Introduction to Survey Quality. Paul P. Biemer and Lars E. Lyberg

ISBN: 0-471-19375-5

CHAPTER 2

The Survey Process and Data Quality

In this chapter we review the survey process and describe the major sources of error associated with each stage of the process. Then our focus will shift to developing a means for quantifying the level of error in survey data using a measure referred to as the *mean squared error*. This measure of survey accuracy will guide all our efforts throughout this book to identify the major sources of survey error and to reduce them to the extent possible within the budgetary and scheduling constraints of the survey. The mean squared error will also serve as a device for comparing alternative methods in order to choose the best, most accurate method. Thus, the concept of mean squared error as a measure of survey error is fundamental to the study of data quality.

2.1 OVERVIEW OF THE SURVEY PROCESS

As mentioned in Chapter 1, there is an insatiable need today for timely and accurate information in government, business, education, science, and in our personal lives. To understand the present and to plan for the future, data are needed on the preferences, needs, and behaviors of people in society as well as other entities, such as business establishments and social institutions. For many researchers and planners, sample surveys and censuses are major sources of this information.

The word *survey* is often used to describe a method of gathering information from a *sample of units*, a fraction of the persons, households, agencies, and so on, in the population that is to be studied. For example, to measure the size of the workforce the government may ask a sample of people questions about their current employment. A business may use information from a survey to compare the costs of its production against the costs of other similar businesses (see Chapter 1). In this section we present an overview of the process for plan-

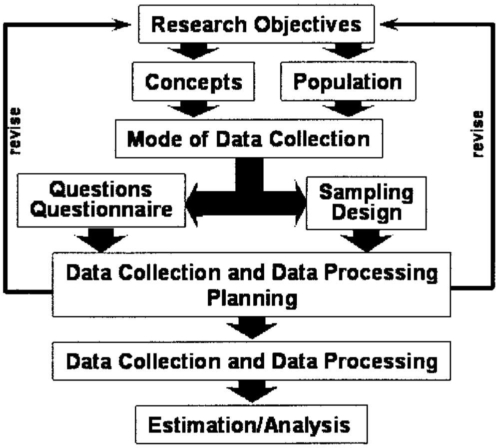


Figure 2.1 Survey process. The planning stages of the survey process are largely iterative. At each stage of planning, new information may be revealed regarding the feasibility of the design. The research objectives, questionnaire, target population, sampling design, and implementation strategy may be revised several times prior to implementing the survey.

ning and conducting sample surveys. Understanding the survey process is central to measuring and controlling survey quality.

As shown in Figure 2.1, the survey process is composed of a number of steps that are executed more or less sequentially, from determining the research objectives to analyzing the data. In what follows, we discuss each major step of the survey process in the context of a hypothetical study that might be commissioned by a government entity to draft new legislation or possibly to evaluate existing legislation. As a means of illustrating the concepts, let us assume that this agency is the U.S. Health Care Finance Administration (HCFA).

HCFA is responsible for administering the Medicare Health Insurance program, which provides health care benefits to U.S. citizens aged 65 and older and citizens with disabilities. Suppose that HCFA is interested in monitoring the health status over time of people who receive Medicare, referred to as *Medicare beneficiaries*. They are particularly interested in measuring the general health of new recipients of Medicare benefits (i.e., recipients who recently reached their sixty-fifth birthday) and how the health characteristics of this population change over time as the population ages and continues to receive Medicare benefits. This study is mandated by the U.S. Congress, which has specified a time frame for starting the study (two years) and a total budget that should not be exceeded for the first three years of the survey. Using this example, we consider the process for designing and conducting a survey to obtain information on the health of this population of older U.S. citizens.

Determining the Research Objectives

The first step in the survey process is to determine the research objectives (i.e., the primary estimates that will be produced from the survey results or the key data analyses that are to be conducted using the survey data). A well-specified set of research objectives is a critical component of the survey process and will facilitate many of the decisions involved in survey design. Defining the research objectives is often accomplished best by identifying a small set of key research questions to be answered by the survey. This is usually done in collaboration with the survey sponsor or researcher(s) commissioning the survey—in this case the survey sponsor is HCFA.

As an example, an important but very general question for Medicare analysts is whether and how the Medicare program contributes to the health and well-being of its beneficiaries. Before a questionnaire can be designed to obtain information on these abstract concepts, a series of steps must be taken. First, HCFA might convene a meeting of experts on the health and well-being of senior adults. The experts would determine the various dimensions of the concepts that should be measured to describe and evaluate the concepts adequately. For example, they may decide that data on food intake, exercise, medical diagnoses, quality of life, and so on, should be collected. They may also identify a number of existing measures, instruments, or scales to assess these concepts that have been validated in other studies and are therefore well understood. The experts may decide further that this information should be collected for all persons 65 years of age and older as they enter the Medicare system and then, following them over time, collected again at periodic intervals to determine how these characteristics change as the beneficiaries age.

The subject matter experts may also recommend that data be collected on visits to the doctor; medications received; current medical conditions; personal characteristics such as height, weight, blood pressure, and functional status of the respondent (i.e., sight, hearing, mobility, mental health); life satisfaction; frequency of depression; and other mental conditions. Further, they may decide that the survey should be repeated for the same sample of persons to determine how these characteristics change over time as the need for medical services increases.

Quite often, the time spent in the development of a comprehensive set of research questions is time saved in the questionnaire design step since eventually each question posed for the research can be linked to one or more data elements to be collected in the data collection phase of the process. These data elements or items are in turn linked to one or more questions on the survey questionnaire or form. In fact, it is good practice to ensure that every question on the questionnaire corresponds to at least one research question, to avoid the situation where questions that are superfluous and really not needed for the purposes of the survey somehow find their way onto the questionnaire. This process of linking research objectives and survey questions also ensures that all survey questions necessary to address the research objectives fully are included in the questionnaire (Table 2.1). As we will see later in this chapter, adherence to this approach will minimize the risk of *specification error* in the results. Specification errors are errors that arise when the survey questions fail

Table 2.1 Correspondence Between Research Questions and Survey Questions[a]

Research Questions	SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7	
RQ1	✓	✓						← SQ7 is an unnecessary question; could be deleted
RQ2	✓		✓					
RQ3				✓				
RQ4					✓	✓		
RQ5								← No questionnaire item to address RQ5

[a] A table such as this is useful for identifying redundant or unnecessary questions in the questionnaire or unaddressed research questions.

to ask respondents about what is essential to answer the research questions (i.e., the subject-matter problem).

