CHAPTER TWO

INFERENCE AND ERROR IN SURVEYS

2.1 INTRODUCTION

Survey methodology seeks to understand why error arises in survey statistics. Chapters 3 through 11 describe in detail strategies for measuring and minimizing error. In order to appreciate those chapters, and to understand survey methodology, it is first necessary to understand thoroughly what we mean by "error."

As the starting point, let us think about how surveys work to produce statistical descriptions of populations. Figure 2.1 provides the simplest diagram of how they work. At the bottom left is the raw material of surveys—answers to questions by an individual. These have value to the extent they are good descrip-

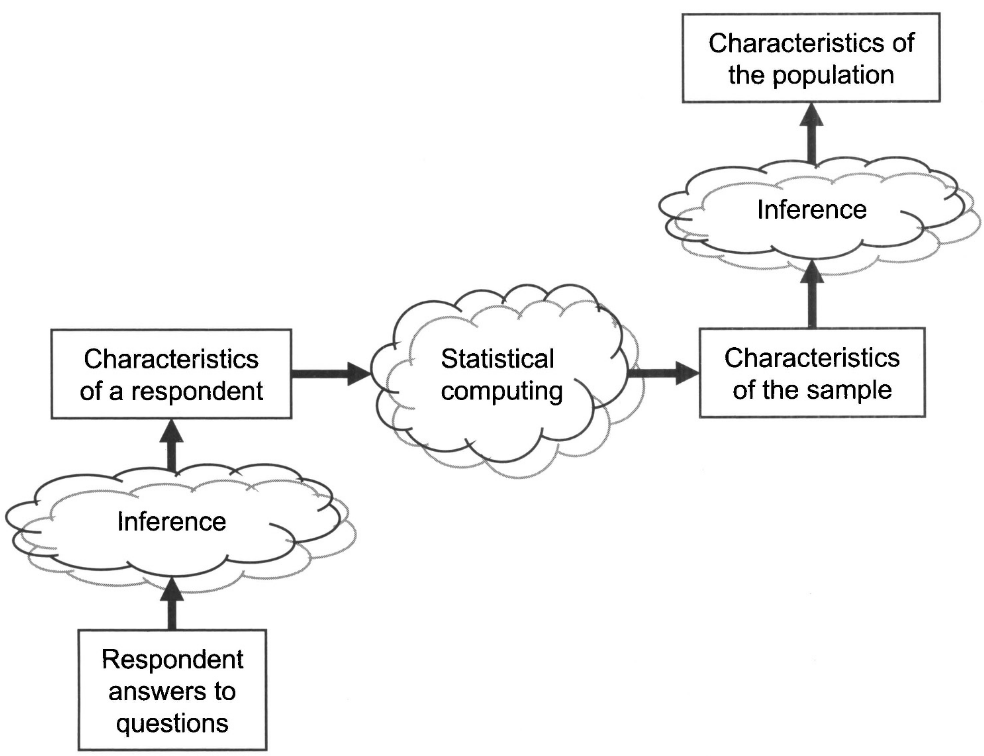

Figure 2.1 Two types of survey inference.

tors of the characteristics of interest (the next-higher box on the left). Surveys, however, are never interested in the characteristics of individual respondents per se. They are interested in statistics that combine those answers to summarize the characteristics of groups of persons. *Sample* surveys combine the answers of individual respondents in statistical computing steps (the middle cloud in Figure 2.1) to construct statistics describing all persons in the sample. At this point, a survey is one step away from its goal—the description of characteristics of a larger population from which the sample was drawn.

statistic

The vertical arrows in Figure 2.1 are "inferential steps." That is, they use information obtained imperfectly to describe a more abstract, larger entity. "Inference" in surveys is the formal logic that permits description of unobserved phenomena based on observed phenomena. For example, inference about unobserved mental states, like opinions, is made based on answers to specific questions related to those opinions. Inference about population elements not measured is made based on observations of a sample of others from the same population. In the jargon of survey methodology, we use an answer to a question from an individual respondent to draw inferences about the characteristic of interest to the survey for that person. We use statistics computed on the respondents to draw inferences about the characteristics of the larger population.

inference

These two inferential steps are central to the two needed characteristics of a survey:

1) Answers people give must accurately describe characteristics of the respondents.
2) The subset of persons participating in the survey must have characteristics similar to those of a larger population.

When either of these two conditions is not met, the survey statistics are subject to "error." The use of the term "error" does not imply mistakes in the colloquial sense. Instead, it refers to deviations of what is desired in the survey process from what is attained. "Measurement errors" or "errors of observation" will pertain to deviations from answers given to a survey question and the underlying attribute being measured. "Errors of nonobservation" will pertain to the deviations of a statistic estimated on a sample from that on the full population.

error

measurement error

error of observation

error of nonobservation

Let us give an example to make this real. The Current Employment Statistics (CES) program is interested in measuring the total number of jobs in existence in the United States during a specific month. It asks individual sample employers to report how many persons were on their payroll in the week of the 12th of that month. (An error can arise because the survey does not attempt to measure job counts in other weeks of the month.) Some employer's records are incomplete or out of date. (An error can arise from poor records used to respond.) These are problems of inference from the answers obtained to the desired characteristic to be measured (the leftmost vertical arrows in Figure 2.1).

The sample of the employers chosen is based on lists of units of state unemployment compensation rolls months before the month in question. Newly created employers are omitted. (An error can arise from using out-of-date lists of employers.) The specific set of employers chosen for the sample might not be a good reflection of the characteristics of the total population of employers. (An error can arise from sampling only a subset of employers into the survey.) Further, not all selected employers respond. (An error can arise from the absence of answers from some of the selected employers.) These are problems of inference from statistics on the respondents to statistics on the full population.

One's first reaction to this litany of errors may be that it seems impossible for surveys ever to be useful tools to describe large populations. Do not despair! Despite all these potential sources of error, carefully designed, conducted, and analyzed surveys have been found to be uniquely informative tools to describe the world. Survey methodology is the study of what makes survey statistics more or less informative.

Survey methodology has classified these various errors illustrated with the CES example above into separate categories. There are separate research literatures for each error because each seems to be subject to different influences and have different kinds of effects on survey statistics.

One way of learning about surveys is to examine each type of error in turn, or studying surveys from a "quality" perspective. This is a perspective peculiar to survey methodology. Another way of learning about surveys is to study all the survey design decisions that are required to construct a survey; identification of the appropriate population to study, choosing a way of listing the population, selecting a sampling scheme, choosing modes of data collection, and so on. This is an approach common to texts on survey research (e.g., Babbie, 1990; Fowler, 2001).

2.2 The Life Cycle of a Survey From a Design Perspective

In this and the next section, we will describe the two dominant perspectives about surveys: the design perspective and the quality perspective. From the design perspective, discussed in this section, survey designs move from abstract ideas to concrete actions. From the quality perspective, survey designs are distinguished by the major sources of error that affect survey statistics. First, we tackle the design perspective.

A survey moves from design to execution. Without a good design, good survey statistics rarely result. As the focus moves from design to execution, the nature of work moves from the abstract to the concrete. Survey results, therefore, depend on inference back to the abstract from the concrete. Figure 2.2 shows that there are two parallel aspects of surveys: the measurement of constructs and descriptions of population attributes. This figure elaborates the two dimensions of inference shown in Figure 2.1. The measurement dimension describes what data are to be collected about the observational units in the sample: what is the survey about? The representational dimension concerns what populations are described by the survey: who is the survey about? Both dimensions require forethought, planning, and careful execution.

observational unit

Because Figure 2.2 contains important components of survey methods, we will spend some time discussing it. We will do so by defining and giving examples of each box in the figure.

2.2.1 Constructs

construct

"Constructs" are the elements of information that are sought by the researcher. The Current Employment Statistics survey attempts to measure how many new

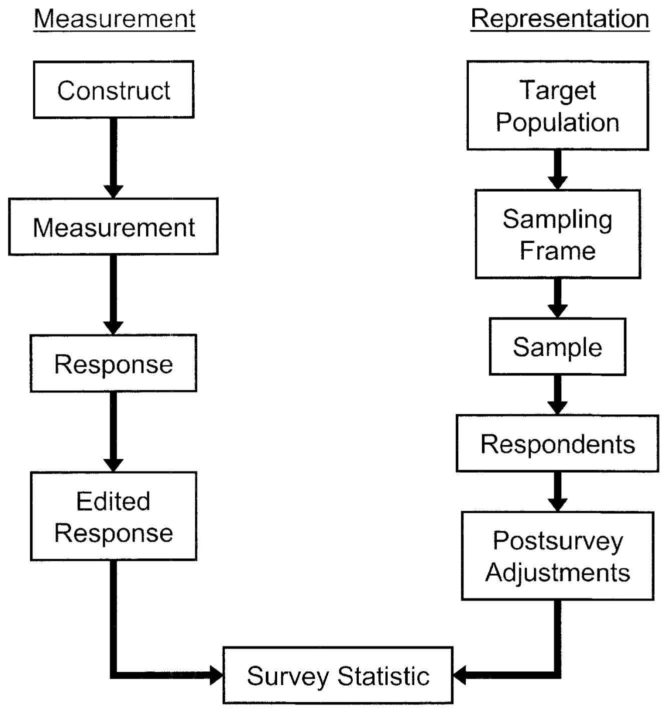

Figure 2.2 Survey lifecycle from a design perspective.

jobs were created in the past month in the United States, the National Assessment of Education Progress measures knowledge in mathematics of school children, and the National Crime Victimization Survey (NCVS) measures how many incidents of crimes with victims there were in the last year. The last sentence can be understood by many; the words are simple. However, the wording is not precise; it is relatively abstract. The words do not describe exactly what is meant, nor exactly what is done to measure the constructs. In some sense, constructs are ideas. They are most often verbally presented.

