APPLIED SOCIAL RESEARCH METHODS SERIES

Series Editors

1. **SURVEY RESEARCH METHODS (Second Edition)** by FLOYD J. FOWLER, Jr.
2. **SYNTHESIZING RESEARCH (Third Edition)** by HARRIS M. COOPER
3. **METHODS FOR POLICY RESEARCH** by ANN MAJCHRZAK
4. **SECONDARY RESEARCH (Second Edition)** by DAVID W. STEWART and MICHAEL A. KAMINS
5. **CASE STUDY RESEARCH (Second Edition)** by ROBERT K. YIN
6. **META-ANALYTIC PROCEDURES FOR SOCIAL RESEARCH (Revised Edition)** by ROBERT ROSENTHAL
7. **TELEPHONE SURVEY METHODS (Second Edition)** by PAUL J. LAVRAKAS
8. **DIAGNOSING ORGANIZATIONS (Second Edition)** by MICHAEL I. HARRISON
9. **GROUP TECHNIQUES FOR IDEA BUILDING (Second Edition)** by CARL M. MOORE
10. **NEED ANALYSIS** by JACK McKILLIP
11. **LINKING AUDITING AND METAEVALUATION** by THOMAS A. SCHWANDT and EDWARD S. HALPERN
12. **ETHICS AND VALUES IN APPLIED SOCIAL RESEARCH** by ALLAN J. KIMMEL
13. **ON TIME AND METHOD** by JANICE R. KELLY and JOSEPH E. McGRATH
14. **RESEARCH IN HEALTH CARE SETTINGS** by KATHLEEN E. GRADY and BARBARA STRUDLER WALLSTON
15. **PARTICIPANT OBSERVATION** by DANNY L. JORGENSEN
16. **INTERPRETIVE INTERACTIONISM** by NORMAN K. DENZIN
17. **ETHNOGRAPHY (Second Edition)** by DAVID M. FETTERMAN
18. **STANDARDIZED SURVEY INTERVIEWING** by FLOYD J. FOWLER, Jr. and THOMAS W. MANGIONE
19. **PRODUCTIVITY MEASUREMENT** by ROBERT O. BRINKERHOFF and DENNIS E. DRESSLER
20. **FOCUS GROUPS** by DAVID W. STEWART and PREM N. SHAMDASANI
21. **PRACTICAL SAMPLING** by GARY T. HENRY
22. **DECISION RESEARCH** by JOHN S. CARROLL and ERIC J. JOHNSON
23. **RESEARCH WITH HISPANIC POPULATIONS** by GERARDO MARIN and BARBARA VANOSS MARIN
24. **INTERNAL EVALUATION** by ARNOLD J. LOVE
25. **COMPUTER SIMULATION APPLICATIONS** by MARCIA LYNN WHICKER and LEE SIGELMAN
26. **SCALE DEVELOPMENT** by ROBERT F. DeVELLIS
27. **STUDYING FAMILIES** by ANNE P. COPELAND and KATHLEEN M. WHITE
28. **EVENT HISTORY ANALYSIS** by KAZUO YAMAGUCHI
29. **RESEARCH IN EDUCATIONAL SETTINGS** by GEOFFREY MARUYAMA and STANLEY DENO
30. **RESEARCHING PERSONS WITH MENTAL ILLNESS** by ROSALIND J. DWORKIN
31. **PLANNING ETHICALLY RESPONSIBLE RESEARCH** by JOAN E. SIEBER

Other volumes in this series are listed at the back of the book

PRACTICAL SAMPLING

Gary T. Henry

Applied Social Research Methods Series
Volume 21

SAGE Publications
International Educational and Professional Publisher
Newbury Park London New Delhi

decreases. Stratification can be added at any or all stages, which generally improves precision, but adds cost and complexity. A total of five stages is utilized in the National Household Survey conducted by the Survey Research Center at the University of Michigan. The stages and the stratifications are described in Chapter 4.

CONCLUSION

Sampling approaches fall into two categories, probability and nonprobability. Choosing between these two types of approaches is a matter of weighing the requirements for validity and credibility against a realistic assessment of the requirements for timeliness and effort of the alternative approaches. Probability sampling, carefully designed and carried out, has greater validity and credibility than nonprobability sampling. Often, although not always, the costs and time required to conduct a probability sample are greater than for a nonprobability sample.

The remainder of this book will deal exclusively with probability sampling. While nonprobability samples are useful in some situations, an underlying premise of this book is that probability samples are clearly the preferred alternative, and only in cases where probability samples cannot be used are nonprobability samples viable.

The next chapter brings the rationale behind this premise clearly into focus. The bias and likely error stemming from the use of a probability sample can be rigorously examined and estimated. No comparable examination and estimation can be conducted for nonprobability samples. Thus, the range of uncertainty stemming from the use of a sample can be estimated with a specified degree of confidence for probability samples, but not for nonprobability samples. Sampling theory, well developed and tested for probability sampling, accounts for this difference.

Chapter 3 and, in more detail, the subsequent chapters provide the researcher with the practical sample design approach, a framework to be used in designing and executing probability sampling. The framework and examples should be useful in helping the researcher develop a practical design given the study goals and the resources available. Understanding these tools is especially important to produce a sample with the best possible fit for the study.

3

Practical Sample Design

Practical sample design seeks to produce valid and credible sample data and statistics that match the precision needed for the study. Practical sample design is an approach that integrates sampling design and execution into the overall research process using the concept of total error for the assessment of validity, credibility, and precision. The validity of the data affects the accuracy of inferring sample results to the population and drawing correct conclusions from the tests of hypotheses. Credibility, in large measure, rests on the sample selection process. Using random selection removes subjective judgment from the selection of the sample and enhances credibility.

The concern for precision arises as an intrinsic consequence of sampling. Selecting a subset of the population means that some members of the population are not included in a sample. Repeated selection of samples using the same procedure will yield different results because of this. Fortunately, sampling theory provides us with much useful information about the fluctuation of sample results. Using sampling theory, the probable amount of fluctuation or sampling variability can be calculated. The relationship between sampling variability and precision is well established: precision of a sample statistic decreases as sampling variability increases. Sampling theory provides the researcher with knowledge of the factors that contribute to variability and, therefore, ways to reduce the variability to obtain an acceptable level of precision.

To begin to understand how choices should be made in the design process and to use the practical sample design approach, the sources of threats to validity and sampling variability must first be understood. The threats to external validity, described in Chapter 1, are usually known as bias in the sampling context. Bias in sample selection causes systematic differences between the sample and the population that the sample represents. Sources of bias mentioned in the previous chapter are differences between the target population and the study population and sampling methods that cause some subpopulations to be overrepresented in the sample.

Threats to statistical conclusion validity usually manifest themselves as increases in sampling variability. Increasing sampling variability does not cause systematic differences between the population and the sample, but it affects the precision of the estimates and may retard the ability to reach accurate conclusions. Bias and sampling variability together represent total error for the sample. Total error is systematically decomposed and analyzed in the next section. The final section of this chapter presents the practical sample design framework, which provides guidance for making sampling choices that reduce, to the extent possible, bias and sampling variability.

