gathering simple random samples of size n_1, n_2, and n_3 from Strata 1, 2, and 3, respectively. The combined sample of size $n_1 + n_2 + n_3 = n$ elements would comprise the stratified random sample. Nugent and Paddock (1996) provide an example of stratified random sampling in social work research.

Once a stratified random sample has been obtained, the stratified random sample estimate of the total population mean will be given by

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{L} N_i \bar{y}_i \qquad [9]$$

where N is the total population size, N_i is the size of the ith stratum, and there are L strata. The error bound will be given by

$$B = \pm 2\sqrt{\frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \left(\frac{N_i - n_i}{N_i}\right)\frac{s_i^2}{n_i}} \qquad [10]$$

where n_i is the sample size from stratum i, and s_i^2 is the estimated variance of scores in stratum i. Again, if the FPC is equal to or greater than .95, then it can be dropped.

Let us select a stratified random sample of $n = 15$ scores, with 5 scores randomly selected from each stratum, from the population in Table 3.1. The random number table in Scheaffer et al. (1996) is used to obtain simple random samples of $n = 5$ scores from each stratum. The random numbers used to select the sample from Stratum 1 are 191, 196, 069, 210, and 114, and the scores 2, 2, 1, 3, and 2 are obtained. The estimated mean score in Stratum 1 is [(2 + 2 + 1 + 3 + 2)/5] = 10/5 = 2.0 (variance = 0.5). The random numbers obtained for selecting the simple random sample from Stratum 2 are 396, 443, 425, 337, and 306, and the scores 4, 5, 5, 4, and 4 are obtained. The estimated mean score in Stratum 2 is [(4 + 5 + 5 + 4 + 4)/5] = 22/5 = 4.4 (variance = 0.3). The random numbers obtained for selecting the simple random sample from Stratum 3 are 649, 781, 749, 898, and 696, and the scores 7, 8, 8, 9, and 7 are obtained. The estimated mean score in Stratum 3 is [(7 + 8 + 8 + 9 + 7)/5] = 39/5 = 7.8 (variance = 0.7).

The estimated population mean will be

$$\frac{1}{900}[(300 \times 2) + (300 \times 4.4) + (300 \times 7.8)] = 4.73$$

and the error bound will be, since the FPC factor is greater than .95 for each stratum,

$$B = \pm 2\sqrt{\frac{1}{900^2}\left[\left(300^2 \times \frac{.5}{5}\right) + \left(300^2 \times \frac{.3}{5}\right) + \left(300^2 \times \frac{.7}{5}\right)\right]} = \pm 0.36$$

Thus, the 95% confidence interval for the population mean will be

$$4.73 \pm 0.36$$

or 4.37 to 5.09.

Notice how much narrower this confidence interval is compared to the confidence interval from the simple random sample. This shows how much more efficient stratified random sampling can be relative to simple random sampling *if the within-stratum variances are less than the between-strata variance,* as is the case for the population in Table 3.1. The variance of scores within each stratum is .667, whereas the between-strata variance is 6.000. Stratified random samples also are useful if estimates of stratum parameters, as well as estimates of the overall population parameters, are desired (Scheaffer et al., 1996).

Systematic Sampling

Systematic sampling sometimes is used in lieu of simple random sampling. First, we list each element in the population and assign the elements numeric identifiers, which will range in value from 1 to n. We then form the ratio $k = N/n$, where N = population size and n = sample size. The number k is the sampling interval (Rubin & Babbie, 1997). Next, we use a random number table and find a single random number between 1 and k and then find the element in the first k elements with the numeric identifier that matches this random number. Starting with this element, we go through the population and select every kth element until we have a sample of size n. A sample obtained like this is called a systematic sample. A good example of systematic sampling in social work research is provided by Glisson (1994).

Let us obtain a systematic sample of size $n = 15$ from the population in Table 3.1.