Defining the Target Population

The next step in the survey process is to define the population to be studied or the target population for the survey. The target population is the group of persons or other units for whom the study results will apply and about which inferences will be made from the survey results. In the Medicare study, the target population is defined as "persons living in the United States who are aged 65 years or older and are enrolled in the Medicare system." Note that this definition does not include persons under the Medicare system living outside the United States or persons older than 65 years who do not receive Medicare benefits. However, it does include persons enrolled in Medicare whether or not they receive Medicare benefits. Decisions about whom to include and exclude in the target population are important. As we will see later, these decisions guide other important decisions about the survey design in virtually every subsequent stage of the survey process (see also Chapter 3).

Determining the Mode of Administration

Having specified the research objectives and defined the target population, the next step in the process is to determine the mode of administration for the survey. Here we consider whether to use mail questionnaires, telephone, or face-to-face interviewing or some other mode of collecting the data. These decisions must be made before designing the questionnaire, since different modes of data collection often require very different types of questionnaires. The mode of administration will also constrain the sampling design choices that can be used for the survey. Face-to-face interviewing will usually require a sample that is highly clustered (i.e., a sample composed of clusters of units such as persons living within the same neighborhoods). This is done to reduce interviewer travel costs. Telephone and mail survey samples are usually dispersed geographically or unclustered since interviewer travel costs are not incurred for these modes. However, a telephone survey requires either a fairly complete list of telephone numbers for the persons in the target population or a practical and cost-effective method to generate a random sample of telephone numbers that is representative of the target population. A mail survey requires a fairly complete list of addresses for the persons in the target population. If an adequate address list is not available, a mail survey may not be possible.

In deciding on the mode of administration, one of the first constraints one encounters is costs. Face-to-face interviewing, even with highly clustered samples, can be several times more costly than collecting data by telephone or by mail. The budget available for the survey often limits the choices regarding administration mode. Another important consideration relates to the topics to be surveyed. Interviewers can affect the responses to questions on sensitive topics, so if this is a concern, a more private mode of administration such as a mail self-administered questionnaire may be preferable.

One should also consider how important it is to have visual communication with the respondent during the interview. Is it important to use flash cards, for example, to identify the pills and other medications respondents may take? Or are there long lists of medical problems or procedures from which the respondent will be asked to choose? The timing of the survey is also an important consideration in deciding the mode of administration. How quickly are the data needed? If less than two months is available for data collection, a mail survey may not be the best choice.

After some discussion, the Medicare survey design team may determine that a self-administered mail survey is feasible and cost-effective since the questionnaire could be kept simple enough for a sample member to complete without the aid of an interviewer, and since mailing out questionnaires is less expensive than the other modes under consideration. Further, the current addresses of all target population members are available on the Medicare database, so mailing questionnaires to the appropriate addresses would not pose any difficulties.

Finally, to ensure an adequate response rate (about 75% is a typical minimum rate for U.S. government surveys), a telephone follow-up of the mail nonrespondents should also be included as part of the data collection design. Specifying one mode as a primary mode of administration and another mode as a secondary or follow-up mode is a common feature of data collection designs. Referred to as mixed-mode data collection, such strategies are often necessary to maximize response rates for the survey. Additional considerations for determining the best mode of administration are discussed in Chapter 6.

Developing the Questionnaire

The next step of the survey process is the development of the questionnaire or instrument. In this step, the research objectives developed previously are used to determine the data elements to be collected in the survey (i.e., the variables that will be used to address each research question). Each data element corresponds to a single response to a question on the questionnaire. For example, one data element may be the date of birth or a response to a question about medication.

A set of data elements may be used to create a new data element during the postsurvey processing stage. For example, suppose that a research question relates to the mental well-being of Medicare beneficiaries and how this changes over time. Specifically, the researchers may wish to know whether Medicare beneficiaries are generally depressed or contented and how these attributes vary as a person ages. The primary measure for this question is actually a *score*, which is a summary measure derived from a group of data elements on the questionnaire. The score then summarizes the information about a person's mental state into a single measure. The result is a mental health status score which is a single, continuous variable that increases as a person's level of happiness increases and decreases with the onset of depression. Note, however, that this measure requires not just one data element but multiple data elements, all of which provide some information on individual mental health. Thus, it is not uncommon that a number of data elements are needed to address a single research question (see Table 2.1).

As mentioned previously, the design of the questionnaire should also take into account the mode of administration and the capabilities of the target population members to provide information under the desired mode of administration. For example, if in our study a mail self-administered questionnaire is chosen, the design of the questionnaire may use a larger font and incorporate special features to help the oldest respondents complete the questionnaire. These and other considerations for instrument development are discussed further in Chapter 4.

Designing the Sampling Approach

Having defined the target population and the research objectives and determined the mode of administration, the next stage of the survey process can begin, that of specifying the sampling design. The sampling design specification describes the sampling frame (i.e., the list of population members) to be used for the survey, the methods used for randomly selecting the sample from the frame, and the sample sizes that are required. The *sampling frame* is simply the list of target population members from which the sample will be drawn. It may also be a combination of several lists, a map, or any other device that can be used to select the sample. As mentioned previously, the frame chosen for sampling depends to a large extent on the mode of administration for the survey. For our survey, a logical frame is the Medicare list of all persons who are registered in the Medicare program. The coverage of this frame (i.e., the proportion of target population members contained in it) is approximately 100%. This means that every member of the target population has a chance of being selected for the survey. Further, all the information needed to mail the questionnaires to the sample members is available on the Medicare frame.

Since the survey is to be conducted by mail with telephone follow-up of the nonrespondents, interviewer travel costs are not a consideration for the Medicare survey. Thus, the sample could be drawn completely at random without attempting to cluster the sample. However, a sampling plan that involves stratifying the frame into homogeneous groups (e.g., by age) might be used since such a design results in better precision in the survey estimates with no appreciable increase in survey costs.

Finally, after considering the required precision in the estimates for the most important population characteristics to be measured in the study, the sample size is determined. Determining the required sample size for the Medicare survey should take into consideration the loss of sample units that is inevitable as a result of refusals to respond, death, incorrect location information, and loss of sample members resulting from other types of nonresponse.

As shown in Figure 2.1, the process to this point is somewhat iterative. For example, quite often in the process of developing the questionnaire, it is necessary to rethink the survey objectives since to address them all would require a questionnaire or interview that is either longer than can be afforded with the available budget or too burdensome for the sample members, who are thus likely to refuse to participate in the survey. Further, it may be determined that some objectives cannot be addressed adequately with the chosen mode of administration. Consequently, it is necessary either to drop some research questions from the study or to reconsider the mode of data collection.