For example, one ambiguity is the identity of the victim of the crime. When acts of vandalism occur for a household (say, a mailbox being knocked down), who is the victim? (In these cases, NCVS distinguishes crimes against a household from crimes against a person.) When graffiti is spray painted over a public space, who is the victim? Should "victimization" include only those crimes viewed as eligible for prosecution? When does an unpleasant event rise to the level of a crime? All of these are questions that arise when one begins to move from a short verbal description to a measurement operation. Some constructs more easily lend themselves to measurements than others.

Some constructs are more abstract than others. The Survey of Consumers (SOC) measures short-term optimism about one's financial status. This is an attitudinal state of the person, which cannot be directly observed by another person. It is internal to the person, perhaps having aspects that are highly variable within and across persons (e.g., those who carefully track their current financial status may have well-developed answers; those who have never thought about it may have to construct an answer de novo). In contrast, the National Survey on Drug Use and Health (NSDUH) measures consumption of beer in the last month. This is a construct much closer to observable behaviors. There are a limited number of ways this could be measured. The main issues are simply to decide what kinds of drinks count as beer (e.g., does nonalcoholic beer count?) and what units to count (12-ounce cans or bottles is an obvious choice). Thus, the consumer optimism construct is more abstract than the construct concerning beer consumption.

2.2.2 Measurement

measurement

Measurements are more concrete than constructs. "Measurements" in surveys are ways to gather information about constructs. Survey measurements are quite diverse: soil samples from the yards of sample households in surveys about toxic contamination, blood pressure measurements in health surveys, interviewer observations about housing structure conditions, electronic measurements of traffic flow in traffic surveys. However, survey measurements are often questions posed to a respondent, using words (e.g., "During the last 6 months, did you call the police to report something that happened to YOU that you thought was a crime?"). The critical task for measurement is to design questions that produce answers reflecting perfectly the constructs we are trying to measure. These questions can be communicated orally (in telephone or face-to-face modes) or visually (in paper and computer-assisted self-administered surveys). Sometimes, however, they are observations made by the interviewer (e.g., asking the interviewer to observe the type of structure of the sample housing unit or to observe certain attributes of the neighborhood). Sometimes they are electronic or physical measurements (e.g., electronic recording of prices of goods in a sample retail store, taking a blood or hair sample in a health-related survey, taking a sample of earth in a survey of toxic waste, taking paint samples). Sometimes questions posed to respondents follow their observation of visual material (e.g., streaming video presentation of commercials on a laptop, presentation of magazine covers).

2.2.3 Response

response

The data produced in surveys come from information provided through the survey measurements. The nature of the responses is determined often by the nature of the measurements. When questions are used as the measurement device, respondents can use a variety of means to produce a response. They can search their memories and use their judgment to produce an answer [e.g., answering the question, "Now looking ahead, do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?" from the SOC]. They can access records to provide an answer (e.g., looking at the employer's personnel records to report how many nonsupervisory employees were on staff on the week of the 12th, as in the CES). They can seek another person to help answer the question (e.g., asking a spouse to recall when the respondent last visited the doctor).

Sometimes, the responses are provided as part of the question, and the task of the respondent is to choose from the proffered categories. Other times, only the question is presented, and the respondents must generate an answer in their own words. Sometimes, a respondent fails to provide a response to a measurement attempt. This complicates the computation of statistics involving that measure.

2.2.4 Edited Response

In some modes of data collection, the initial measurement provided undergoes a review prior to moving on to the next. In computer-assisted measurement, quantitative answers are subjected to range checks, to flag answers that are outside acceptable limits. For example, if the question asks about year of birth, numbers less than 1890 might lead to a follow-up question verifying the stated year. There may also be consistency checks, which are logical relationships between two different measurements. For example, if the respondent states that she is 14 years old and has given birth to 5 children, there may be a follow-up question that clarifies the apparent discrepancy and permits a correction of any errors of data. With interviewer-administered paper questionnaires, the interviewer often is instructed to review a completed instrument, look for illegible answers, and cross out questions that were skipped in the interview.

outlier detection

After all of the respondents have provided their answers, further editing of data sometimes occurs. This editing may examine the full distribution of answers and look for atypical patterns of responses. This attempt at "outlier detection" often leads to more careful examination of a particular completed questionnaire.

To review, edited responses try to improve on the original responses obtained from measurements of underlying constructs. The edited responses are the data from which inference is made about the values of the construct for an individual respondent.

2.2.5 The Target Population

target population

We are now ready to move to the right side of Figure 2.2, moving from the abstract to the concrete with regard to the representational properties of a survey. The first box describes the concept of a "target population." This is the set of units to be studied. As denoted in Figure 2.2, this is the most abstract of the population definitions. For many U.S. household surveys, the target population may be "the U.S. adult population." This description fails to mention the time extents of the group (e.g., the population living in 2004). It fails to note whether to include those living outside traditional households, fails to specify whether to include those who recently became adults, and fails to note how residency in the United States would be determined. The lack of specificity is not damaging to some discussions, but is to others. The target population is a set of persons of finite size, which will be studied. The National Crime Victimization Survey targets those aged 12 and over who are not in active military service and reside in noninstitutionalized settings (i.e., housing units, not hospitals, prisons, or dormitories). The time extents of the population are fixed for the month in which the residence of the sample person is selected.

2.2.6 The Frame Population

The frame population is the set of target population members that has a chance to be selected into the survey sample. In a simple case, the "sampling frame" is a listing of all units (e.g., people and employers) in the target population. Sometimes, however, the sampling frame is a set of units imperfectly linked to population members. For example, the SOC has as its target population the U.S. adult household population. It uses as its sampling frame a list of telephone numbers. It associates each person to the telephone number of his/her household. (Note that there are complications in that some persons have no telephone in their household and others have several different telephone numbers.) The National Survey on Drug Use and Health uses a sampling frame of county maps in the United States. Through this, it associates each housing unit with a unique county. It then associates each person in the target population of adults and children age 12 or older with the housing unit in which they live. (Note that there are complications for persons without fixed residence and those who have multiple residences.)

Illustration—Populations of Inference and Target Populations

Often, survey statistics are constructed to describe a population that cannot easily be measured. For example, the Surveys of Consumers attempts to estimate consumer sentiment among U.S. adults in a specific month. Each minute, households are being formed through family or rent-sharing arrangements; being dissolved through death, divorce, and residential mobility; being merged together, and so on. The household population of a month is different at the beginning of the month than at the end of the month. Sometimes, the phrase "population of inference" is used for the set of persons who at any time in the month might be eligible. The "target population" describes the population that could be covered, given that the frame is set at the beginning of the month and contact with sample households occurs throughout the month.