First, we compute the sampling interval k. Because there are 900 elements in this population,

$$k = N/n = 900/15 = 60 \qquad [11]$$

We then use the random number table in Scheaffer et al. (1996) to find a two-digit random number between 1 and k—in this case, between 1 and 60. The random number we find is 10, so we select the 10th element in the population and every 60th element thereafter. Thus, the second element to be selected has the numeric identifier 10 + 60 = 70, the third element has the numeric identifier 70 + 60 = 130, the fourth element has the numeric identifier 130 + 60 = 190, and so on. The 15 scores obtained in this manner are 1, 1, 2, 2, 3, 4, 4, 5, 5, 6, 7, 7, 8, 8, and 9. We then can use Equation (4) to compute our sample estimate of the population mean,

$$\hat{\mu} = [(1 + 1 + 2 + 2 + 3 + 4 + 4 + 5 + 5 + 6 + 7 + 7 + 8 + 8 + 9)/15] = 72/15 = 4.8$$

One limitation with systematic sampling is that we can use Equation (5) to estimate the error bound *only* if the elements in the population are in random order. Such a population is called a *random population* (Scheaffer et al., 1996). If the elements in the population are *ordered* with respect to magnitude, then Equation (5) will overestimate the error bound. If the elements are *periodic,* showing some form of cyclic variation, then

Equation (5) will underestimate the error bound (Scheaffer et al., 1996). Unless we know how the elements of the population are ordered in the sampling frame, we cannot use the formulae for error bounds to estimate the sampling error.

Other Methods of Random Sampling

There are other random sampling methods, most notably single-stage and multistage cluster sampling. Cluster sampling is useful because in many circumstances, it is logistically more practical than the random sampling techniques discussed previously. Scheaffer et al. (1996) discuss cluster sampling in some depth. A good example of the use of cluster sampling in social work research is provided by White (1998).

Nonprobability Sampling

Probability samples can be contrasted with *nonprobability* samples. When using nonprobability sampling methods, we cannot estimate the probabilities associated with selection of different samples of size *n*. This means that any of a plethora of biasing factors may be operative, leading some samples to have greater (or lesser) probabilities of being selected. The major consequences of this are that (a) the sample will most likely be biased in unknown ways, and (b) we cannot estimate the error involved in our use of sample statistics as estimates of population parameters. Thus, inferences we make from nonprobability samples are very risky.

Perhaps the most commonly used nonprobability sampling method is convenience (or accidental) sampling (Cook & Campbell, 1979; Scheaffer et al., 1996). In convenience sampling, the researcher uses whomever he or she can find who meets the eligibility criteria for being involved in the research and who agrees to participate. The greatest advantage of convenience sampling is that it is easy to use. However, a convenience sample is not representative of the population, and if population estimates are made from the sample, then the researcher (or anyone else) is likely to reach erroneous conclusions (Scheaffer et al., 1996). Convenience samples can be very useful, however, in program evaluation and treatment outcome studies. The program evaluation done by Sprang (1997) makes good use of a convenience sample.

Convenience sampling can be illustrated as follows. Suppose that a politician, Generic Joe, represents the population of persons whose attitude scores are shown in Table 3.1. Generic Joe tells a television reporter, "I have received 108 phone calls from my constituents, and the overall attitude in my district toward the proposed policy is one of 'strongly in favor.' This tells me that I absolutely must support this proposal." Let us suppose that the 108 calls that Generic Joe has received all have been made by the persons whose attitude scores are in the bold font in the third stratum of Table 3.1. As can be verified from these scores, the mean attitude toward the new policy held by the persons who called is 8.0, indicating that they strongly favor it (compare this estimate to the random sample estimates and to the true population mean of 5.0). Generic Joe is using this sample of convenience as if it is representative of his district when in fact it is not. Thus, the inferences that Generic Joe is making are incorrect.

There are other forms of nonprobability sampling, including purposive sampling, sampling for heterogeneity, impressionistic modal sampling, quota sampling, and snowball sampling. These methods are discussed in Cook and Campbell (1979) and Scheaffer et al. (1996).