SOURCES OF ERROR IN SAMPLING DESIGN

The goal of practical sampling design can be achieved by minimizing the amount of total error in the sample selection to an acceptable level given the purpose and resources available for the research. Total error has three distinct components:

Nonsampling bias: systematic error due to differences in population definition or measurement error, for example.

Sampling bias: systematic error that results from a sampling approach that over-represents a portion of the study population.

Sampling variability: the fluctuation of sample estimates around the study population parameters that results from the random selection process.

Each component generates concerns for the researcher. However, only the final two components are generally considered in the domain of the sampling design. Frequently, discussions of error in samples dwell exclusively on sampling variability and its estimate, the standard error or sampling error. All three sources of error should be explicitly considered for effective sampling designs.

Total error is defined as the difference between the true population value for the target population and the estimate based on the sample data. Total error for the mean is:

$$E = \overline{X}_T - \overline{x}$$

where X_T is the true population value, and $\bar{x}$ is the sample mean.

Because total error affects the extent to which the study objectives are met, practical sample design must take all three components into account. Each of the three components of total error, the point in the research process they can arise, and some examples of the sources of each are graphically illustrated in Figure 3.1. Furthermore, sample design takes place under

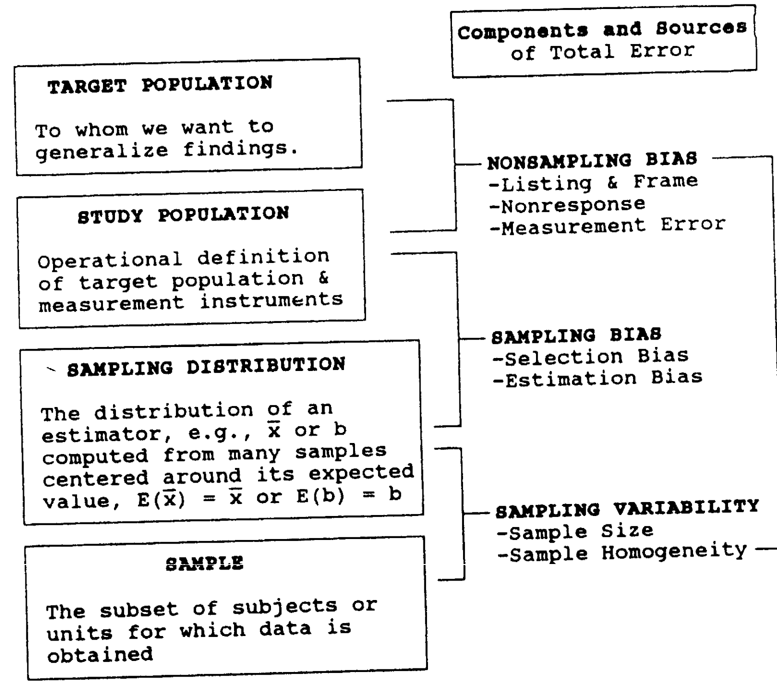

Figure 3.1. Practical Sampling Design

resource constraints. Decisions that allocate resources to reduce error from one component necessarily affect the resources available for reducing error from the other two components. Trade-offs are essential in practical sampling design. Limited resources, which are the case in practical design situations, force the researcher into making trade-offs between reducing the components of total error. The researcher must be fully aware of the three components of error to make decisions based on the trade-offs to be considered in reducing total error.

Nonsampling bias. Nonsampling bias is the difference between the true target population value and the population value obtained if the data collection operations were carried out on the entire population. Nonsampling bias results from explicit decisions and implementation of decisions during data collection efforts that are not directly related to the selection of the sample. For example, the definition of the study population may exclude some members of the target population that the researcher would like to include in the study findings. Even if data were collected on the entire *study* population in this case, the findings would be biased because of the exclusion of some *target* population members. The formula for nonsampling bias using the mean as an example of the population value is:

$$\text{NSB} = \bar{X}_T - \bar{X}_S$$

where $\bar{X}_T$ is the true mean of the target population, and
$\bar{X}_S$ is the mean obtained from the study population.

Differences in the true mean of the population and the survey population mean arise from several sources. The principal difference relevant to sample design is the difference between the target population and the study population. The target population is the group about which the researcher would like to make statements. The target population can be defined based on conditions and concerns that arise from the theory being tested or by concerns generated from the policy being examined. For instance, in a comprehensive needs assessment for the homeless, the target population should include all homeless, whether served by current programs or not. On the other hand, an evaluation of the community mental health services provided to the homeless should include only homeless recipients of community mental health care. The target population for the needs assessment is more broadly defined and inclusive of all homeless, in contrast to the portion of that population that is the target population for the evaluation. The definition of the target population must be worked out from the study objectives.

The study population operationalizes the target population definition. Target populations are often dynamic, with new members being added and former members departing constantly. Target population listings can be incomplete and some members may not be identifiable. In other cases, study population members include individuals not within the target population. For example, researchers may be interested in the impact of development programs for 4-year-olds. The study population may include children from 3.5 to 5.5 years old in a pilot program conducted for the evaluation. The researchers would like to infer their research findings to 4-year-olds, even though the study population includes some children older than 4 and some younger. Since the age at which the developmental program is given may influence its impact, the expanded cohort may present a problem. However, problems with expanding the cohort in this case are lessened if the operational definition for the pilot program is the same definition that would be used if the program were implemented on a large scale. This issue should be brought up and discussed at the initiation of the program to insure that the pilot is consistent with the theory that underlies the implemented program and the intent of program sponsors.

In addition to definition problems that cause differences between the target and study population, nonresponse bias also creates differences. Nonresponse results from the inability to contact certain members of the population or the refusal of requests for data by some members. Absence from the class when cognitive tests were administered could cause nonresponse in the example of the developmental programs. If the nonresponse is truly random, it does not represent a bias. But this is frequently *not* the case. More frequently, the nonrespondents come from a definable subgroup of the population, and the omission of this subgroup from the data that are actually collected creates a bias in the results.

The researcher can never simply assume that nonresponse is unbiased. Usually nonresponses that result from research procedures (e.g., only contacting members of the population during the day rather than also attempting to contact them during the evening hours) or from refusals have a systematic pattern that leaves the nonrespondents underrepresented by the sample. The best way to deal with nonresponse bias is to reduce the existence of nonresponse, thereby reducing the proportion of the population that is underrepresented (Kalton, 1983, p. 64). Follow-ups to initial data collection attempts, trying multiple methods of contacting sample members, and using methods that minimize respondent refusals to participate are practical means of reducing nonresponse.