2.2.7 The Sample

A sample is selected from a sampling frame. This sample is the group from which measurements will be sought. In many cases, the sample will be only a very small fraction of the sampling frame (and, therefore, of the target population).

sampling frame

2.2.8 The Respondents

In almost all surveys, the attempt to measure the selected sample cases does not achieve full success. Those successfully measured are commonly called "respondents" ("nonrespondents" or "unit nonresponse" is the complement). There is usually some difficulty in determining whether some cases should be termed "respondents" or "nonrespondents," because they provide only part of the information that is sought. Decisions must be made when building a data file about when to include a data record with less than complete information and when to exclude a respondent altogether from the analytic file. "Item missing data" is the term used to describe the absence of information on individual data items for a sample case successfully measured on other items. Figure 2.3 is a visual portrayal of the type of survey and frame data and the nature of unit and item nonresponse.

respondents

non-respondents

unit non-response

item missing data

The figure portrays a data file; each line is a data record of a different sample person. The left columns contain data from the sampling frame, on all sample

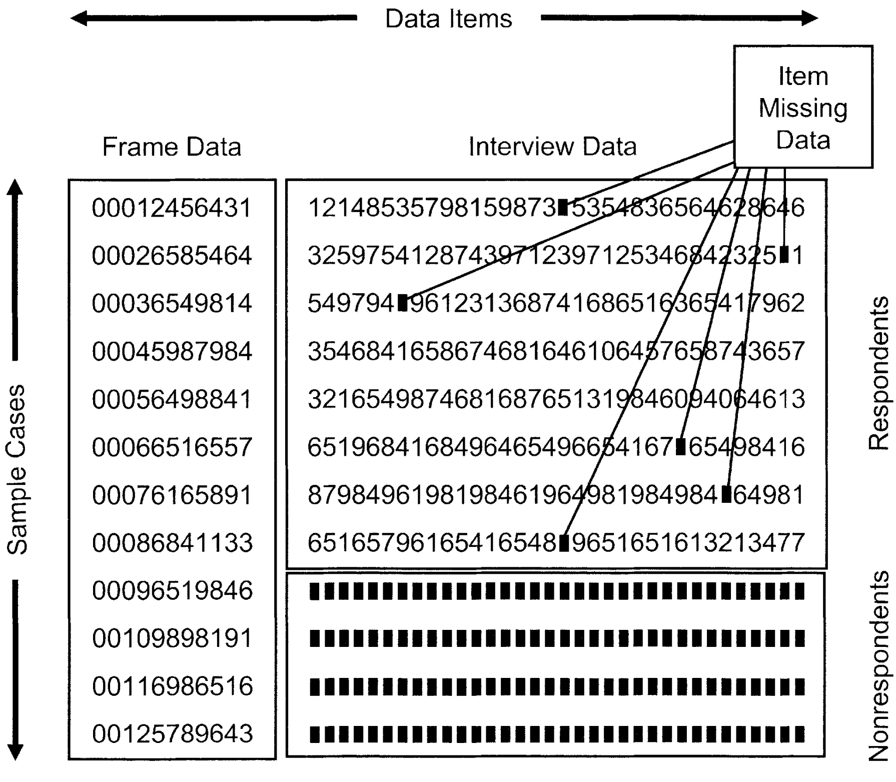

Figure 2.3 Unit and item nonresponse in a survey data file.

cases. Respondents have longer data records containing their answers to questions. The nonrespondents (at the end of the file) have data only from the sampling frame. Here and there throughout the respondent records are some individual missing data, symbolized by a "▮." One example of item missing data from the CES is missing payroll totals for sample employers who have not finalized their payroll records by the time the questionnaire must be returned.

2.2.9 Postsurvey Adjustments

After all respondents provide data and a set of data records for them is assembled, there is often another step taken to improve the quality of the estimates made from the survey. Because of nonresponse and because of some coverage problems (mismatches of the sampling frame and the target population), statistics based on the respondents may depart from those of the full population the statistics are attempting to estimate. At this point, examination of unit nonresponse patterns over different subgroups (e.g., the finding that urban response rates are lower than rural response rates) may suggest an underrepresentation of some groups relative to the sampling frame. Similarly, knowledge about the type of units not included in the sampling frame (e.g., new households in the SOC or new employers in the CES) may suggest an underrepresentation of certain types of target population members. We will learn later that "weighting" up the underrepresented in our calculations may improve the survey estimates. Alternatively, data that are missing are replaced with estimated responses through a process called "imputation." There are many different weighting and imputation procedures, all labeled as "postsurvey adjustments."

weighting

imputation

postsurvey adjustments

2.2.10 How Design Becomes Process

The design steps described above typically have a very predictable sequence. It is most common to array the steps of a survey along the temporal continuum in which they occur, and this is the order of most texts on "how to do a survey."

Figure 2.4 shows how the objectives of a survey help make two decisions, one regarding the sample and another regarding the measurement process. The decision on what mode of data collection to use is an important determinant of how the measurement instrument is shaped (e.g., "questionnaire" in Figure 2.4). The questionnaire needs a pretest before it is used to collect survey data. On the right-hand track of activities, the choice of a sampling frame, when married to a sample design, produces the realized sample for the survey. The measurement instrument and the sample come together during a data collection phase, during which attention is paid to obtaining complete measurement of the sample (i.e.,

mode of data collection

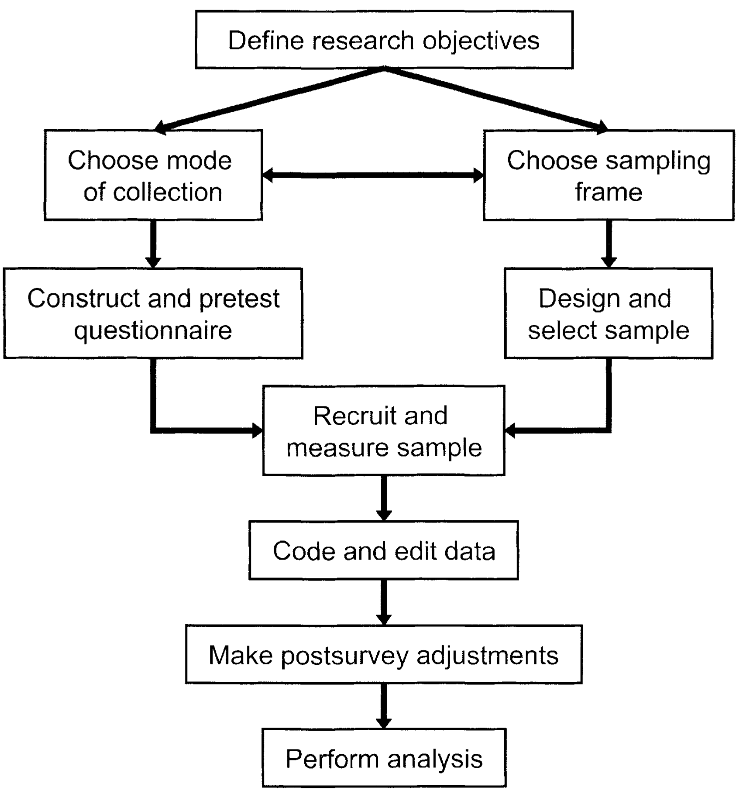

Figure 2.4 A survey from a process perspective.

avoiding nonresponse). After data collection, the data are edited and coded (i.e., placed into a form suitable for analysis). The data file often undergoes some post-survey adjustments, mainly for coverage and nonresponse errors. These adjustments define the data used in the final estimation or analysis step, which forms the statistical basis of the inference back to the full target population. This book takes the perspective that good survey estimates require simultaneous and coordinated attention to the different steps in the survey process.