Similarly, during the sample design development step, it may be realized that an adequate sampling frame does not exist or is too expensive to develop. This could require the use of more than one sampling frame or modifying the definition of the target population to exclude those groups that are too difficult to reach. A common occurrence is that the sample size must be reduced as a result of cost considerations. Thus, several iterations of the foregoing steps of the design process may be necessary before the final design is determined. Additional aspects of the sample design specification are considered in Chapter 9.

Developing Data Collection and Data Processing Plans

Once the initial, basic design decisions are made, the data collection and data processing plans can be developed. These steps involve specifying the process of fielding the survey, collecting the data, converting the data to computer-readable format, and editing the data both manually and by computer. For the Medicare survey, the process would also involve developing procedures for controlling the flow of cases, checking in the mail returns, moving cases to the telephone follow-up operation, keying or scanning the data from paper questionnaires, and merging the data from the mail operation and the telephone operation. Plans are also developed for editing the survey data (i.e., for correcting stray or inappropriate marks on the questionnaire returns, errors that occur during keying or scanning the paper questionnaires, inconsistent responses, and other problems with the data). The structure of the final data files should also be determined so that data analysis would be facilitated.

In the Medicare survey design process, there may be concerns about whether the elderly will complete the forms accurately; whether the response rates will be adequate using the mail mode; how to efficiently handle persons in institutions; whether to accept information from informants other than the sample persons on behalf of the sample persons; and so on. To address these questions and others, the initial design should be tested in a pretest of the survey procedures and questionnaire. The pretest can indicate whether certain aspects of the design do not function well so those aspects of the design can be modified for the main study. As an example, there may be problems in the design of the questionnaire or in the methods used for determining the telephone numbers of the mail nonrespondents for the telephone follow-up operation (see Chapter 10).

Collecting and Processing the Data

The next step of the survey process involves implementing the data collection and data processing plans developed in the previous steps. Interviewers must be recruited, trained, and sent into the field or to a telephone center to collect the data. If the survey is to be conducted by mail, the questionnaires must be mailed and plans for following up nonrespondents must be implemented. Even in a well-planned survey, unforeseen problems can develop which require deviations from the plans. Here it is important for the project staff to monitor carefully the progress of the data collection operations via measurements on key process variables to identify potential problems before they develop into real problems. Thus, an important aspect of the data collection plan is a process for routine monitoring of data collection and obtaining feedback from the supervisory staff. For the Medicare study, this would involve developing the procedures for mailing the questionnaires and checking in the returns, training the telephone interviewers who will contact sample members who do not return their questionnaires, scanning the mail questionnaires into the computer, and conducting quality control operations to ensure that these activities are conducted as planned.

Once the data are in computer-readable form, they can be edited, cleaned, and prepared for estimation and analysis. Editing the data involves correcting out-of-range or inconsistent responses, possibly recontacting respondents to obtain additional information, and generally, cleaning the data of many discernible errors. Information obtained from an open-ended question—that is, a question that elicits an unstructured response—is often converted into code numbers that summarize the verbal information provided by the respondent (see Chapter 7).

Estimation and Data Analysis

Finally, the data are *weighted* to compensate for unequal probabilities of selection, missing data, and frame problems, and the estimates are computed following the plans previously developed for estimation and analysis. Weighting the data essentially involves determining an appropriate multiplier for each observation so that the sample estimates better reflect the true population parameter. The estimation and analysis plan lists the major research questions that should be addressed in the analysis, the estimates that will be computed, and the statistical analyses that will be performed. The latter includes detailed specifications for weighting the data and compensating for nonresponse in the final estimates.

In remaining chapters of the book we discuss many of the decisions that must be made in the survey design process and provide a general background for understanding how these decisions are made. Unfortunately, there are no absolute criteria to dictate the best choice of mode, questionnaire design, data collection protocol, and so on, to use in each situation. Rather, survey design is guided more by past experience, theories, and good advice on the advantages and disadvantages of alternative design choices so that we can make intelligent decisions for each situation we encounter. As will become apparent, the emphasis will be on the general theory of good design rather than on specific guidelines to follow for each set of special circumstances. The aim of good design is to use practical and reliable processes whose outcomes are reasonably predictable. Thus, our guiding philosophy is that it is more useful to learn a few basic techniques for dealing with the underlying causes of survey error and the general theories leading to their development rather than to learn numerous ad hoc methods that essentially treat the same causes of survey error but under a variety of special circumstances.

2.2 DATA QUALITY AND TOTAL SURVEY ERROR

To many users of survey data, data quality is purely a function of the amount of error in the data. If the data are perfectly accurate, the data are of high quality. If the data contain a large amount of error, the data are of poor quality. For estimates of population parameters (such as means, totals, proportions, correlation coefficients, etc.), essentially the same criteria for data quality can be applied. Assuming that a proper estimator of the population parameter is used, an estimate of a population parameter is of high quality if the data on which the estimate is based are of high quality. Conversely, if the data themselves are of poor quality, the estimates will also be of poor quality. However, in the case of estimates, the sample size on which the estimates are based is also an important determinant of quality. Even if the data are of high quality, an estimate based on too few observations will be unreliable and potentially unusable. Thus, the quality of an estimator of a population parameter is a function of the *total survey error*, which includes components of error that arise solely as a result of drawing a sample rather than conducting a complete census called *sampling error components*, as well as other components that are related to the data collection and processing procedures called *nonsampling error components*.

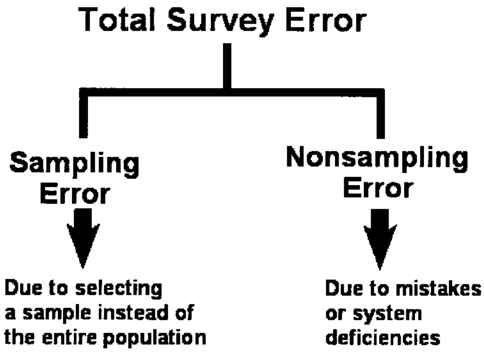


Figure 2.2 Total survey error. Total survey error can be partitioned into two types of components: sampling error and nonsampling error.

In what follows we use the term *estimator* to refer to the formula or rule by which estimates from a survey are produced. For example, an estimator of the population mean for some characteristic in the survey is the sum of the values of the characteristic for all sample members who responded divided by the number of sample members who responded. Suppose that for one particular implementation of the survey design, the value of the estimator is 22. Then 22 is called the *estimate* of the population mean.