Other sources of nonsampling bias are measurement error and errors that occur in recording, coding, or transferring data. These two topics are beyond the scope of practical sampling design, although they are important considerations in reducing total error (see Raj, 1972). Implications of measurement error are discussed in Cook and Campbell (1979). Bradburn and Sudman (1980) provide insight from careful empirical research on reducing response bias in survey research through careful wording of items. Advice on reducing both sources of bias is provided in Fowler (1984) for survey research in general and in Lavrakas (1986) for telephone surveys in particular.

Sampling bias. Sampling bias is the difference between the study population value and the expected value for the sample:

$$\text{SB} = \bar{X}_S - E(\bar{x})$$

where SB is the sampling bias, and
$E(\bar{x})$ is the expected value of the mean.

The expected value of the mean is the average of the means obtained by repeating the sampling procedures on the study population. The expected value of the mean is equal to the study population value if the sampling and calculation procedures are unbiased.

Sampling bias can be subdivided into two components: (1) selection bias and (2) estimation bias. Selection bias occurs when not all members of the study population have an equal probability of selection. Estimation procedures can adjust for the unequal probabilities. The adjustments are made by using weights to compensate for the unequal probabilities of selection.

Selecting a sample from a study population list that contains duplicate entries for some members of the population provides an illustrative example of a selection bias. In the citizen survey example presented in Chapter 4, two lists are combined to form the study population list: state income tax returns and medicaid-eligible clients. An individual appearing on both lists would have twice the likelihood of being selected for the sample. It is impossible from a practical standpoint to purge the combined list of any duplicate listings. However, it is possible to adjust for the unequal probability of selection that arises.

To adjust for this unequal probability of selection, a weight (w) equal to the inverse of the increase in the probability of selection should be applied in the estimation process:

$$w = 1/p = 1/2 = .5$$

The probability of selection for this individual was twice the probability of selection for the study population appearing on the list only once. Therefore, this type of individual would receive only one-half of the weight of the other population members to compensate for their increased likelihood of appearing in the sample.

Estimation bias occurs when the average calculated using an estimation technique on all possible simple random samples from a population does not equal the study population value. For example, the median is a biased estimate of the population mean. Selection bias and estimation bias are intrinsically linked by using adjustments in the estimation to compensate for selection bias.

The concept of the expected value of the statistic undergirds sampling theory and provides the basis for many of the practical solutions used in modern sampling practice. The mean provides a convenient example to examine this concept in more detail, although any number of other estimators, such as the standard deviation or a regression coefficient, could be used. The expected value of the mean is the average of the means computed from repeated samples of the study population. Means computed from each sample form a distribution of values around the expected value of the mean. This distribution is known as the sampling distribution (Figure 3.2).

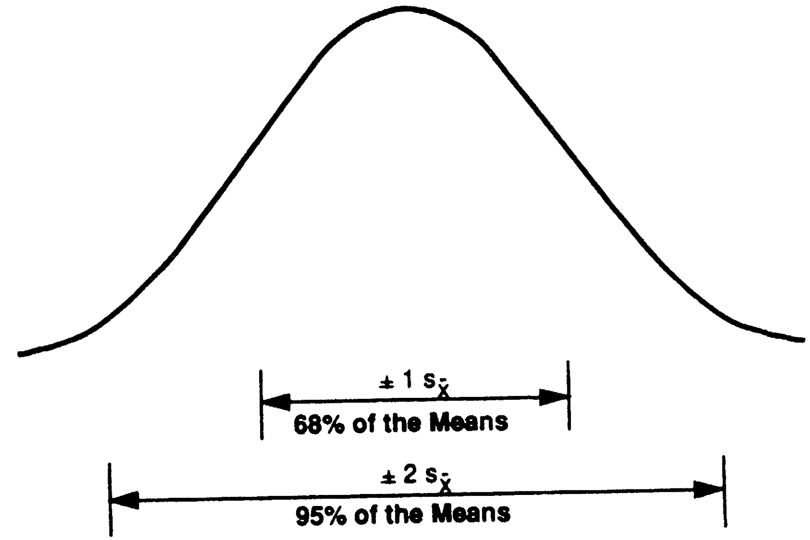

Figure 3.2. Sampling Distribution of the Mean

The sampling distribution when sample sizes reach 30 or more has the shape of the normal distribution, the familiar bell-shaped curve seen in Figure 3.2. This proves to be true no matter what the shape of the frequency distribution of the study population, as long as a large number of samples are selected.

Along with the mean (expected value), the standard deviation of the sampling distribution is another important attribute of the distribution. The properties of the normal distribution allow us to calculate the proportion of the sample means within a defined number of standard deviation units of the study population mean. Figure 3.2 illustrates the sampling distribution, its mean, and the percentage of the means within one or two standard deviation units of the average of the sample means.

Statistical theory also shows that the standard deviation of the sampling distribution ($s_{\bar{x}}$) is inversely related to the sample size. That is, the larger the sample size, the smaller the standard deviation of the sampling distribution. Figure 3.3 graphically shows this relationship.

Tables presenting the area under the normal curve can be used to calculate the percentage of sample means falling within specified standard deviation units. For smaller samples, the student t distribution can be used. For sample sizes below 30, values (the number of the standard deviation units) become larger for smaller sample sizes. For example, for a sample size