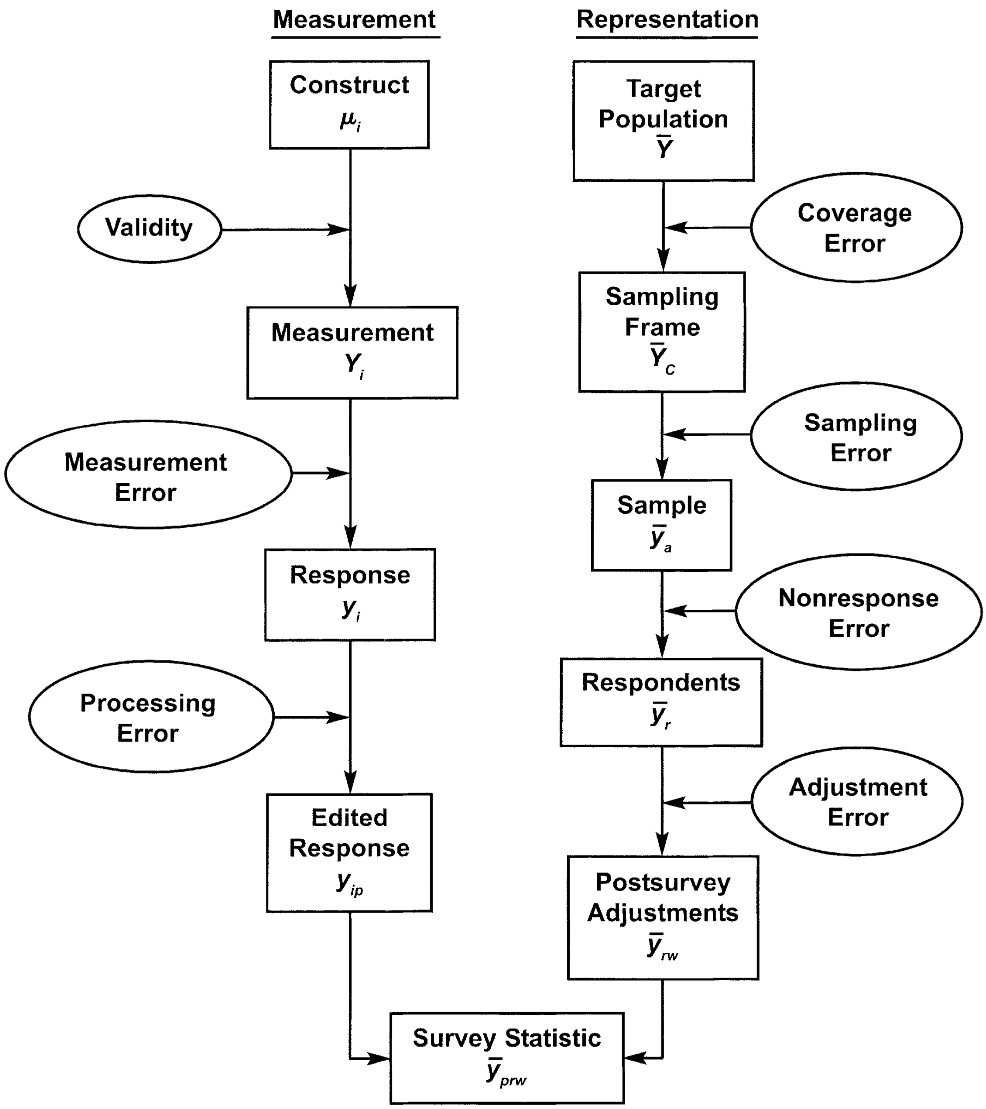

Figure 2.5 Survey life cycle from a quality perspective.

2.3 The Life Cycle of a Survey from a Quality Perspective

We used Figure 2.2 to describe key terminology in surveys. The same figure is useful to describe how survey methodologists think about quality. Figure 2.5 has added in ovals a set of quality concepts that are common in survey methodology. Each of them is placed in between successive steps in the survey process, to indicate that the quality concepts reflect mismatches between successive steps. Most of the ovals contain the word "error" because that is the terminology most commonly used. The job of a survey designer is to minimize error in survey statistics by making design and estimation choices that minimize the gap between two successive stages of the survey process. This framework is sometimes labeled the "total survey error" framework or "total survey error" paradigm.

total survey error

There are two important things to note about Figure 2.5:

1) Each of the quality components (ovals in Figure 2.5) has verbal descriptions and statistical formulations.
2) The quality components are properties of individual survey statistics (i.e., each statistic from a single survey may differ in its qualities), not of whole surveys.

The next sections introduce the reader both to the concepts of different quality components and to the simple statistical notation describing them. Since the quality is an attribute not of a survey but of individual statistics, we could present the statistics for a variety of commonly used statistics (e.g., the sample mean, a regression coefficient between two variables, estimates of population totals). To keep the discussion as simple as possible, we will describe the error components for a very simple statistic, the sample mean, as an indicator of the average in the population of some underlying construct. The quality properties of the sample mean will be discussed as a function of its relationship to the population mean.

We will use symbols to present a compact form of description of the error concepts, because that is the traditional mode of presentation. The Greek letter μ (mu) will be used to denote the unobservable construct that is the target of the measurement. The capital letter Y will be used to denote the measurement meant to reflect μ (but subject to inevitable measurement problems). When the measurement is actually applied, we obtain a response called y (lower case).

The statistical notation will be

μ_i = the value of a construct (e.g., reported number of doctor visits) for the ith person in the population, $i = 1,2,\ldots, N$

Y_i = the value of a measurement (e.g., number of doctor visits) for the ith sample person, $i = 1,2,\ldots, n$

y_i = the value of the response to application of the measurement (e.g., an answer to a survey question)

y_{ip} = the value of the response after editing and other processing steps.

In short, the underlying target attribute we are attempting to measure is μ_i, but instead we use an imperfect indicator, Y_i, which departs from the target because of imperfections in the measurement. When we apply the measurement, there are problems of administration. Instead of obtaining the answer Y_i, we obtain instead y_i, the response to the measurement. We attempt to repair the weakness in the measurement through an editing step, and obtain as a result y_{ip}, which we call the edited response (the subscript p stands for "postdata collection").

2.3.1 The Observational Gap between Constructs and Measures

construct validity

The only oval in Figure 2.5 that does not contain the word "error" corresponds to mismatches between a construct and its associated measurement. Measurement theory in psychology (called "psychometrics") offers the richest notions relevant to this issue. Construct "validity" is the extent to which the measure is related to the underlying construct ("invalidity" is the term sometimes used to describe the extent to which validity is not attained). For example, in the National Assessment of Educational Progress, when measuring the construct of mathematical abilities of 4th graders, the measures are sets of arithmetic problems. Each of these problems is viewed to test some component of mathematical ability. The notion of validity is itself conceptual; if we knew each student's true mathematical ability, how related would it be to that measured by the set of arithmetic problems? Validity is the extent to which the measures reflect the underlying construct.

true value

In statistical terms, the notion of validity lies at the level of an individual respondent. It notes that the construct (even though it may not be easily observed or observed at all) has some value associated with the ith person in the population, traditionally labeled as μ_i, implying the "true value" of the construct for the ith person. When a specific measure of Y is administered (e.g., an arithmetic problem given to measure mathematical ability), simple psychometric measurement theory notes that the result is not μ_i but something else:

$$Y_i = \mu_i + \varepsilon_i$$

That is, the measurement equals the true value plus some error term, ε_i, the Greek letter epsilon, denoting a deviation from the true value. This deviation is the basis of the notion of validity. For example, in the NAEP we might conceptualize mathematics ability as a scale from 0 to 100, with the average ability at 50. The model above says that on a particular measure of math ability, a student (i) who has a true math ability of, say, 57, may achieve a different score, say, 52. The error of that measure

The Notion of Trials

What does it mean when someone says that a specific response to a survey question is just "one trial of the measurement process?" How can one really ask the same question of the same respondent multiple times and learn anything valuable? The answer is that "trials" are a concept, a model of the response process. The model posits that the response given by one person to a specific question is inherently variable. If one could erase all memories of the first trial measurement and repeat the question, somewhat different answers would be given.

of math ability for the ith student is $(52 - 57) = -5$, because $Y_i = 52 = \mu_i + \varepsilon_i = 57 + (-5)$.

One added feature of the measurement is necessary to understand notions of validity: a single application of the measurement to the ith person is viewed as one of an infinite number of such measurements that could be made. For example, the answer to a survey question about how many times one has been victimized in the last six months is viewed as just one incident of the application of that question to a specific respondent. In the language of psychometric measurement theory, each survey is one trial of an infinite number of trials.

Thus, with the notion of trials, the response process becomes

$$Y_{it} = \mu_i + \varepsilon_{it}$$

Now we need two subscripts on the terms: one to denote the element of the population (i) and one to denote the trial of the measurement (t). Any one application of the measurement (t) is but one trial from a conceptually infinite number of possible measurements. The response obtained for the one survey conducted (Y_{it} for the tth trial) deviates from the true value by an error that is specific to the one trial (ε_{it}). That is, each survey is one specific trial, t, of a measurement process, and the deviations from the true value for the ith person may vary over trials (requiring the subscript t, as in ε_{it}). For example, on the math ability construct, using a particular measure, sometimes the ith student may achieve a 52 as above, but on repeated administrations might achieve a 59 or a 49 or a 57, with the corresponding error being +2 or –8 or 0, respectively. We do not really administer the test many times; instead, we envision that the one test might have achieved different outcomes from the same person over conceptually independent trials.

Now we are very close to defining validity for this simple case of response deviations from the true value. Validity is the correlation of the measurement, Y_i, and the true value, μ_i, measured over all possible trials and persons: validity

$$E_{it}\left[(Y_{it} - \bar{Y})(\mu_i - \mu)\right] \Big/ \left[\sqrt{E_{it}(Y_{it} - \bar{Y})^2}\sqrt{E_{it}(\mu_i - \mu)^2}\right]$$

where μ is merely the mean of the μ_{it} over all trials and all persons and $\bar{Y}$ is the average of the Y_{it}. The E at the beginning of the expression denotes an expected or average value over all persons and all trials of measurement. When y and μ covary, moving up and down in tandem, the measurement has high construct validity. A valid measure of an underlying construct is one that is perfectly correlated to the construct. expected value

Later, we will become more sophisticated about this notion of validity, noting that two variables can be perfectly correlated but produce different values of some of their univariate statistics. Two variables can be perfectly correlated but yield different mean values. For example, if all respondents underreport their weight by 5 pounds, then true weight and reported weight will be perfectly correlated, but the mean reported weight will be 5 pounds less than the mean of the true weights. This is a point of divergence of psychometric measurement theory and survey statistical error properties.