Simply stated, the *total survey error* (Figure 2.2) of an estimate is the difference between the estimate and the true value of the population parameter. To illustrate the concept of total survey error, consider a very simple survey aimed at estimating the average annual income of all the workers in a small community of 5000 workers. Thus, the population parameter in this case is the average income over all 5000 workers. Suppose the survey designer determines that a sample of 400 employees drawn at random from the community population should be sufficient to provide an adequate estimate of the population average income. The designer also determines that the best estimator of average annual income is just the simple average of the incomes of the 400 workers in the sample. Thus, the sample is drawn, interviewers are hired and trained, the data are collected, and the sample average is computed from the survey data.

Suppose that average annual income for the persons in the sample is $32,981. Thus, $32,981 is the survey estimate of the population parameter. Finally, suppose that the actual population average income (i.e., the population parameter) for this community is $35,181. This value, of course, is not known since otherwise there would be no need for a survey to estimate it; however, for purposes of this illustration, assume it is known so that we can compute the error in the sample estimate. The difference between the survey estimate of annual income and the unknown true annual income for the community is the total survey error in the estimate of annual income. In this case, the total survey error in the estimate is $32,981 – $35,181 = –$2200.

Total survey error is the difference between a population mean, total, or other population parameter and the estimate of the parameter based on the sample survey (or census).

As noted previously, the true value of the population parameter is not known, but sometimes it can be approximated using the methods discussed in Chapter 8. Therefore, the total survey error in an estimate is also not known but may be approximated using special methods for evaluating surveys. Next, we examine some of the reasons why survey error is unavoidable.

One major reason that a survey estimate will not be identical to the population parameter is *sampling error*, the difference between the estimate and the parameter as a result of only taking a sample of 400 workers in the community instead of the entire population of 5000 workers (i.e., a complete census). Another sample of 400 workers would very likely have different incomes and would therefore produce a different estimate from the first sample estimate. The only way to eliminate the sampling error from the estimation process is to take a complete census of the community. In that case, the average income for all 5000 workers in the "sample" should be the same as the population average income.

However, even if we could afford to observe the entire community in an attempt to measure the true annual income without sampling error, our estimate would not be exactly $35,181 because of another type of error, referred to as *nonsampling error*. Each step of the survey process is a potential source of nonsampling error. *Nonsampling error* encompasses all the various kinds of errors that can be made during data collection, data processing, and estimation except sampling error. The cumulative effect of these errors constitutes the nonsampling error component of the total survey error. In our example, nonsampling errors could arise from the following sources:

- *The respondent.* Respondents may not want to reveal their true income or may unintentionally exclude some sources of income in their response to the survey, such as tips, gifts, bonuses, winnings, and so on.
- *The interviewer.* Interviewers may make mistakes in entering the information on the survey form, or may cause the respondent to make an error, for example, by giving the respondent incorrect information about what to include as income.
- *Refusals to participate.* Some of the 400 persons contacted from the survey may refuse to reveal their incomes or even refuse to participate in the interview.
- *Data entry.* The income values entered on the survey questionnaire may be miskeyed during the data-entry process.

Any and all of these errors could result in the wrong income being recorded and thus cause the estimate of annual income to deviate from the true value of the population parameter. Thus, nonsampling errors can be viewed as mistakes or unintentional errors that can be made at any stage of the survey process. Despite our best efforts to avoid them, nonsampling errors are inevitable particularly in large-scale data collections. Sampling errors, on the other hand, are *intentional* errors in the sense that we can control their magnitude by adjusting the size of the sample. With a sample size of 1, sampling error is at its maximum, and as we increase the sample size to the population size (5000 in our example), sampling error becomes smaller and smaller. When the sample size is the same as the population size (as in a census), the sampling error is zero, and completely absent from the estimates. Thus, sampling error can be made as small as we wish (or can afford) to make it by manipulating the sample size. Later in this chapter we see further illustrations of the sampling error. A more thorough treatment of sampling error is left for Chapter 9.

Nonsampling error, on the other hand, is unpredictable and not so easily controlled. For example, the expected level of nonsampling error may actually increase with increases in the sample size. This may be the result of having to hire a larger staff of interviewers who may be less experienced, who are more prone to certain types of error, or who receive less adequate supervision. Alternatively, the scale of the survey operations may become such that the quality control systems become overloaded and less effective at preventing some types of error. Processing the survey data may be subject to similar control problems, thus resulting in larger data processing errors.

$$\boxed{\text{total survey error} = \text{sampling error} + \text{nonsampling error}}$$

Only in the last 50 years have survey researchers realized that, in many cases, nonsampling error can be much more damaging than sampling error to estimates from surveys. As stated previously, an important goal of this chapter, as well as this book, is to explain how this can happen and why we need to be just as concerned about controlling nonsampling errors in surveys as we are the sampling errors.

There is a considerable literature on nonsampling errors in surveys: the sources and causes of nonsampling error, the design of surveys to minimize them in the final results, statistical methods and models for assessing their effects on the survey results, methods for making postsurvey adjustments to reduce their effects on the estimates, and so on. In this book we try to cover all of these aspects to some extent since the key to survey data quality is understanding the root causes of nonsampling errors and how to minimize them. As mentioned in the preface, our goal is breadth of coverage of these topics, not depth of any specific topic. However, depending on the interests of the reader, more depth of coverage of each topic can be obtained through readings in the extensive literature on survey error, particularly the references that are provided throughout the book.

2.3 DECOMPOSING NONSAMPLING ERROR INTO ITS COMPONENT PARTS

The objective of any survey design is to minimize the total survey error in the estimates subject to the constraints imposed by the budget and other resources available for the survey. As we shall see later in this chapter, reducing nonsampling error while controlling survey costs sometime means increasing sampling error (by reducing the sample size) to reduce some important sources of nonsampling error. Optimizing a survey design means finding a balance between sampling errors and nonsampling errors so that the overall total survey error is as small as possible for the budget available for the survey. This entails allocating the survey resources to the various stages of the survey process so that the major sources of error are controlled to acceptable levels. It does not entail conducting every stage of the process as accurately as possible (without considering the costs involved) since this could result in exceeding the survey budget by a considerable margin.

To stay within the survey budget, training interviewers adequately may require eliminating or limiting the quality control activities conducted at the data processing stage. Increasing the response rate to the survey to an acceptable level may require substantial cuts in the sample size, and so on. How should these decisions be made? Making these trade-offs wisely requires an understanding of the sources of nonsampling error and how they can be controlled. To this end, in the next section, we consider each of the major sources of nonsampling error in surveys in some detail.

2.3.1 The Five Components of Nonsampling Error

Table 2.2 shows a decomposition of nonsampling error into five major sources: specification error, frame error, nonresponse error, measurement error, and processing error. All of the nonsampling errors that we consider in this book can be classified as originating from one of these five sources.