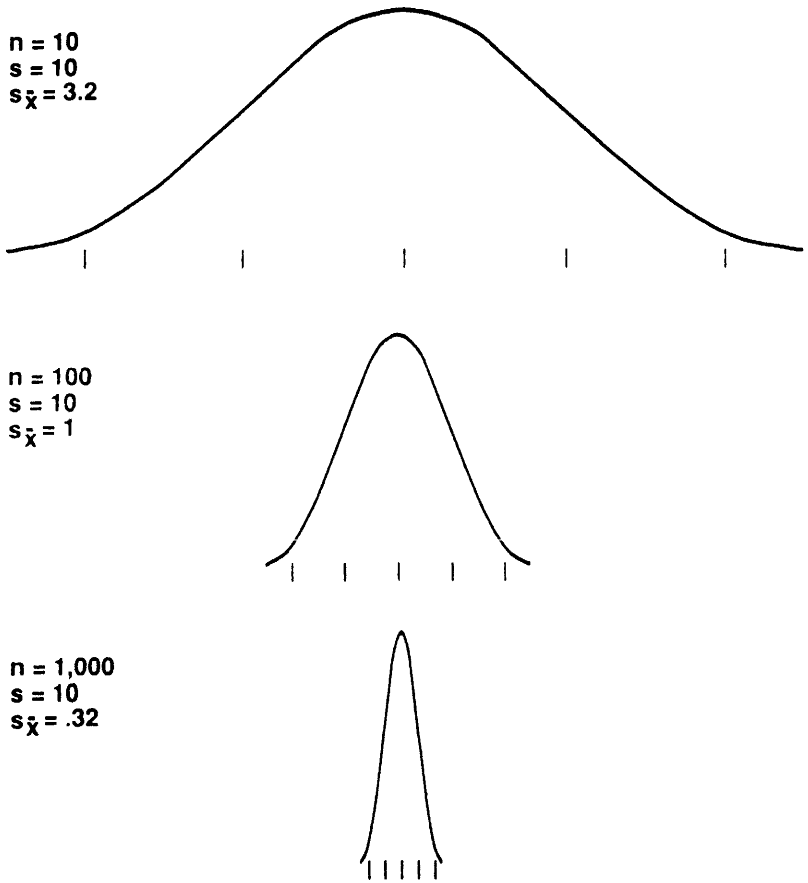

Figure 3.3. Comparing Sampling Distributions with Alternate Sample Sizes

of 100, 95% of the sample means fall within ± 1.96 standard deviation units. A sample size of 10 (9 degrees of freedom) would require ± 2.26 standard deviation units to contain 95% of the sample means. As sample sizes decrease below approximately 30 units, the number of standard deviation units required to contain a specified proportion of sample means must increase. Thus, confidence intervals are larger for smaller samples. These t-values can be found in most tables showing the t distribution. First, compute the degrees of freedom (*df*) by subtracting one from the sample size ($n - 1$). In the case of a sample size of 10, the degrees of freedom would be 9. Then, select the column that indicates the level of one minus the

probability level. That is for 95% probability, an $\alpha = .05$ is used. The column selected should be a two-tailed test. Make sure of this by referring to the directions associated with the table that you are using. If the table indicates a one-tailed test or only one side of the graphic is shaded, double the α. For example, for a table presenting one-tailed values, $\alpha = .025$ would be used for 95% confidence.

Finally, find the cell in the table at the intersection of the row with the appropriate degrees of freedom and the column with the selected probability level. For a sample size of 10, the t-value in the cell to be used for 95% confidence is 2.26, as shown below:

df	$\alpha = .05$	$\alpha = .01$
1	12.71	63.66
.	.	.
.	.	.
.	.	.
9	2.26	3.25
10	2.23	3.17

This means that 2.26 standard deviation units on each side of the mean will include 95% of the sample means. For a 99% confidence level, 3.25 standard deviation units are required.

Sampling variability. The final component of total error in a sample is directly attributable to the fact that statistics from randomly selected samples will vary from one sample to the next due to chance. In any particular sample, some members of the study population will be included and others will be excluded, producing this variation. For this reason, statistics are often noted to be random variables, varying by chance. Because statistics are not usually exactly equal to the study population value, it is useful to have an estimate of their proximity to the population value, in other words, their precision.

Using the properties of the sampling distribution, the interval in which 95% of the sample means fall can be computed:

$$E(\bar{x}) \pm 1.96(s_{\bar{x}})$$

where $s_{\bar{x}}$ is the standard deviation of the sampling distribution (standard error).

However, the practical problem a researcher is presented with is different than this. First, the researcher wishes to know how close the statistic calculated from sample data is to the population value. Second, the standard deviation of the sampling distribution is seldom known in actual research situations because only one sample rather than repeated samples are chosen.

The first problem is quickly overcome. The property, cited above, which indicates that a known percentage of the means in the sampling distribution fall within a related number of standard deviation units of the mean of the sampling distribution (e.g., 95% of the means fall within $\pm 2\ s_{\bar{x}}$) is used here. When the sample mean is calculated, an interval can be drawn around it such that the interval has a predetermined probability of containing the study population mean:

$$\bar{x} \pm t(s_{\bar{x}})$$

where t is the t-statistic for the predetermined probability level, and
$s_{\bar{x}}$ is the standard error of the mean.

The interval drawn around the sample mean in this case is the interval within which the researcher expects that the true mean falls. The degree of confidence associated with this expectation is based on the predetermined probability of the selected t-statistic. The interval is often described as a confidence interval.

The second problem involves estimating the standard deviation of the sampling distribution or the standard error. Properties of the standard deviation of the population, its sample estimate, and the standard deviation of the sampling distribution must be used to solve this problem. The standard deviation of the population is the square root of the sum of squared deviations from the population mean divided by the number of units in the population:

$$S = (\Sigma(x_i - \bar{x})^2/N)^{1/2}$$

The population standard deviation can be estimated from the sample data using the following formula:

$$s = (\Sigma(x_i - \bar{x})^2 / (n - 1))^{1/2}$$

The standard deviation of the sampling distribution is the square root of the sum of the squared deviations of the sample means from the average of the sample means divided by the number of sample means:

$$S_{\bar{x}} = (\Sigma(\bar{x} - \bar{X})^2/m)^{1/2}$$

where m = the number of sample means.

This is referred to as the standard error or sampling error.

Sampling theory shows that the standard deviation of the sampling distribution is related to the sample estimate of standard deviation of the population by the following formula:

$$s_{\bar{x}} = s/(n)^{1/2}$$

From the formula, it can be deduced that two factors have an influence on sampling variability: the variability of the variable (standard deviation) and the size of the sample. Smaller standard deviations reduce the sampling error of the mean. The larger the sample, the smaller the standard deviation of the sampling distribution.

Since the standard deviation for the population can be estimated from the sample information and the sample size is known, a formula can be used to estimate the standard deviation of the sampling distribution, referred to hereafter as the standard error of the estimate, in this particular case, the standard error of the mean:

$$s_{\bar{x}} = s/(n)^{1/2}$$

$$s = (\Sigma(x_i - \bar{x})^2/(n - 1))^{1/2}$$

where $s_{\bar{x}}$ is the estimate of the standard error,
s is the estimate of the standard deviation,
n is the sample size,
x_i are the sample observations, and
$\bar{x}$ is the mean of the sample.

Using this formula allows the researcher to estimate the standard error, the statistic used to measure the final component of total error, based solely on information from the sample.

Probability sampling design discussions in this book assume that the sample would be selected without replacement; that is, once a unit has been randomly drawn from the population to appear in the sample, it is set aside and not eligible to be selected again. Sampling without replacement limits the cases available for selection as more are drawn from the population. The finite nature of the population under this condition may cause a finite population correction factor (FPC) to be needed in the computation of the standard error of the estimate.

For the standard error of the mean, the formula using the FPC is:

$$s_{\bar{x}} = (1 - n/N)^{1/2} s/(n)^{1/2}$$

As a rule of thumb, the sample must contain over 5% of the population to necessitate using the FPC. This is based on the fact that the finite population correction factor is so close to 1 when the sampling fraction is less than .05 that it does not appreciably affect the standard error calculation.

Standard error calculations are specific to the particular statistics being estimated. For example, the standard error for proportions is also commonly used:

$$s_p = ((pq)/n)^{1/2}$$

Most statistics textbooks present formulas for the standard error of several estimators. Also, they are calculated for the statistic being used by almost any statistical software package. These formulas, like the formulas presented above, assume that a simple random sample design has been used to select the sample. Formulas for more complex sampling techniques will be presented in Chapter 7.

One further note on terminology: sampling error and standard error are used interchangeably in the literature. They are specific statistics that measure the more general concept of sampling variability. Standard error, however, is the preferred term. The common use of the term sampling error is unfortunate for two reasons. First, it implies an error in procedures rather than a natural occurrence. Second, it often becomes substituted for the total error concept, which is more comprehensive. Examples of actual standard error calculations are shown in the examples in Chapter 4.