2.3.2 Measurement Error: The Observational Gap between the Ideal Measurement and the Response Obtained

measurement error

The next important quality component in Figure 2.5 is measurement error. By measurement error we mean a departure from the true value of the measurement as applied to a sample unit and the value provided. For example, imagine that the question from the National Survey on Drug Use and Health (NSDUH) is "Have you ever, even once, used any form of cocaine?" A common finding (see Sections 5.3.5 and 7.3.7) is that behaviors that are perceived by the respondent as undesirable tend to be underreported. Thus, for example, the true value for the response to this question for one respondent may be "yes," but the respondent will answer "no" in order to avoid the potential embarrassment of someone learning of his/her drug use.

To the extent that such response behavior is common and systematic across administrations of the question, there arises a discrepancy between the respondent mean response and the true sample mean. In the example above, the percentage of persons reporting any lifetime use of cocaine will be underestimated. In statistical notation, we need to introduce a new term that denotes the response to the question as distinct from the true value on the measure, Y_i, for the ith person. We call the response to the question y_i, so we can denote the systematic deviation from true values as $(y_i - Y_i)$. Returning to the CES example, the count for nonsupervisory employees might be 12 for some employer but the response to the question is 15 employees. In the terminology of survey measurement theory, a response deviation occurs to the extent that $y_i \neq Y_i$, in this case, $y_i - Y_i = 15 - 12 = 3$.

response bias

bias

In our perspective on measurement, we again acknowledge that each act of application of a measure is but one of a conceptually infinite number of applications. Thus, we again use the notion of a "trial" to denote the single application of a measurement. If response deviations described above are systematic, that is, if there is a consistent direction of the response deviations over trials, then "response bias" might result. "Bias" is the difference between the expected value (over all conceptual trials) and the true value being estimated. Bias is a systematic distortion of a response process. There are two examples of response biases from our example surveys. In the NSDUH, independent estimates of the rates of use of many substances asked about, including cigarettes and illegal drugs, suggest that the reporting is somewhat biased; that is, that people on average tend to underreport how much they use various substances. Part of the explanation is that some people are concerned about how use of these substances would reflect on how they are viewed. It also has been found that the rates at which people are victims of crimes are somewhat underestimated from survey reports. One likely explanation is that some individual victimizations, particularly those crimes that have little lasting impact on victims, are forgotten in a fairly short period of time. Whatever the origins, research has shown that survey estimates of the use of some substances and of victimization tend to underestimate the actual rates. Answers to the survey questions are systematically lower than the true scores; in short, they are biased. In statistical notation, we note the average or expected value of the response over trials as $E_t(y_{it})$, where, as before, t denotes a particular trial (or application) of the measurement. Response bias occurs when

$$E_t\left(y_{it}\right) \neq Y_i$$

The Notion of Variance or Variable Errors

Whenever the notion of errors that are variable arises, there must be an assumption of replication (or trials) of the survey process. When the estimates of statistics vary over those replications, they are subject to variable error. Variability at the response step can affect individual answers. Variability in frame development, likelihood of cooperation with the survey request, or characteristics of samples, can affect survey statistics.

Usually, variance is not directly observed because the replications are not actually performed.

In addition to systematic underreporting or overreporting that can produce biased reports, there can be an instability in the response behavior of a person, producing another kind of response error. Consider the case of a survey question in the SOC: "Would you say that at the present time, business conditions are better or worse than they were a year ago?" A common perspective on how respondents approach such a question is that, in addition to the words of the individual question and the context of prior questions, the respondent uses all other stimuli in the measurement environment. But gathering such stimuli (some of which might be memories generated by prior questions) is a haphazard process, unpredictable over independent trials. The result of this is variability in responses over conceptual trials of the measurement, often called "variability in response deviations." For this kind of response error, lay terminology fits rather well; this is an example of low "reliability" or "unreliable" responses. (Survey statisticians term this "response variance" to distinguish it from the error labeled above, "response bias.") The difference between response variance and response bias is that the latter is systematic, leading to consistent overestimation or underestimation of the construct in the question, but response variance leads to instability in the value of estimates over trials.

reliability

response variance

2.3.3 Processing Error: The Observational Gap between the Variable Used in Estimation and that Provided by the Respondent

What errors can be introduced after the data are collected and prior to estimation? For example, an apparent outlier in a distribution may have correctly reported a value. A respondent in the National Crime Victimization Survey may report being assaulted multiple times each day, an implausible report that, under some editing rules, may cause a setting of the value to missing data. However, when the added information that the respondent is a security guard in a bar is provided, the report becomes more plausible. Depending on what construct should be measured by the question, this should or should not be altered in an editing step. The decision can affect processing errors.

processing error

Another processing error can arise for questions allowing the respondent to phrase his or her own answer. For example, in the Surveys of Consumers, if a respondent answers "yes" to the question, "During the last few months, have you heard of any favorable or unfavorable changes in business conditions?" the interviewer then asks, "What did you hear?" The answer to that question is entered into a text field, using the exact words spoken by the respondent. For example, the respondent may say, "There are rumors of layoffs planned for my plant. I'm worried about whether I'll lose my job." Answers like this capture the rich diversity of situations of different respondents, but they do not lend themselves to coding

quantitative summary, which is the main product of surveys. Hence, in a step often called "coding," these text answers are categorized into numbered classes. For example, this answer might be coded as a member of a class labeled "Possible layoffs at own work site/company." The sample univariate summary of this measure is a proportion falling into each class (e.g., "8% of the sample reported possible layoffs at own work site/company").

What errors can be made at this step? Different persons coding these text answers can make different judgments about how to classify the text answers. This generates a variability in results that is purely a function of the coding system (e.g., coding variance). Poor training can prompt all coders to misinterpret a verbal description consistently. This would produce a coding bias.

In statistical notation, if we were considering a variable like income, subject to some editing step, we could denote processing effects as the difference between the response as provided and the response as edited. Thus, y_i = response to the survey question, as before, but y_{ip} = the edited version of the response. The processing or editing deviation is simply $(y_{ip} - y_i)$.

2.3.4 Coverage Error: The Nonobservational Gap between the Target Population and the Sampling Frame

The big change of perspective when moving from the left side (the measurement side) of Figure 2.2 to the right side (the representation side) is that the focus becomes statistics, not individual responses. Notice that the terms in Figure 2.5 are expressions of sample means, simple statistics summarizing individual values of elements of the population. Although there are many possible survey statistics, we use the mean as our illustrative example.

finite population

Sometimes, the target population (the finite population we want to study) does not have a convenient sampling frame that matches it perfectly. For example, in the United States there is no updated list of residents that can be used as a sampling frame of persons. In contrast, in Sweden there is a population register, an updated list of names and addresses of almost all residents. Sample surveys of the target population of all U.S. residents often use sampling frames of telephone numbers. The error that arises is connected to what proportion of U.S. residents can be reached by telephone *and* how different they are from others *on the statistics in question*. Persons with lower incomes and in remote rural areas are less likely to have telephones in their homes. If the survey statistic of interest were the percentage of persons receiving unemployment compensation from the government, it is likely that a telephone survey would underestimate that percentage. That is, there would be a coverage bias in that statistic.

undercoverage

Figure 2.6 is a graphical image of two coverage problems with a sampling frame. The target population differs from the sampling frame. The lower and left portions of elements in the target population are missing from the frame (e.g., nontelephone households, using a telephone frame to cover the full household population). This is labeled "undercoverage" of the sampling frame with respect to the target population. At the top and right portions of the sampling frame are a set of elements that are not members of the target population but are members of the frame population (e.g., business telephone numbers in a telephone frame trying to cover the household population). These are "ineligible units," sometimes labeled "overcoverage" and sometimes the existence of "foreign elements."

ineligible units

overcoverage

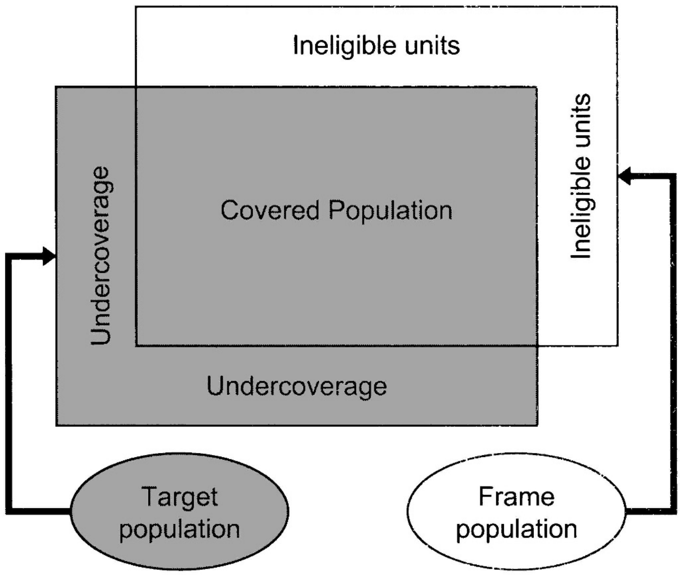

Figure 2.6 Coverage of a target population by a frame.