Specification Error

Specification error occurs when the concept implied by the survey question and the concept that should be measured in the survey differ. When this occurs, the wrong parameter is being estimated in the survey, and thus inferences based on the estimate may be erroneous. Specification error is often caused by poor communication between the researcher, data analyst, or survey sponsor and the questionnaire designer. For example, in an agricultural survey, the researcher or sponsor may be interested in the value of a parcel of land if it were sold at fair market value. That is, if the land were put up for sale today,

Table 2.2 Five Major Sources of Nonsampling Error and Their Potential Causes

Sources of Error	Types of Error
Specification error	Concepts
	Objectives
	Data elements
Frame error	Omissions
	Erroneous inclusions
	Duplications
Nonresponse error	Whole unit
	Within unit
	Item
	Incomplete information
Measurement error	Information system
	Setting
	Mode of data collection
	Respondent
	Interview
	Instrument
Processing error	Editing
	Data entry
	Coding
	Weighting
	Tabulation

what would be a fair price for the land? However, the survey question may simply ask: "For what price *would you sell* this parcel of land?" Thus, instead of measuring the market value of the parcel, the question may instead be measuring how much the parcel is worth *to the farm operator*. There may be quite a difference in these two values. The farm operator may not be ready to sell the land unless offered a very high price for it, a price much higher than market value. Since the survey question does not match the concept (or construct) underlying the research question, we say that the question suffers from specification error.

To take this example a step further, suppose that the survey analyst is interested only in the value of the parcel without any of the capital improvements that may exist on it, such as fences, irrigation equipment, airfields, silos, outbuildings, and so on. However, the survey question may be mute on this point. For example, it may simply ask: "What do you think is the current market value of this parcel of land?" Note that this question does not explicitly exclude capital improvements made to the land, and thus the value of the land may be inflated by these improvements without the knowledge of the researcher. A more appropriate question might be: "What do you think is the current market value of this parcel of land? Do not include any capital improvements in your estimate, such as fences, silos, irrigation equipment, and so on."

The question, "What do you think is the current market value of this parcel of land?" is not necessarily a poorly worded question. Rather, it is the wrong question to ask considering the research objectives. A questionnaire designer who does not clearly understand the research objectives and how data on land values will be used by agricultural economists and other data users may not recognize this specification error. For that reason, identifying specification errors usually requires that the questions be reviewed thoroughly by the research analyst or someone with a good understanding of the concepts that need to be measured to address the research objectives properly. The research analyst should review each question relative to the original intent as it relates to the study objectives and determine whether the question reflects that intent adequately. For the land values example, the agricultural economist or other analyst who will use the data on land values would be the best person to check the survey questionnaire for specification errors. In general, detecting specification error usually requires a review of the survey questions by researchers who are responsible for analyzing the data to address the research objectives and who know best about what concepts should be measured in the survey.

Note that in some disciplines (e.g., econometrics), specification error means including the wrong variables in a model, such as a regression model, or leaving important variables out of the model. In our terminology, specification error does not refer to a model but to a question on the questionnaire.

Frame Error

The next source of nonsampling error is error that arises from construction of the sampling frame(s) for the survey. The sampling frame is usually a list of target population members that will be used to draw the sample. In the Medicare survey example above, the frame was the list of all persons receiving Medicare benefits. However, the frame may also be an area map, as in the agricultural land values example, where the sample for the survey is selected by a random selection of parcels of land delineated on the map.

A sampling frame may not even be a physical list, but rather, a conceptual list. For example, telephone survey samples are often selected using a method referred to as random-digit dialing (RDD). For RDD surveys conducted in the United States and Canada, the frame is a conceptual list of all 10-digit numbers that are potential telephone numbers. No physical lists may exist. Instead, telephone numbers are randomly generated as needed using an algorithm for generating random 10-digit numbers.

To ensure that samples represent the entire population, every person, farm operator, household, establishment, or other element in the population should be listed on the frame. Further, to weight the responses using the appropriate probabilities of selection, the number of times that each element is listed on the frame should also be known.

There are a number of errors that can occur when the frame is constructed. Population elements may be omitted or duplicated an unknown number of times. There may be elements on the frame that should not be included (e.g., businesses that are not farms in a farm survey). Erroneous omissions often occur when the cost of creating a complete frame is too high. Quite often, we must live with omissions due to the survey budget. Duplications on the frame are a common problem when the frame is a combination of a number of lists. For the same reason, erroneous inclusions on the frame usually occur because the available information about each frame member is not adequate to decide which entry is in the target population and which is not. In Chapter 3 we discuss how the problems of omissions, erroneous inclusions, and duplications affect the error in a survey estimate.

Nonresponse Error

Nonresponse error, the next error in Table 2.2, is a fairly general source of error encompassing unit nonresponse, item nonresponse, and incomplete response. A *unit nonresponse* occurs when a sampling unit (household, farm, establishment, etc.) does not respond to any part of the questionnaire. For example, a household refuses to participate in the survey, or a mail questionnaire is never returned from an establishment in the survey. *Item nonresponse* occurs when the questionnaire is only partially completed (i.e., some items are skipped or left blank that should have been answered). As an example, in a household survey, questions about household income are typically subject to a great deal of item nonresponse because respondents frequently refuse to reveal their incomes even though they may answer many other questions on the questionnaire. Finally, *incomplete responses* to open-ended questions are also a type of nonresponse error. Here, the respondent may provide some information, but the response is very short and inadequate. As an example, the open-ended question "What is your occupation?" that appears on all labor force surveys around the world is subject to this type of nonresponse. The respondent may provide some information about his or her occupation, but perhaps not enough information to allow an occupation and industry coder to assign an occupation code number later during the data processing stage. This type of error is discussed in detail in Chapter 7.

Measurement Error

Measurement error has been studied extensively and reported in the survey methods literature, perhaps more than any other source of nonsampling error. For many surveys, measurement error is also the most damaging source of error. The key components of measurement error are the respondent, the interviewer, and the survey questionnaire. Respondents may either deliberately or unintentionally provide incorrect information. Interviewers can cause errors in a number of ways. They may falsify data, inappropriately influence responses, record responses incorrectly, or otherwise fail to comply with the survey procedures. The questionnaire can be a major source of error if it is poorly designed. Ambiguous questions, confusing instructions, and easily misunderstood terms are examples of questionnaire problems that can lead to measurement error.