Total error. Sample design is a conscious process of making trade-offs to minimize the three components of total error. Too frequently, reducing the standard error becomes the exclusive focus of sample design because it can be estimated. Because the two bias components cannot be easily calculated, they are often given short shrift during the design process. However, failing sufficiently to consider and attempt to reduce all three components of total error can lessen the validity and credibility of the study findings.

The total error concept is summarized graphically in Figure 3.4. The top of the figure shows the frequency distribution of the target population, based on perfect information. The distribution includes a value for every member of the population arranged from low to high (left to right). Nonsampling bias is illustrated by the difference between the true mean of the target population and the observed mean of the study population. Nonsampling bias includes differences arising from the definition of the population, as well as errors attributable to instrumentation problems and field operations.

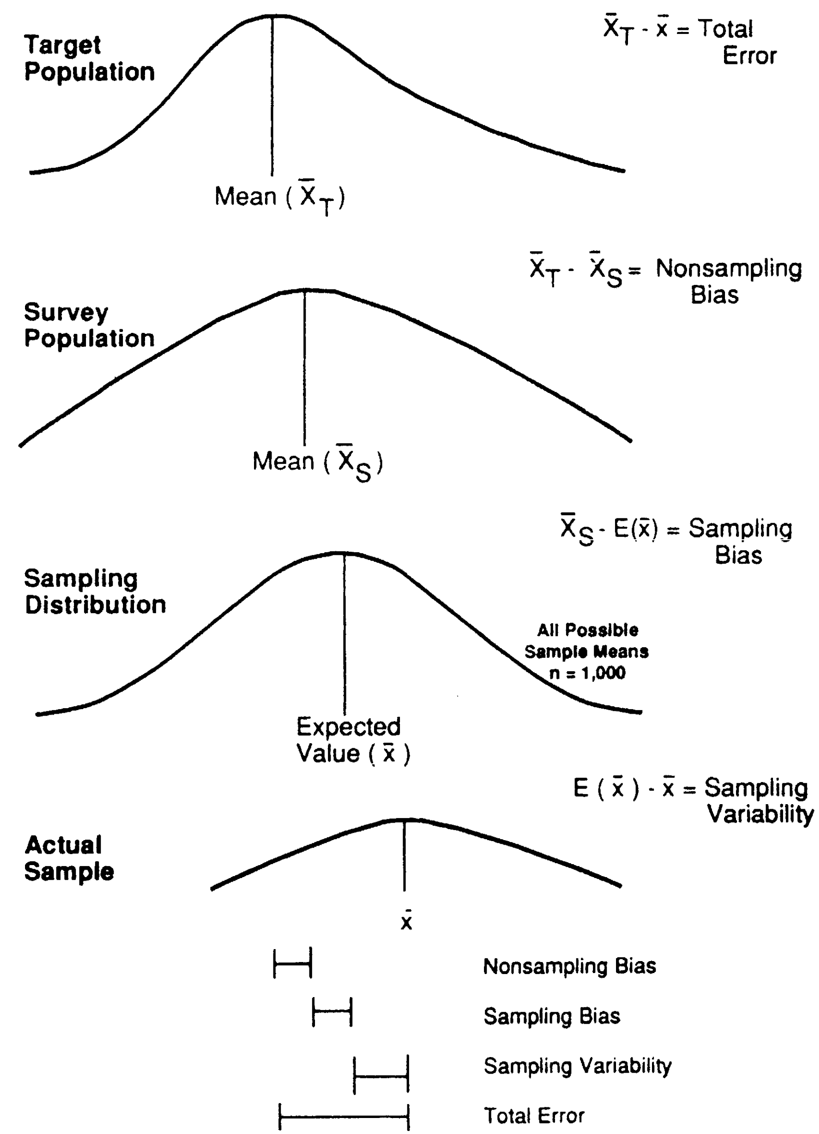

Figure 3.4. Total Error in Sample Design

Sampling bias is illustrated by the difference between the observed mean of the study population and the mean of the sampling distribution, $E(\bar{x})$. Differences here are due to bias in selection or bias in the estimation process. Finally, sampling variability is illustrated. It is the difference between the mean of the sampling distribution and its sample estimate, $\bar{x}$ in this case.

The next section presents the practical sample design framework that will assist the researcher in concentrating on total error and provide criteria for the decisions that are required in the design process.

PRACTICAL SAMPLING DESIGN FRAMEWORK

The framework for practical sampling design is a heuristic tool for researchers to use in sample design. The framework is, in essence, a series of choices that must be made, with each choice having implications for the integrity of the study. The purpose of providing the framework is to help researchers and consumers of research structure their thinking about design choices and the effects of their choices on total error.

The framework includes three phases of the overall design of the research project:

1. Presampling choices
2. Sampling choices
3. Postsampling choices

Presampling choices are the choices that researchers should make when the study is being designed. These choices set the stage for sampling choices, which follow and provide a base for analyzing total error. For example, one presampling choice is the definition of the target population. The definition of the target population relates to the study purpose. If the study is an evaluation of community mental-health services for the homeless, the target population could logically be defined as homeless who received any community mental health services in the state of Virginia between January 1987 and December 1988. This definition will be used as sampling choices concerning the listing of the target population from which the sample will be selected are made. Differences between the target population and the study population result in nonsampling bias.

Sampling choices are the usual fare for discussions of sampling. Each of these choices has direct and immediate impact on total error. One sampling choice is whether the probability of selection of the sample will be equal or unequal. If the probability is unequal and it is not adjusted through the use of weights, then sampling bias occurs.

Postsampling choices require decisions about procedures that are required after data are collected. One postsampling choice is the selection of a method to calculate standard errors. Usually, standard error calculations are based on sampling choices that have been made, particularly the type of sampling

TABLE 3.1
Questions for Sample Design

Presampling Choices
1. What is the nature of the study—exploratory, descriptive, or analytical?
2. What are the variables of greatest interest?
3. What is the target population for the study?
4. Are subpopulations or special groups important for the study?
5. How will the data be collected?
6. Is sampling appropriate?

Sampling Choices
1. What listing of the target population can be used for the sampling frame?
2. What is the tolerable error or estimated effect size?
3. What type of sampling technique will be used?
4. Will the probability of selection be equal or unequal?
5. How many units will be selected for the sample?

Postsampling Choices
1. How can nonresponse be evaluated?
2. Is weighting necessary?
3. What are the standard errors and related confidence intervals for the study estimates?

technique. Postsampling choices, such as this one, help to evaluate the extent of total error resulting from the sample, and, in other cases, reduce the total error.

This brief introduction to a few of the choices in the practical sampling framework demonstrates some interrelationships between choices made throughout the study process. The practical sampling design approach integrates sampling-related choices into research design and execution, rather than separating the sampling choices from the other study choices. A listing of the questions to be asked relevant to sampling choices appears in Table 3.1. Each of these questions is explained in the next sections of this chapter.