coverage bias

coverage error

In statistical terms for a sample mean, coverage bias can be described as a function of two terms: the proportion of the target population not covered by the sampling frame, and the difference between the covered and noncovered population. First, we note that coverage error is a property of a frame and a target population on a specific statistic. It exists *before* the sample is drawn and thus is not a problem arising because we do a *sample* survey. It would also exist if we attempted to do a census of the target population using the same sampling frame. Thus, it is simplest to express coverage error prior to the sampling step. Let us express the effect of the mean of the sampling frame:

$\overline{Y}$ = Mean of the entire target population
$\overline{Y}_C$ = Mean of the population on the sampling frame
$\overline{Y}_U$ = Mean of the target population not on the sampling frame
N = Total number of members of the target population
C = Total number of eligible members of the sampling frame ("covered" elements)
U = Total number of eligible members not on the sampling frame ("not covered" elements)

The bias of coverage is then expressed as

$$\overline{Y}_C - \overline{Y} = \frac{U}{N}\left(\overline{Y}_C - \overline{Y}_U\right)$$

That is, the error in the mean due to undercoverage is the product of the noncoverage rate (U/N) and the difference between the mean of the covered and noncovered cases in the target population. The left side of the equation merely shows that the coverage error for the mean is the difference between the mean of the covered population and the mean of the full target population. The right side is the result of a little algebraic manipulation. It shows that the coverage bias is a function of the proportion of the population missing from the frame and the difference between those on and off the frame. For example, for many statistics on the U.S. household population, telephone frames describe the population well, chiefly because the proportion of nontelephone households is very small, about 5% of the total population. Imagine that we used the Surveys of Consumers, a telephone survey, to measure the mean years of education, and the telephone households had a mean of 14.3 years. Among nontelephone households, which were missed due to this being a telephone survey, the mean education level is 11.2 years. Although the nontelephone households have a much lower mean, the bias in the covered mean is

$$\overline{Y}_C - \overline{Y} = 0.05(14.3 \text{ years} - 11.2 \text{ year}) = 0.16 \text{ years}$$

or, in other words, we would expect the sampling frame to have a mean years of education of 14.3 years versus the target population mean of 14.1 years.

Coverage error on sampling frames results in sample survey means estimating the $\overline{Y}_C$ and not the $\overline{Y}$ and, thus, coverage error properties of sampling frames generate coverage error properties of sample-based statistics.

2.3.5 Sampling Error: The Nonobservational Gap between the Sampling Frame and the Sample

One error is deliberately introduced into sample survey statistics. Because of cost or logistical infeasibility, not all persons in the sampling frame are measured. Instead, a sample of persons is selected; they become the sole target of the measurement. All others are ignored. In almost all cases, this deliberate "nonobservation" introduces deviation from the achieved sample statistics and the same statistic on the full sampling frame.

For example, the National Crime Victimization Survey sample starts with the entire set of 3067 counties within the United States. It separates the counties by population size, region, and correlates of criminal activity, forming separate groups or strata. In each stratum, giving each county a chance of selection, it selects sample counties or groups of counties, totaling 237. All the sample persons in the survey will come from those geographic areas. Each month of the sample selects about 8300 households in the selected areas and attempts interviews with their members.

sampling error

sampling variance

sampling bias

As with all the other survey errors, there are two types of sampling error: sampling bias and sampling variance. Sampling bias arises when some members of the sampling frame are given no chance (or reduced chance) of selection. In such a design, every possible set of selections excludes them systematically. To the extent that they have distinctive values on the survey statistics, the statistics will depart from the corresponding ones on the frame population. Sampling variance arises because, given the design for the sample, by chance many different sets of frame elements could be drawn (e.g., different counties and households in the NCVS). Each set will have different values on the survey statistic.

Just like the notion of trials of measurement (see Section 2.3.1), sampling variance rests on the notion of conceptual replications of the sample selection. Figure 2.7 shows the basic concept. On the left appear illustrations of the different possible sets of sample elements that are possible over different samples. The figure portrays S different sample "realizations," or different sets of frame elements, with frequency distributions for each (the x axis is the value of the variable and the y axis is the number of sample elements with that value). Let us use our example of the sample mean as the survey statistic of interest. Each of the S samples produces a different sample mean. One way to portray the sampling variance of the mean appears on the right of the figure. This is the sampling distribution of the mean, a plotting of the frequency of specific different values of the sample mean (the x axis is the value of a sample mean and the y axis is the number of samples with that value among the S different samples). The dispersion of this distribution is the measure of sampling variance normally employed. If the average sample mean over all S samples is equal to the mean of the sampling frame, then there is no sampling bias for the mean. If the dispersion of the distribution on the right is small, the sampling variance is low. (Sampling variance is zero only in populations with constant values on the variable of interest.)

realization

sampling distribution

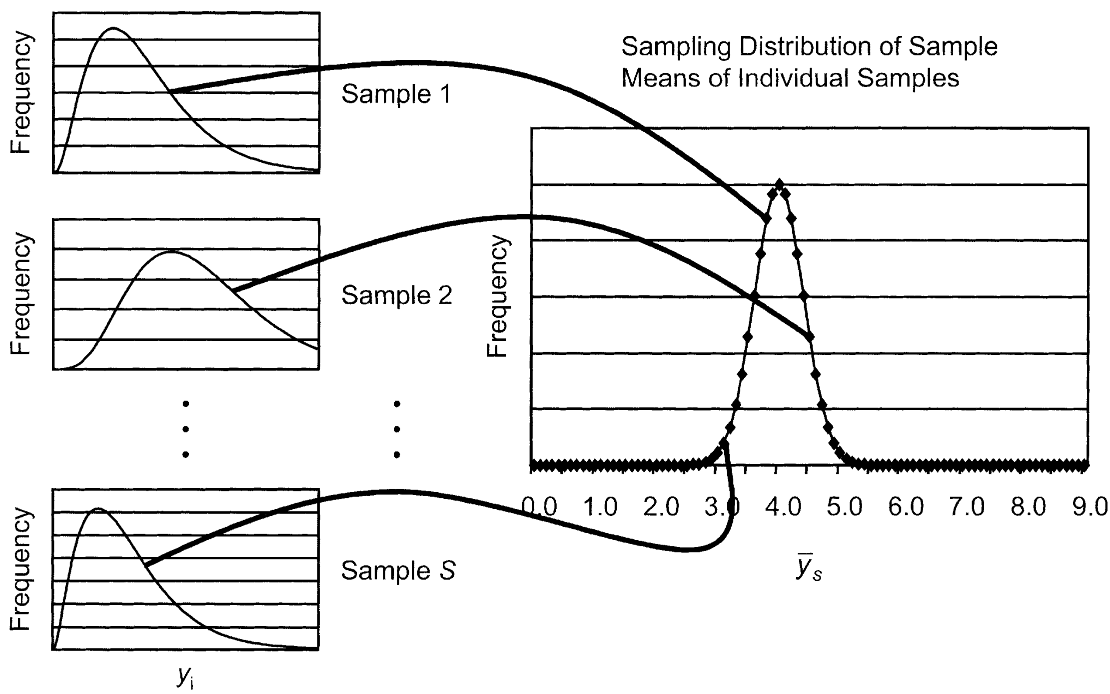

Figure 2.7 Samples and the sampling distribution of the mean.

The extent of the error due to sampling is a function of four basic principles of the design:

probability sampling

stratification

element sample

cluster sample

1) Whether all sampling frame elements have known, nonzero chances of selection into the sample (called "probability sampling")
2) Whether the sample is designed to control the representation of key subpopulations in the sample (called "stratification")
3) Whether individual elements are drawn directly and independently or in groups (called "element" or "cluster" samples)
4) How large a sample of elements is selected

Using this terminology, the NCVS is thus a stratified, clustered probability sample of approximately 8500 households per month.

Sampling bias is mainly affected by how probabilities of selection are assigned to different frame elements. Sampling bias can be easily removed by giving all elements an equal chance of selection. Sampling variance is reduced with big samples, with samples that are stratified, and samples that are not clustered.

In statistical terms,

y_S = Mean of the specific sample draw, sample s; $s = 1, 2,..., S$
Y_c = Mean of the total set of C elements in the sampling frame

These means (in simple sample designs) have the form:

$$\bar{y}_s = \frac{\sum_{i=1}^{n_s} y_{si}}{n_s}, \text{ and } \bar{Y}_C = \frac{\sum_{i=1}^{C} Y_i}{C}$$

sampling variance

"Sampling variance" measures how variable the $\bar{y}_s$ are over all sample realizations. The common measurement tool for this is to use squared deviations of the sample means about the mean of the sampling frame, so that the "sampling variance" of the mean is

$$\frac{\sum_{s=1}^{S} \left(\bar{y}_s - \bar{Y}_C\right)^2}{S}$$

When sampling variance is high, the sample means are very unstable. In that situation, sampling error is very high. That means that for any given survey with that kind of design, there is a larger chance that the mean from the survey will be comparatively far from the true mean of the population from which the sample was drawn (the sample frame).