Sampling and External Validity

External validity concerns the extent to which research results can be generalized, and sampling is a critical issue in establishing the external validity of research results (Bracht & Glass, 1968; Cook & Campbell, 1979). Cook and Campbell (1979) make the important distinction between generalizing to a specific well-defined population and generalizing across specific subgroups of a population. Generalizing to a population involves making inferences about overall population parameters without any concern about parameter values for specific subgroups within the population. An example of this situation would be the researcher who is interested in the "typical" (i.e., mean) response to some treatment in a well-defined population of persons but who is not interested in knowing which subgroups within the population benefit, which are unaffected, and which get worse. By contrast, generalizing across specific subgroups involves making inferences about parameter values for specific subgroups without any concern about overall population parameters. An example of this situation would be the researcher who is interested only in showing that the treatment is beneficial for males of a certain age range in the population and is not interested in the typical effect of the treatment in the population. Random sampling is critical for generalizing to a well-defined population. However, it is not necessary for generalizing across subgroups. The replication of results across multiple nonprobability samples may, in fact, form a more sound basis for generalizing across subgroups than do results based on a large random sample. The reader is referred to Cook and Campbell (1979) and Johnston and Pennypacker (1993) for in-depth discussions of these issues.

Conclusion

The reader is encouraged to work through (and replicate) the examples given in this chapter and to select different random samples from the population in Figure 3.1 and compute sample estimates of the population mean as well as the associated error bounds. This practice can help the reader to develop a deeper understanding of the sampling methods presented in the chapter.

Probability and sampling are broad and complex topics, and this chapter only skimmed the surface of these important subjects. The reader is referred to the references cited for more detailed presentations. The text by Scheaffer et al. (1996) is recommended for a study of sampling theory, whereas Ash (1993) is an excellent (and understandable) treatment of probability theory.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders DSM-IV-TR*. Washington, DC: American Psychiatric Association.

Ash, C. (1993). *The probability tutoring book.* New York: IEEE Press.

Bracht, G., & Glass, G. (1968). The external validity of experiments. *American Educational Research Journal, 5,* 437–474.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Glisson, C. (1994). The effect of services coordination teams on outcomes for children in state custody. *Administration in Social Work, 18,* 1–23.

Johnston, J., & Pennypacker, H. (1993). *Strategies and tactics of human behavioral research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-R disorders in high school students. *Journal of Abnormal Psychology, 102,* 133–144.

Murphy, J., Monson, N., Laird, A., Sobol, A., & Leighton, A. (2000). A comparison of diagnostic interviews for depression in the Stirling county study. *Archives of General Psychiatry, 57,* 230–236.

Nugent, W. (2004). The role of prevalence rates, sensitivity, and specificity on assessment accuracy: Rolling the dice in social work process. *Journal of Social Service Research, 31*(2), 51–75.

Nugent, W. (2005). The probabilistic nature of diagnosis: Rolling the dice in social work practice. In S. Kirk (Ed.), *Mental disorders in the social environment* (pp. 96–119). New York: Columbia University Press.

Nugent, W., & Paddock, J. (1996). Evaluating the effects of a victim-offender reconciliation program on reoffense. *Research on Social Work Practice, 6,* 155–178.

Robins, L., Helzer, J., Croughan, J., & Ratcliff, K. (1981). National Institute of Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry, 38,* 381–389.

Rubin, A., & Babbie, E. (1997). *Research methods for social work* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Scheaffer, R., Mendenhall, W., & Ott, R. (1996). *Elementary survey sampling* (5th ed.). New York: Duxbury.

Sprang, G. (1997). Victim impact panels: An examination of the effectiveness of this program on lowering recidivism and changing offenders' attitudes about drinking and driving. *Journal of Social Service Research, 22*(3), 73–84.

White, L. (1998). *An examination of quality of worklife in public service organizations.* Unpublished doctoral dissertation, University of Tennessee, Knoxville.

USEFUL WEB SITES

http://www.rapid-diagnostics.org/accuracy.htm

Provides a further discussion of sensitivity, specificity, PPV, and negative predictive value (NPV).

http://uhavax.hartford.edu/bugl/treat.htm

A Web site that illustrates the use of sampling and probability theory as it relates to HIV testing and treatment.

http://www.statistics.com

One of the leading providers of online training in the field of statistics, including a number of introductory courses.

DISCUSSION QUESTIONS

1. Why is it important to try and obtain a genuinely random sample of research participants?
2. When is it justifiable to not use random sampling methods in your research?
3. Describe two realistic methods of obtaining a probability sample.
4. Look up a recent article published in a social work journal, and try and determine what type of sampling method was used in that study.