The practical sampling design choices represented by these questions bear directly on total error. Presampling choices relate primarily to nonsampling error. The choice concerning the need for analysis of subpopulations is one exception. If subpopulation analysis is a study objective, the sampling variability for the subpopulation analysis will be affected by the number of subpopulation members likely to be in the sample. A presampling choice to conduct subpopulation analysis can impact later sampling choices, such as the type of sampling technique. The sampling technique affects the number of subpopulation members in the sample and, therefore, the sampling variability for the subpopulation analysis.

Sampling choices have impact on all three components of total error. For example, the listing of the target population will affect the amount of nonsampling bias. A choice of unequal probability of selection will introduce bias into the sample. The number of units or subjects in the sample influences the sampling variability. Postsampling choices are ways of estimating, adjusting, or analyzing the components of total error.

PRESAMPLING CHOICES

What Is the Nature of the Study—Exploratory, Descriptive, or Analytical?

Exploratory research is generally conducted to provide an orientation or familiarization with the topic under study. It serves to enlighten the researcher about salient issues, helps focus future research on important variables, and generates hypotheses to be tested. Descriptive research is the core of many survey research projects where estimates of population characteristics, attributes, or attitudes are study objectives. Analytical research tests hypotheses, examining relationships between groups and/or relationships between variables. In practice, many studies attempt both descriptive and analytical tasks. It is important to understand this and establish priorities for the two tasks in these cases.

Exploratory research is often a preliminary activity leading to a more rigorous descriptive or analytical study. The sampling approach is usually severely limited by resource and time constraints. Small stratified probability samples or purposeful quota samples are frequently used to ensure that a wide range of groups are covered in the study. Broad coverage is more important in many exploratory research projects than reducing error since estimates, such as averages and proportions, are not reasonable study products.

Both descriptive and analytical studies are concerned with reducing total error. While they have similar objectives for reducing bias, the standard error component of total error is addressed differently. For descriptive studies, the focus is on the precision needed for the estimates. Needs assessments are one type of policy-related study where concern for precision of estimates is manifest. To make program decisions about the number of elderly that need services and the particular types of services needed, for instance, the samples are only useful if the estimates are sufficiently precise for budgetary and service delivery planning. For analytical studies, determining if the study will be powerful enough to reject the null hypothesis

for the expected effect size is the relevant concern. This is done through a power test (Lipsey, 1989). Evaluation research, which often assesses the impact of a program on a target population, is likely to fall into the category of an analytical study.

What Are the Variables of Greatest Interest?

Selecting variables and estimating their variability is an important precursor to the sampling design. Studies often have multiple purposes. The researcher may envision many tables appearing in the write-up or using several statistical tools to test hypotheses. The selection of the variables of greatest interest will have important implications in the determination of sample size later on in the design process.

What Is the Target Population for the Study?

The target population for a study is the group for which the researchers would like to make general statements. The population can be individuals (residents of North Carolina or homeless in Los Angeles), groups of individuals (households in Richmond or schools in Wisconsin), or other elements (invoices, state-owned cars, or dwelling units). The theory being tested or the policy under study will help define the target population. Specific time frames, geographic location, age, and other relevant criteria are often used in the definition.

Are Subpopulations or Special Groups Important for the Study?

Often a researcher will choose to focus on a part of the target population for additional analysis of the phenomena being scrutinized. For example, households headed by single, working females were of particular interest to some scientists examining the impact of the income maintenance experiments (Skidmore, 1983). Design choices are available for instances when subgroups are known to be important for the study. A sample designed without taking the subpopulation into account can yield too few of the subpopulation members in the sample for reliable analysis. Increasing the overall sample size or disproportionately increasing the subgroup sample, if they can be identified before sampling, are potential remedies.

How Will the Data Be Collected?

Certain sampling choices can only be used in conjunction with certain data collection choices. For example, random digit dialing, a technique that generates a sample of randomly selected phone numbers, is an option when interviews are to be conducted over the phone. A probability sample of dwelling units is mainly useful for studies where on-site fieldwork, usually personal interviews, are to be used. Collecting data from administrative records or mailed questionnaires also poses specific sampling concerns. Mailed questionnaires typically have a high proportion of nonrespondents that affects the sampling variability and may cause nonsampling bias if the sample members that choose not to respond have similar characteristics.

In addition, the reliability of the responses achieved by the administration of the data collection instrument affects sampling considerations. Assuming that the instrument is unbiased, the central concern is that the less reliable an instrument is, the greater the standard error. Reliability of the instrument will be a consideration in selecting the sample size to avoid problems with statistical conclusion validity because of the inflation of the standard error.

Is Sampling Appropriate?

The decision to sample should be consciously made. Sampling is generally required to meet resource constraints. Furthermore, in many cases it will produce more accurate results than a population-census type study. Often, resources for studies of an entire population are consumed in attempting to contact all population members. Response to the first contact is often less than 50%, causing substantial nonsampling bias. Sampling the population would require fewer resources for the initial contacts and allow more resources to be invested in follow-up activities designed to increase response, paying dividends in lowering nonsampling bias. Also, having fewer cases by using sampling allows greater attention to accuracy in handling data. On the other hand, small populations and use of the information in the political environment may weigh against sampling. Careful consideration of the reasons for not sampling can be useful in designing a sample that will overcome most of the objections. For example, the political objection that every community in the state was not included when reviewing the program can be partially overcome if the sample can provide regional estimates.

SAMPLING CHOICES

What Listing of the Target Population Can Be Used for the Sampling Frame?

The sampling frame is the list from which the sample is selected. The sampling frame provides the definition of the study population and differences between the target population and sampling frame constitute a nonsampling bias. The sampling frame is the operational definition of the population, the group about which the researchers can reasonably speak. A telephone directory can provide the sampling frame for a study of the population in a community. It is an explicit sampling frame consisting of all households within the service area with working phones and listed numbers. For random digit dialing, the sampling frame is implicit, rather than explicit, that is, residences with working phones; no physical list of the population is obtained. Often, this frame is further refined by using a screen to determine if the residence has a member of the target population or to select a household member at random rather than selecting the individual that answers the phone by default.

Also, systematic, multistage, and cluster sampling do not necessarily require a physical listing of the entire target population. Systematic sampling is often done with the objects themselves, such as pulling invoices from a file cabinet or selecting individuals from a line in a soup kitchen. For cluster and multistage sampling, only a complete listing of the clusters or the sampling units is needed at each stage. Members of the target population are only listed in the final stage of multistage sampling and only for the sampling units selected in the stage immediately before.

Imperfect sampling frames can lead to nonsampling error. The most difficult imperfection to overcome is the omission of part of the target population from the sampling frame. This can lead to a bias that cannot be estimated for the sample data. Using multiple listings to formulate a frame, choosing a technique that does not require a frame, or conducting a special supplemental study to tap the omitted portion of the population are alternatives to ameliorate the problem or at least to estimate its size.