2.3.6 Nonresponse Error: The Nonobservational Gap between the Sample and the Respondent Pool

Despite efforts to the contrary, not all sample members are successfully measured in surveys involving human respondents. Sometimes, 100% response rates are obtained in surveys requiring sampling of inanimate objects (e.g., medical records of persons, housing units). Almost never does it occur in sample surveys of humans. For example, in the SOC, about 30–35% of the sample each month either eludes contact or refuses to be interviewed. In the main (nationwide) National Assessment of Educational Progress (NAEP), about 17% of the schools refuse participation, and within cooperating schools about 11% of the students are not measured, either because of absences or refusal by their parents. In even years, schools are required by law to participate in reading and mathematics components, grades 4 and 8. As a result, in 2007, there was 100% school participation.

Nonresponse error arises when the values of statistics computed based only on respondent data differ from those based on the entire sample data. For example, if the students who are absent on the day of the NAEP measurement have lower knowledge in the mathematical or verbal constructs being measured, then NAEP scores suffer nonresponse bias, they systematically overestimate the knowledge of the entire sampling frame. If the nonresponse rate is very high, then the amount of the overestimation could be severe.

nonresponse error

Most of the concern of practicing survey researchers is about nonresponse bias, and its statistical expression resembles that of coverage bias described in Section 2.3.1:

$\overline{y}_s$ = Mean of the entire specific sample as selected
$\overline{y}_r$ = Mean of the respondents within the sth sample
$\overline{y}_m$ = Mean of the nonrespondents within the sth sample
n_s = Total number of sample members in the sth sample
r_s = Total number of respondents in the sth sample
m_s = Total number of nonrespondents in the sth sample

The nonresponse bias is then expressed as an average over all samples of

$$\overline{y}_r - \overline{y}_s = \frac{m_s}{n_s}\left(\overline{y}_r - \overline{y}_m\right)$$

Thus, nonresponse bias for the sample mean is the product of the nonresponse rate (the proportion of eligible sample elements for which data are not collected) and the difference between the respondent and nonrespondent mean. This indicates that response rates alone are not quality indicators. High response rate surveys can also have high nonresponse bias (if the nonrespondents are very distinctive on the survey variable). The best way to think about this is that high response rates reduce the *risk* of nonresponse bias.

nonresponse bias

2.3.7 Adjustment Error

The last step in Figure 2.5 on the side of errors of nonobservation is postsurvey adjustments. These are efforts to improve the sample estimate in the face of coverage, sampling, and nonresponse errors. (In a way, they serve the same function as the edit step on individual responses, discussed in Section 2.3.3.)

The adjustments use some information about the target or frame population, or response rate information on the sample. The adjustments give greater weight to sample cases that are underrepresented in the final dataset. For example, some adjustments pertain to nonresponse. Imagine that you are interested in the rate of personal crimes in the United States and that the response rate for urban areas in the National Crime Victimization Survey is 85% (i.e., 85% of the eligible sample persons provide data on a specific victimization), but the response rate in the rest of the country is 96%. This implies that urban persons are underrepresented in the respondent dataset. One adjustment weighting scheme counteracts this by creating two weights, $w_i = 1/0.85$ for urban respondents and $w_i = 1/0.96$ for other respondents. An adjusted sample mean is computed by

$$\bar{y}_{rw} = \frac{\sum_{i=1}^{r} w_i y_{si}}{\sum_{i=1}^{r} w_i}$$

which has the effect of giving greater weight to the urban respondents in the computation of the mean. The error associated with the adjusted mean relative to the target population mean is

$$(\bar{y}_{rw} - \bar{Y})$$

which would vary over different samples and applications of the survey. That is, adjustments generally affect the bias of the estimates and the variance (over samples and implementations of the survey). Finally, although postsurvey adjustments are introduced to reduce coverage, sampling, and nonresponse errors, they can also increase them, as we will learn later in Chapter 10.

2.4 Putting It All Together

The chapter started by presenting three perspectives on the survey. The first, portrayed in Figure 2.2, showed the stages of survey design, moving from abstract concepts to concrete steps of activities. Next, in Figure 2.4, we presented the steps of a survey project, from beginning to end. Finally, in Figure 2.5, we presented the quality characteristics of surveys that the field associates with each of the steps and the notation used for different quantities. The quality components focus on two properties of survey estimates: errors of observation and errors of nonobservation. Errors of observation concern successive gaps between constructs, measures, responses, and edited responses. Errors of nonobservation concern successive gaps between statistics on the target population, the sampling frame, the sample, and the respondents from the sample.

For some of these errors we described a systematic source, one that produced consistent effects over replications or trials (e.g., nonresponse). We labeled these "biases." For others, we described a variable or random source of error (e.g., validity). We labeled these as "variance." In fact, as the later chapters in the book will show, all of the error sources are both systematic and variable, and contain both biases and variances.

The reader now knows that the quantitative products of surveys have their quality properties described in quantitative measures. Look again at Figure 2.5, showing how the notation for a simple sample mean varies over the development of the survey. This notation will be used in other parts of the text. Capital letters will stand for properties of population elements. Capital letters will be used for the discussions about measurement, when the sampling of specific target population members is not at issue. In discussions about inference to the target population through the use of a sample, capital letters will denote population elements and lower case letters will denote sample quantities. The subscripts of the variables will indicate membership in subsets of the population (e.g., *i* for the *i*th person, or the existence of adjustment, such as *w* for weighting).

2.5 Error Notions in Different Kinds of Statistics

The presentation above focused on just one possible survey statistic—the sample mean—to illustrate error principles. There are many other statistics computed from surveys (e.g., correlation coefficients and regression coefficients).

Two uses of surveys, linked to different statistics, deserve mention:

1) Descriptive uses (i.e., how prevalent an attribute is in the population, how big a group exists in the population, or the average value on some quantitative measure)
2) Analytic uses (i.e., what causes some phenomenon to occur or how two attributes are associated with one another)

Many surveys are done to collect information about the distribution of characteristics, ideas, experiences, or opinions in a population. Often, results are reported as means or averages. For example, the NCVS might report that 5% of the people have had their car stolen in the past year.

In contrast, statements such as "Women are more likely to go to a doctor than men," "Republicans are more likely to vote than Democrats," or "Young adults are more likely to be victims of crime than those over 65" are all statements about relationships. For some purposes, describing the degree of relationship is important. So a researcher might say that the correlation between a person's family income and the likelihood of voting is 0.23. Alternatively, the income rise associated with an investment in education might be described by an equation, often called a "model" of the income generation process:

$$\ln(y_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$$

where y_i is the value of the *i*th person's income, and x_i is the value of the *i*th person's educational attainment in years.

The hypothesis tested in the model specification is that the payoff in income of educational investments is large for the early years and then diminishes with additional years. If the coefficient β_1 is positive and β_2 is negative, there is some support for the hypothesis. This is an example of using survey data for causal analysis. In this case, the example concerns whether educational attainment causes income attainment.

Are statistics like a correlation coefficient and a regression coefficient subject to the same types of survey errors as described above? Yes. When survey data are used to estimate the statistics, they can be subject to coverage, sampling, nonresponse, and measurement errors, just as can simple sample means. The mathematical expressions for the errors are different, however. For most survey statistics measuring relationships, the statistical errors are properties of crossproducts between the two variables involved (e.g., covariance and variance properties).

The literatures in analytic statistics and econometrics are valuable for understanding errors in statistics related to causal hypotheses. The language of error used in these fields is somewhat different from that in survey statistics, but many of the concepts are similar.

Thus, the kind of analysis one is planning, and the kinds of questions one wants to answer can be related to how concerned one is about the various kinds of error that we have discussed. Bias, either in the sample of respondents answering questions or in the answers that are given, is often a primary concern for those focused on the goal of describing distributions. If the main purpose of a survey is to estimate the percentage of people who are victims of particular kinds of crime, and those crimes are systematically underreported, then that bias in reporting will have a direct effect on the ability of the survey to achieve those goals. In contrast, if the principal concern is whether old people or young people are more likely to be victims, it could be that the ability of the survey to accomplish that goal would be unaffected if there were consistent underreporting of minor crimes. If the bias was really consistent for all age groups, then the researcher could reach a valid conclusion about the relationship between age and victimization, even if all the estimates of victimization were too low.