We also consider the errors that arise from the information systems that respondents may draw on to formulate their responses. For example, a farm operator or business owner may consult records that may be in error, and thus cause an error in the data reported. It is also well known that the mode of administration can have an effect on measurement error. For example, information collected by telephone interviewing is, in some cases, less accurate than the same information collected by face-to-face interviewing. Finally, the setting or environment within which the survey is conducted can also contribute to measurement error. For example, for collecting data on sensitive topics such as drug use, sexual behavior, fertility, and so on, a private setting for the interview is often more conducive to obtaining accurate responses than one in which other members of the household are present. In establishment surveys, topics such as land use, loss and profit, environmental waste treatment, and resource allocation can also be sensitive. In these cases, assurances of confidentiality may reduce measurement errors due to intentional misreporting.

These sources of nonsampling error can have a tremendous effect on the accuracy of a survey estimate. To illustrate, consider the previous example of a survey to estimate the income in a community where the unknown, true average income is \$35,181. With a sample of 400 persons, we might expect the error in our estimate due to sampling error to be around \$500. (See Chapter 9 for the details on how this sampling error prediction is constructed.) That is, the estimate from the survey could be as low as \$34,681 and as high as \$35,681. However, as a consequence of nonsampling errors from all the sources described above, the level of error in the survey estimate could be much higher. For example, it is not unreasonable to expect the error to be \$1000—twice the size of the error for sampling alone! As a result, the survey estimate could be as low as \$34,181 and as high as \$36,181 when the true parameter value is \$35,181 (see Figure 2.3).

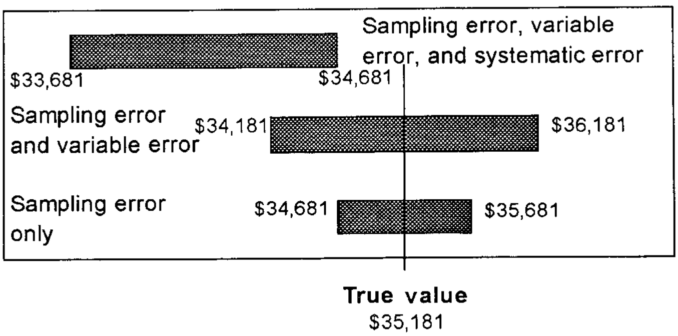


Figure 2.3 Range of estimates produced by a sample survey subject to sampling error, variable error, and systematic error. Shown is the range of possible estimates of average income for a sample of size 400. The range is much smaller with sampling error only, and when systematic nonsampling error is introduced, the range of possible estimates may not even cover the true value.

Even more damaging errors in the estimate can result when the errors of respondents who overreport their incomes do not balance against the errors of respondents underreporting their incomes; that is, if reporting errors tend to be in one direction, which tends to bias the estimate. For example, in the case of income, the negative errors may be the dominant errors since respondents, in general, may have a greater tendency to underreport their income than to overreport it. This type of situation leads to a *negative bias* in the estimates, which means that we expect that the survey estimate will always be lower than the true population parameter value by some unknown amount. In this case, the expected range for the income estimate might be more like \$33,681 to \$34,681 when the actual value is \$35,181. The concepts of biasing *or systematic errors* and nonbiasing or *variable errors* are discussed further in the next section.

Processing Error

The fifth and final source of error in Table 2.2 is processing error, errors that arise during the data processing stage, including errors in the editing of data, data entry, coding, the assignment of survey weights, and the tabulation of survey data. As an example of editing error, suppose that a data editor is instructed to call back the respondent to verify the value of a budget item whenever the value of the item exceeds a specified limit. In some cases, the editor may fail to apply this rule correctly, thus causing an error in the data.

For open-ended items that are coded, coding error is another type of data processing error. The personnel coding the data may make mistakes or deviate from prescribed procedures. The system for assigning the code numbers—for variables such as place of work, occupation, industry in which the respondent is employed, field of study for college students, and so on—may itself be quite ambiguous and very prone to error. As a result, code numbers may be assigned inconsistently and inappropriately, resulting in significant levels of coding error.

The survey weights that compensate statistically for unequal selection probabilities, nonresponse error, and frame coverage errors may be calculated erroneously or there may be programming errors in the estimation software that computes the weights. Errors in the tabulation software may also affect the final data tables. For example, a spreadsheet used to compute the estimates may contain a cell-reference error that goes undetected. As a result, the weights are applied incorrectly and the survey estimates are in error. See Chapters 7 and 9 for details.

2.4 GAUGING THE MAGNITUDE OF TOTAL SURVEY ERROR

As we saw in Section 2.3, the development of a survey design involves many decisions that can affect the total error of a survey estimate. These are decisions regarding the sample size, mode of administration, interviewer training and supervision, design of the questionnaire, and so on, that ultimately will determine the quality of the survey data. Further, these decisions are often influenced by the costs of the various options and their effects on the duration of the survey. A mail survey may be less expensive than a survey conducted by personal visit, but the time allowed for data collection may be such that the mail survey option is not feasible. Face-to-face interviewing may not be affordable due to interviewer costs and other field costs, and less expensive options for collecting the data within the time limits available for the survey must be considered. Telephone interviewing may be both affordable and timely; however, the quality of the data for some items may not be adequate. For example, questions requiring the respondent to consider visual information such as pill cards or magazine covers he or she may have seen are not feasible by telephone. Thus, in determining the design of a survey, one must consider and balance several factors simultaneously to arrive at the design that is best in terms of data quality while meeting the schedule, budget, and other resource constraints for the survey. The resulting design is then a compromise which reflects the priorities attributed to the multiple users and uses of the data.

Making the correct design decisions requires the simultaneous consideration of many quality, cost, and timeliness factors and choosing the combination of design elements that minimizes the total survey error while meeting the budget and schedule constraints. An important aid in the design process is a means of quantifying the total error in a survey process. In this way, alternative survey designs can be compared not only on the basis of cost and timeliness, but also in terms of their total survey error.

As an example, consider two survey designs, design A and design B, and suppose that both designs meet the budget and schedule constraints for the survey. However, for the key characteristic to be measured in the study (e.g., the income question), the total error in the estimate for design A is + or − $3780, while the total error in estimate for design B is only + or − $1200. Since design B has a much smaller error, this is the design of choice, all other things being equal. In this way, having a way of summarizing and quantifying the total error in a survey process can provide a method for choosing between competing designs.

Such a measure would have other advantages as well. For example, suppose that we could establish that most of the error in the survey process under design B is due to nonresponse error. This indicates that nonresponse is the most important source of error for design B, and thus, efforts to improve the quality of the survey data further under design B should focus on the reduction of nonresponse error. To free up resources to reduce nonresponse error, the survey design could consider substituting less expensive procedures for more costly ones in other areas of the survey process. Even though other sources of error may increase by these modifications, the overall effect would be to reduce survey error by the reduction in nonresponse error. In this way, the total error associated with design B could be reduced without increasing the total costs of the survey.