What Is the Tolerable Error or the Estimated Effect Size?

To begin consideration of sampling variability in the context of choices, the tolerable error or estimated effect size must be addressed. For descriptive studies, the tolerable error of the estimates must be determined for the variables of greatest interest. The tolerable error relates to the size of the interval around the estimate that is expected, with a specified degree of confidence, to include the population value. For example, an estimate of the percentage of the homeless population that have been released from mental hospitals may need to be within 5% of the actual value. For a candidate in a close election, the tolerable error in a pre-election poll may be 1%. Tolerable error includes an assumption about the likelihood that the true value is contained in the interval.

The estimated size effect must be determined for analytical studies. The estimated size effect is the amount of difference the treatment or independent variable is likely to make in the dependent variable(s). Consider the case of legislation that requires an additional one-year sentence to be added to a sentence when a handgun is used in the commission of a crime. The effect of the legislation is expected to be a 12-month increase in sentences when crimes are committed using a handgun. Research evaluating the actual increase in sentences would need to be sensitive enough to detect a change of sentence length of 12 months. The effect size for a developmental program for 4-year-olds could be estimated in terms of increases in standardized test scores or decreases in the percentage of students held back in the first grade or both. The effect size used in this case could be the minimum amount of change that program sponsors considered to be sufficient to justify costs of the program.

The likely effect size and the tolerable error are used in the calculation of efficient sample sizes. The objective of the researcher is to produce estimates within the limits of tolerable error or to conduct an analytical test that is sensitive enough to detect the estimated effect. Sample size is a principal means by which the researcher can achieve these objectives. But the efficiency of the sampling technique can have considerable impact on the amount of sampling error and the estimate of the desired sample size.

What Type of Sampling Technique Will Be Used?

In Chapter 2, the five basic probability sampling techniques were described:

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Cluster sampling
- Multistage sampling

While all the techniques are probability techniques, they can each result in sampling bias. The choice of a technique will depend on several factors, including the availability of an adequate sampling frame; the availability of prior information about the target population; the need for greater efficiency; the need to conduct interviews on-site, or alternatively, over the phone; and the location of the target population. However, the choices do not end with the selection of a technique.

Choices branch off independently for each technique. If stratified sampling is chosen, how many strata should be used? If researchers choose cluster sampling, how should the clusters be defined? For multistage samples, can the researcher reduce sampling variability by selecting more of the primary sampling units with fewer secondary units for each primary unit than by selecting fewer primary units with more secondary units for each primary unit? Chapter 4 presents examples of how these choices have been made in practice. Chapter 6 discusses implications of the alternatives.

Will the Probability of Selection Be Equal or Unequal?

Choices about the probability of selection will also affect sampling bias. For simple random sampling, the probability of selecting any individual unit is equal to the sampling fraction or the proportion of the population selected for the sample (n/N). The probability of selecting any unit is equal. For stratified sampling, the probability of selection for any unit is the sampling fraction for the stratum in which the unit is placed. Probabilities using a stratified sampling technique can be either equal or unequal. Multistage samples have the most complex probability of selection calculation. The overall probability of selection is the product of the probabilities of selection in each stage. The calculation must be done separately for each stratum and sampling unit in each stage.

A sample with equal probability of selection is termed a self-weighting sample, indicating that no weights are needed to adjust for unequal probabilities. Unequal probabilities of selection are biased and require weights to be used for estimation and analysis when a design-based approach is being followed. A design-based approach is an approach to estimation and analysis that follows the structure of the design in terms of using weights to compensate for unequal probabilities created by the design. Since the design-based approach to estimation and analysis prevails among sampling practitioners, the design-based approach will be followed throughout this text. Readers interested in an alternative, the model-based approach, are referred to Smith (1976) and Kalton (1986).

How Many Units Will Be Selected for the Sample?

Researchers directly affect the amount of sampling variability in their choice of sample size. Sample size depends on a number of factors. Estimating an efficient sample size is a reasonable place to begin the researcher's decision process. Efficient sample size calculations are ways to estimate the size of the sample needed to fulfill the study objectives once a particular selection technique has been chosen and specified in operational terms. Efficient sample size is computed in one of two ways depending on the nature of the study.

For descriptive studies, the question posed is, what sample size will produce estimates that are precise enough to be useful, given the sampling technique? This relates directly to the choice of the amount of tolerable error. Recall that tolerable error relates solely to the variability due to sampling and does not include the other components of total error. The standard error of the estimate is the measure of sampling variability. Tolerable error is the allowable standard error of the estimate times the t-value selected for the desired probability that the true value will be contained in the interval around the estimate:

$$te = ts_{\bar{x}}$$

where te is tolerable error,
t is the t-value for the desired probability, and
$s_{\bar{x}}$ is the standard error of the estimate.

Tolerable error is the size of one side of the confidence interval, presented earlier ($\pm ts_{\bar{x}}$). Thus, in computing the efficient sample size, the tolerable error is actually one half of the size of the confidence interval that the researcher can tolerate for the study. The tolerable error relates to the amount of precision needed for the estimates.

Based on the formulas presented earlier for the confidence interval, the size of the interval is primarily influenced by three variables: the standard deviation, the sample size, and the t-statistic. To a lesser extent, it can be influenced by the sampling fraction as a result of the finite population correction (FPC). The researcher directly controls only the sample size: to produce an estimate from the sample that is precise enough for the study objectives the researcher can adjust the sample size. But increasing the sample size means increasing the cost of the data collection. Trade-offs between precision and cost are inherent at this juncture.

For a descriptive study, assuming a simple random sample, the sample size calculation is an algebraic transformation of the standard error calculation:

$$n' = s^2/(te/t)^2$$

$$n = n'/(1 + f)$$

where n' is the sample size computed in the first step,
s is the estimate of the standard deviation,
te is the tolerable error,
t is the t-value for the desired probability level,

n is the efficient sample size using the finite population correction factor, and
f is the sampling fraction.

Tolerable error, discussed above, is used in this equation. Tolerable error is the allowable standard error of the estimate multiplied by the t-value. In other words, it is the value that will be added to and subtracted from the estimate to form the confidence interval. Therefore, in this equation the tolerable error is divided by the t-value so that it is expressed comparably to the standard error. An example of using this formula for sample size decision is presented in Chapter 4 (see Table 4.6 especially).

The most difficult piece of information to obtain for this formula, considering it is used prior to conducting the actual data collection, is the estimate of the standard deviation. A number of options are available, including prior studies, small pilot studies, and estimates using the range. These options are discussed in Chapter 7.

Power tests are used to compute efficient sample sizes for studies that are primarily analytical. A power test is used to indicate whether a particular sample size is sufficiently sensitive to detect the expected effect (Lipsey, 1989). Rejecting the null hypothesis that there is no effect from the program or treatment depends primarily on the size of the effect and the size of the standard error of the estimate. The larger the standard error or the smaller the effect size, the more difficult it is to reject the null hypothesis. Failure to reject the null hypothesis when it is in fact false is described as a Type II error.