2.6 Nonstatistical Notions of Survey Quality

In addition to the components of the total survey error perspective reviewed above, there are three additional notions that are relevant to assessing the quality of a survey estimate. These three notions arise from the overall desire to maximize the "fitness for use" of an estimate. "Fitness for use" acknowledges that different users of the same estimate may have different purposes for the information. For one, highly accurate estimates are necessary for a good decision based on the information; for another, rough orders of magnitude of the population value are sufficient. This viewpoint implies that a "good" estimate for the second user may not be a "good" estimate for the first. High fitness for use means the indicator provides the information needed for the specific use.

fitness for use

credibility

The first notion is "credibility," the extent to which the producer of the information is judged by the user to be free of any particular point of view, a perspective on the phenomena being measured that may influence an outcome of the survey in a known direction. Central government statistical agencies strive to achieve the image of neutrality and objectivity. Scientists using the survey method (a) document each step in their design and implementation, to facilitate replication of their results, then they (b) explicitly note the weaknesses in the estimates that may affect their conclusions. Both of these steps are intended to enhance the credibility of the estimates.

relevance

The second notion is "relevance." A survey estimate is relevant to a user if it measures a construct quite similar in meaning to the user's main concern.

Sometimes, there is gap between the construct measured by the survey and that needed by the user. For example, a user may wish to have a prevalence indicator of economic suffering, the extent to which emotional and physical discomfort arises from economic difficulties among persons. They may use the unemployment rate as an indicator of such suffering. The relevance of this indicator might be criticized by noting that some unemployment does not produce such suffering, due to government support mechanisms. The reader may note a similarity between notions of "construct validity" (see p. 49) and relevance. Relevance focuses on differences among constructs; construct validity relates to differences between a construct and a given measurement.

timeliness

The third notion is "timeliness." One key determinant of whether the survey estimate is fit for a user's purpose is whether the estimate is available at a time needed for the decision based on the information. For example, a survey estimate of the Surveys of Consumers describing the confidence of the U.S. public in March, 2009 is of little value to macroeconomists a year later in March, 2010. Timeliness of an estimate is completely determined by its use.

Indeed, all three of these notions—credibility, relevance, and timeliness—are ones that are well defined only when specific to a particular use of a survey estimate. The notions lie outside the paradigm of total survey error (and we will not discuss them further), and they have not influenced the field of survey methodology in the same way as others. Nonetheless, they are important notions when considering how the same estimates might be used in different ways by different users.

2.7 Summary

Sample surveys rely on two types of inference – from the questions to constructs, and from the sample statistics to the population statistics. The inference involves two coordinated sets of steps: obtaining answers to questions constructed to mirror the constructs, and identifying and measuring sample units that form a microcosm of the target population.

Despite all efforts, each of the steps is subject to imperfections, producing statistical errors in survey statistics. The errors involving the gap between the measures and the construct are issues of validity. The errors arising during application of the measures are called "measurement errors." Editing and processing errors can arise during efforts to prepare the data for statistical analysis. Coverage errors arise when enumerating the target population using a sampling frame. Sampling errors stem from surveys measuring only a subset of the frame population. The failure to measure all sample persons on all measures creates nonresponse error. "Adjustment" errors arise in the construction of statistical estimators to describe the full target population. All of these error sources can have varying effects on different statistics from the same survey.

This chapter introduced the reader to these elementary building blocks of the field of survey methodology. Throughout the text, we will elaborate and add to the concepts in this chapter. We will describe a large and dynamic research literature that is discovering new principles of human measurement and estimation of large population characteristics. We will gain insight into how the theories being developed lead to practices that affect the day-to-day tasks of constructing and implementing surveys. The practices will generally be aimed at improving the quality of the survey statistics (or reducing the cost of the survey). Often, the practices will provide new measurements of how good the estimates are from the survey.

Keywords

bias
cluster sample
coding
construct
construct validity
coverage bias
coverage error
credibility
element sample
error
errors of nonobservation
errors of observation
expected value
finite population
fitness for use
imputation
ineligible units
inference
item missing data
measurement
measurement error
mode of data collection
nonrespondents
nonresponse bias
nonresponse error
observation unit
outlier detection
overcoverage
postsurvey adjustment
probability sampling
processing erro
realization
relevance
reliability
respondents
response
response bias
response variance
sampling error
sampling bias
sampling frame
sampling variance
statistic
stratification
target population
timeliness
total survey error
true values
undercoverage
unit response
validity
weighting

For More In-Depth Reading

Biemer, P. and Lyberg, L. (2003), *Introduction to Survey Quality*, New York: Wiley.

Groves, R. M. (1989), *Survey Errors and Survey Costs*, New York: Wiley.

Lessler, J. and Kalsbeek, W. (1992), *Nonsampling Error in Surveys*, New York: Wiley.

Weisberg, H. (2005), *The Total Survey Error Approach: A Guide to the New Science of Survey Research*, Chicago: University of Chicago Press.

Exercises

1) A recent newspaper article reported that "sales of handheld digital devices (e.g., Blackberries and PDAs) are up by nearly 10% in the last quarter, while sales of laptops and desktop PCs have remained stagnant." This report was based on the results of an on-line survey in which 9.8% of the more than 126,000 respondents said that they had "purchased a handheld digital device between January 1 and April 30 of this year."

   E-mails soliciting participation in this survey were sent to individuals using an e-mail address frame from the five largest commercial Internet service providers (ISPs) in the United States. Data collection took place over a 6-week period beginning May 1, 2002. The overall response rate achieved in this survey was 13%.

   Assume that the authors of this study wanted to infer something about the expected purchases of U.S. adults (18 years old and older).

   a) What is the target population? What is the population in the sample frame?
   b) Based on this chapter and your readings, briefly discuss how the design of this survey might affect the following sources of error:
      - Coverage error
      - Nonresponse error
      - Measurement error
   c) Without changing the duration or the mode of this survey (i.e., computer assisted or self-administered), what could be done to reduce the errors you outlined in 1b? For each source of error, suggest one change that could be made to reduce this error component, making sure to justify your answer based on readings and lecture material.
   d) To lower the cost of this survey in the future, researchers are considering cutting the sample in half, using an e-mail address frame from only the two largest ISPs. What effect (if any) will these changes have on sampling error and coverage error?

2) Describe the difference between coverage error and sampling error in survey estimates.

3) Given what you have read about coverage, nonresponse, and measurement errors, invent an example of a survey design in which attempting to reduce one error might lead to another error increasing. After you have constructed the example, invent a methodological study design to investigate whether the reduction of the one error actually does increase the other.

4) This chapter described errors of observation and errors of nonobservation.

   a) Name three sources of error that affect inference from the sample from which data were collected to the target population.
   b) Name three sources of error that affect inference from the respondents' answers to the underlying construct.

c) For each source of error you mentioned, state whether it potentially affects the variance of estimates, biases the estimates, or both.

5) For each of the following design decisions, identify which error sources described in your readings might be affected. Each design decision can affect at least two different error sources. Write short (2–4 sentences) answers to each point.

a) The decision to include or exclude institutionalized persons (e.g., residing in hospitals, prisons, or military group quarters) from the sampling frame in a survey of the prevalence of physical disabilities in the United States.
b) The decision to use self-administration of a mailed questionnaire for a survey of elderly Social Security beneficiaries regarding their housing situation.
c) The decision to use repeated calls persuading reluctant respondents in a survey of customer satisfaction for a household product manufacturer.
d) The decision to reduce costs of interviewing by using existing office personnel to interview a sample of patients of a health maintenance organization (HMO) and thereby increase the sample size of the survey. The topic of the survey is satisfaction with the medical care they receive.
e) The decision to increase the number of questions about assets and income in a survey of income dynamics, resulting in a lengthening of the interview.
f) The decision to extend interviewing on a survey of use of child care facilities by parents of young children from the originally scheduled period of January 1–May 1, to the new schedule of January 1–August 1.
g) The decision to include prisons and hospitals in the sampling frame for a study of consumer expenditures.
h) The decision to use an existing trained staff of female interviewers (instead of hiring and training some male interviewers) in a survey measuring attitudes toward an amendment to the constitution to provide equal rights under the law to females and males.
i) The decision to change from a face-to-face interview design to a mailed questionnaire mode in a household survey of illegal drug usage.

6) For each of the following questions, state briefly what the construct is that you think the question is most likely designed to measure. In some cases, there may be more than one plausible measurement goal.

a) How old are you?
b) Are you married?
c) Do you own a car?
d) What is your income?
e) Did you vote in the last election for U.S. President?
f) Do you consider yourself to be a Democrat, a Republican, or an Independent?
g) In the next 12 months, do you think that the economy will get better, get worse, or will it stay about the same as it is now?

h) Do you consider yourself to be a happy person?
i) Has a doctor ever told you that you have high blood pressure?
j) How would you rate your doctor in ability to diagnose and propose treatments for medical problems: excellent, good, fair or poor?
k) In the past week, have you prepared any meals for yourself?

7) From an inference perspective, what is the central concern one should have about those who are sampled but do not respond to surveys?