As will become clear in this section, there are many ways of quantifying the total survey error for a survey estimate. However, one measure that is used most often in the survey literature is the total *mean squared error* (MSE). Each estimate that will be computed from the survey data has a corresponding MSE which reflects the effects on the estimate of all sources of error. The MSE gauges the magnitude of total survey error, or more precisely, the magnitude of the effect of total survey error on the particular estimate of interest. A small MSE indicates that total survey error is also small and under control. A large MSE indicates that one or more sources of error are adversely affecting the accuracy of the estimate. As we have said, this information is important since it can influence the way the data are used as well as the way the data are collected in the future should the survey ever be repeated.

One of the primary uses of the MSE is as a measure of the accuracy of the survey data. Unfortunately, it is usually not possible to compute the MSE directly from the survey data, particularly when the data are subject to large nonsampling errors. In most situations, special evaluation studies that are supplemental to the main survey are needed to measure the total MSE. Still, measures of data accuracy are important for the proper interpretation of survey results.

As an example, in the 2000 U.S. population census, a *postenumeration survey* (PES) was conducted following the census to estimate the number of persons missed by the census as well as potentially to use the estimates of number of persons missed for correcting the final census numbers. Special studies were conducted during the census and the PES to measure the MSE of the estimated census total with and without adjustment for the undercount. One important use of the census count is to determine how the 435 seats in the U.S. House of Representatives should be distributed among the 50 states, a process referred to as *congressional apportionment*. The amount of the improvement in the quality of the census counts as measured by the total MSE was an important consideration in the decision not to use the adjusted numbers for apportionment in 2000.

Thus, the concept of a total survey error measure is fundamentally important to the field of survey design and improvement. Indeed, the primary objective of survey design can be stated simply as minimizing the MSEs of the key survey estimates while staying within budget and on schedule. Therefore, the remainder of this chapter is devoted to developing and understanding these critical concepts. In the next section we discuss another way of classifying the nonsampling errors that arise from the survey process: errors that are *variable* and errors that are *systematic*. As we shall see, variable error and systematic error are the essential components of the total MSE since the former determines the variance of a survey estimate and the latter the bias. Later in the chapter we show that the MSE is essentially the sum of variance and bias components contributed by the many sources of error in the survey process.

The *primary objective of survey design* is to minimize the MSEs of the key survey estimates while staying within budget and on schedule.

Variable Errors

In this discussion it will be useful to consider a specific item or question on the survey questionnaire: for example, the income question. For this item, the nonsampling errors that arise from all the various error sources in a survey have a cumulative effect on the survey responses, so that the value of an item for a particular person in a survey is either higher or lower than the true value for the person. In other words, the cumulative effect of the total error for a particular observation is either positive or negative. This is true for all observations: The cumulative effect of all errors will be positive for some persons and negative for others. Suppose that we wish to estimate the mean income for the population using the average of the sample observations (i.e., the sum of the observations in the sample divided by the number of observations). Further suppose that persons in the population are just as likely to make positive errors as they are to make negative errors in reporting their incomes. In this situation, the negative errors will, to some extent, offset the positive errors and the net effect of the errors on the average will be very small. That is, the negative errors in the observations tend to cancel the positive errors, so that nonsampling errors will have essentially no biasing effect on the estimate of the population mean.

Further, if the survey process for collecting income data were to be repeated for the population, a very similar result would occur (i.e., the negative errors would approximately cancel the positive errors). Error sources that produce these types of errors are called *variable error sources* and the errors arising from them are referred to as *variable errors*. When the frequency of variable errors in the data is high, the data are often referred to as *noisy*, since variable error limits our ability to understand what the data are telling us just as a noisy room limits our ability to hear a speaker.

Another concept that is closely related to variable error and often encountered in the survey literature is *data reliability*. *Reliability* refers to the ratio of two types of variation in the observations: the variation in the true values among the population members, and the total variation, which includes the true value variation as well as the additional variation due to variable error. The ratio of these two variances is referred to as the *reliability ratio*. Thus, the reliability ratio ranges from 0.0 to 1.0. *Perfect reliability* occurs when there are no variable errors in the data. In this case, the numerator of the reliability ratio is equal to the denominator and thus the reliability ratio is 1. As the amount of variable error increases, the denominator of the ratio increases and thus the reliability ratio decreases. For example, when the reliability ratio is 0.50 or 50% for a characteristic being measured, the variation in the true values of the characteristic in the population is equal to the variation in the values observed due to variable nonsampling error. This is considered by most standards to be very poor reliability.

In some cases, unreliable data can be recognized by a close examination of the variables that are related in the survey. For example, in an attitudinal survey, the attitudes that a person expresses toward similar issues should show very good agreement. If they do not, this may be a sign of poor reliability on the attitudinal measures. However, in most situations, determining whether the observed data are reliable requires special studies to evaluate the reliability. An example of one type of study to evaluate reliability is a reinterview study, in which the interview is repeated for a sample of households a few days after the original interview. Assuming that the first interview does not influence the responses to the second interview in any way, a comparison of the results of the two interviews will reveal whether the data are reliable. If the data are reliable, there will be good agreement between the first and second responses. Considerable disagreement is an indication of unreliable data. Thus, reliability is often referred to as *test–retest reliability* (referring to first and second measurement), a term that is rooted in the educational psychometric literature (see Lord and Novick, 1968).

Systematic Errors

In many situations, the negative and positive errors do not exactly cancel. For example, positive errors may be much more prevalent than negative errors, and consequently, when the observations are averaged together, the average may be much larger than the true population average. The sample average is then said to exhibit a positive bias. Conversely, the number of respondents in the sample who make negative errors may be considerably larger than the number who make positive errors, and thus the estimate of the mean is smaller than what it would have been without nonsampling error (i.e., it is negatively biased). Errors that do not sum to zero when the sample observations are averaged are referred to as *systematic errors* (Figure 2.4). When the systematic errors are such that the errors in the positive direction dominate (or outnumber) the errors in the negative direction, the sample average will tend to be too high or positively biased. Similarly, when the systematic errors are such that the negative errors dominate, the sample average will be negatively biased.