Once again, sample size is the researcher's main tool in obtaining a plausible chance of rejecting the null hypothesis. Lipsey (1989) cites and explains additional methods of avoiding Type II errors, treating the power issue in the overall context of design sensitivity.

While the sample size is the principal means for influencing the precision of the estimate once the design is set, an iterative process can be used to examine the impact on efficient sample size when altering the design, especially the sampling technique. Improved stratification or the selection of more primary sampling units in multistage sampling can improve the efficiency of the design. Of course, these adjustments are also likely to increase costs, but perhaps less than increasing the sample size would.

In addition, other sample size considerations should be brought to bear at this point. For example, will the number of members of subpopulations that are to be described be sufficient in the design using the efficient sample size? When the members of an important subpopulation are a small fraction of the population, it may not be possible to increase the total sample size to a number that would yield a sufficient number of subpopulation members for the analysis. Disproportionate stratification designed to increase the number of subpopulation members or a supplemental sample are possible alternatives. Both have a cost that must be considered, and they may change the efficiency of the design.

Determining the sample size is generally an iterative process. Numerous factors are considered and analyzed that may alter earlier choices. It is important carefully to review the proposed alternatives in terms of total error, changes in the study population definition from using different sampling frames, ability to meet the study objectives, time, and cost. Usually it is beneficial to review each item that has appeared in the framework to examine the impact of the alternative. Impact on other factors such as the cost and interviewer training and follow-up for nonrespondents must also be considered when making the sampling choices.

POSTSAMPLING CHOICES

How Can the Impact of Nonresponse Be Evaluated?

Nonresponse is the lack of valid responses from some members of the sample. Nonresponse can occur when the respondent refuses to answer a particular question or refuses to participate in the survey, or when the respondent cannot be contacted. Nonresponse is one component of nonsampling bias. Nonresponse leads to differences between the study population and the target population. In essence, the population is divided into two subpopulations, responders and nonresponders. The smaller the subpopulation of nonresponders, the less their impact on biasing the results can be (Kalton, 1983). So the best way of dealing with nonresponse is to eliminate it. Fowler (1984) discusses several ways of reducing nonresponse.

Since entirely eliminating nonresponse is unrealistic, the researcher is left having to assess its impact. This can be done by comparing characteristics of the sample with characteristics of the study population, for variables where population values are available. The comparison can help by showing which types of population members are underrepresented. However, the characteristics of the population members for which data are available may not directly relate to variables that are important for the current study. For example, demographic variables such as age, gender, and race are often the only data available for the population. In addition, the data are often out of date and may not reflect the distribution of these characteristics in the current population.

However, where the population data are reasonably timely, it is often useful to weight the sample data to obtain the population proportions and examine the impact of the weighting on the variables of greatest interest for the study. This practice assumes no differences in the responders and nonresponders. It simply compensates for underrepresentation of certain groups. This assumption is not generally borne out.

To test for the amount of bias, a subsample of nonresponders could be randomly selected and intensively followed up to obtain information on important variables for the study. This is sometimes referred to as an attrition study. In the case of mailed surveys, a phone interview follow-up may be used (Dillman & Tarnai, 1988). The data that are obtained can be compared to the data obtained in the original effort to evaluate the potential bias caused by nonresponse. Evaluating nonresponse is discussed in Chapter 8.

Is It Necessary to Weight the Sample Data?

Weighting is usually required to compensate for sampling bias when unequal probabilities result from the researchers' sampling choices. Sometimes weights are needed in some analyses and not in others. For example, a study may have two units of analysis, households and individuals. From each household one individual is selected by a random process to respond. The households are selected with equal probability, so no weights are needed when households are the object of inquiry. Because the probability of selection of a member of the household to be a respondent depends on the number of individuals living in the household, unequal probabilities occur when individuals are the object of inquiry. Thus, weights are needed. The use of weights is discussed in Chapter 8 and examples are presented in Chapter 4.

What Are the Standard Errors for the Study Variables?

Standard errors are important for both descriptive and analytical studies. The precision of the estimates and the sensitivity of hypothesis tests are determined by the standard errors. Standard errors are the measures of sampling variability. Calculating standard errors is complex, since the formulas are different for every statistic and every technique.

Two approaches are used to estimate standard error: a direct method and a more complex method requiring the approximation of the deviations. In practice, the direct method can be used for simple random, systematic, stratified, and cluster samples. Formulas for the direct method can be illustrated by the standard error of the mean calculation for simple random samples that was presented and discussed earlier in the chapter:

$$s_{\bar{x}} = (1 - f)^{1/2} s/(n^{1/2})$$

These formulas are stock features of statistics texts for a variety of statistics assuming that a simple random sample is used.

Other sampling techniques require modifications to the formula using the direct approach. The ratio of the sampling variance ($s_{\bar{x}}{}^2$) for the actual design to the sampling variance, assuming a simple random sample, is called the design effect (Kish, 1965). This ratio shows the extent to which the actual sample design has increased or decreased the sampling variability component of total error. Stratification lowers the sampling error, all other things held constant, and has a design effect less than one. Sampling error and design effect can be further lowered when larger sampling fractions are allocated to strata that have the highest standard deviations.

Cluster sampling has the opposite impact on the design effect: it causes the design effect to exceed one. This occurs because the number of independent choices is the number of clusters in cluster sampling, not the number of units finally selected. The effect is reduced when:

clusters are internally heterogeneous on the important study variables (large standard deviations within the clusters), or

cluster means do not vary.

Since clusters are often geographically defined, increasing the size of the clusters to contain more heterogeneous groupings increases data collection costs. Reducing the sampling variability can also be achieved through stratification of the clusters before selection. This means that the clusters must be placed into strata before selection, and the variables used to define the strata must be available for all clusters.

The calculation of the sampling error for complex, multistage samples using the Taylor approximation of deviations method or some form of repeated replications is only practical with the aid of a computer. In addition, they require that at least two selections be made from each stratum or primary sampling units. Otherwise selections must be combined across strata. Sampling error calculations for complex samples are presented in Chapter 6.

SUMMARY

The challenge of practical sample design is making trade-offs to reduce total error, while keeping study goals and resources in mind. The sampler must act to make choices throughout the sampling process to reduce error. But reducing the error associated with one choice can increase errors from other sources.

Faced with this complex, multidimensional challenge, the rcsearcher must concentrate on reducing total error. Error can arise systematically from bias or occur due to random fluctuation inherent in sampling. Error cannot be entirely eliminated. Reducing error is the practical objective, which can be achieved through careful design.

This chapter has pointed out three components of total error: nonsampling bias, sampling bias, and sampling variability. Each must be addressed through the design. The chapter also presented some basic questions that must be grappled with throughout the design process. Choices to be made at each stage in the process interact with other choices. The next chapter presents four examples of the way choices were made by researchers conducting applied social research.