It is important to note that in our discussion of nonsampling error, the definitions of systematic error and variable error do not refer to what happens in the one *particular* sample that is selected. Rather, they refer to the collection of samples and outcomes of the same survey process over many repetitions under essentially the same survey conditions. This concept of the survey as a repeatable process is similar to the assumptions made in the literature on statistical process control. For example, consider a process designed for the manufacture of some product, say a computer chip. What is important to the designers of the process is the quality of the chips produced by the process over many repetitions of the process, not what the process yields for a parti-

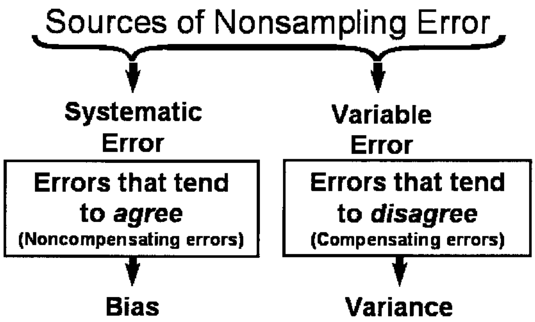


Figure 2.4 Two types of nonsampling error. All nonsampling error sources produce variable error, systematic error, or both. Systematic error leads to biased estimates; variable error affects the variance of an estimator.

cular chip. (Of course, that may be of primary interest to the consumer who purchases the chip!) Similarly, a survey is a process—one that produces data. Although we as consumers of the data are interested primarily in what happens in a particular implementation of the survey, the theory of survey data quality is more concerned about the process and what it yields over many repetitions.

Example 2.4.1 The survey question can be a source of either systematic or variable error in a survey. For example, consider a question that asks about a person's consumption of alcohol in the past week. Respondents may try to estimate their consumption rather than recall exactly the amount they consumed. However, many respondents may deliberately underreport their alcohol consumption to avoid embarrassment, an effect referred to in the literature as *social desirability bias*. As a result of this systematic underreporting, the average amount of alcohol consumed across all sample members will be biased downward, and the estimate of the average amount of alcohol consumed per person will be underestimated.

An example of variable errors occurs when respondents try to estimate events they wish to report accurately, such as the number of trips to the grocery store in the last six months. Rather than try to recall and count the number of trips in six months, many respondents might use some method to estimate the number. For example, some might say that they usually go to the grocery store about twice a week and multiply this number by 26, roughly the number of weeks in the six-month interval. Others may use some other method of estimation. The result is that some respondents may report a number that is slightly higher than the actual number and others may report a slightly lower number than actual. When the entire sample is considered, however, the average number of trips to the grocery store could still be very close to the actual average since the positive and negative errors cancel each other when the data are summed up.

Effects of Systematic and Variable Errors on Estimates

Although both systematic and variable error reduce accuracy, which type of error is more harmful to accuracy? The answer to this question depends on what is being estimated. As we have shown, for *linear estimates* such as estimates of population means, totals, and proportions—in other words, estimates which are sums of the observations in the sample—systematic errors will lead to biases in the estimates, whereas variable errors will tend to cancel one another out and are therefore often nonbiasing. Thus, for linear estimates, systematic errors may be more damaging than variable error, due to their biasing effects. In addition, as we see later, the effect of variable nonsampling errors on linear estimates is very similar to the effect of sampling error on linear estimates; that is, both sampling error and variable nonsampling error can be reduced by increasing the sample size. Therefore, one way to compensate for the effects of variable errors on linear estimates is by increasing the sample size for the survey. However, increasing the sample size will have no effect on systematic error. As mentioned previously, it is possible for the systematic errors to increase as the sample size increases, as a result of increasing the scope of work for the survey and potentially losing some control over the nonsampling error sources.

For *nonlinear estimates* such as estimates of correlation coefficients, regression estimates, standard error estimates, and so on, the answer to the question of what type of error is more damaging is not so simple. For these types of estimates, both systematic errors and variable errors can lead to bias. For example, it can be shown that estimates of regression coefficients are attenuated (i.e., biased toward zero) in the presence of variable error, while for systematic error, the direction of the bias is unpredictable. Therefore, for nonlinear estimates, there is little to choose between systematic error and variable error. Understanding exactly why this is true is not within the scope of this book; however, see Fuller (1987) and Biemer and Trewin (1997) for useful discussions for those interested in pursuing this topic.

Many times, the primary purpose of a sample survey is to report means, totals, and proportions for some target population (i.e., descriptive studies of the population). For this reason, in designing surveys for the reduction of total error, priority is usually given to the identification and elimination of the major sources of systematic error. Although the goal of survey design is the minimization of both types of error, the survey designer often must decide which types of errors are most damaging and control for those while other types of errors are either ignored or controlled much less.

For example, the designer may have to decide whether it is better to allocate more survey resources on interviewer training than on further refinement of the questionnaire. Another decision might be whether it is better to devote more survey resources to the follow-up of survey nonrespondents than to spend those resources on more extensive quality control checks for data-entry errors. In these situations, it is useful to have some idea as to whether a particular error source produces predominately systematic error or variable error.

In most situations, eliminating the source that produces systematic error should take priority over the error sources where the primary risk is variable error. However, there are no hard-and-fast rules about this. Nevertheless, in our subsequent discussion of the error sources in the survey process, some consideration of the risk of each systematic error from the error source will be useful.

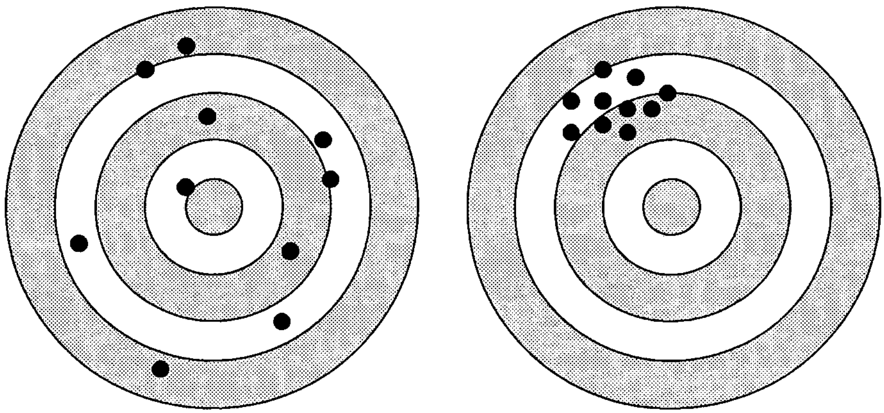
Error Sources Can Produce Both Systematic and Variable Errors

Some error sources produce errors that are primarily systematic. For example, nonresponse error is primarily a systematic error rather than a variable error since the cumulative effect of the nonresponse error on a particular survey estimate is to bias the estimate. As we shall see in Chapter 3, the magnitude of the bias is a function of the nonresponse rate and the difference in the characteristics under study between the average respondent and the average nonrespondent. Frame noncoverage error behaves in much the same way. If a particular type of population member is missing from the frame—for example, that small farms are missing in most agricultural survey frames—repeated implementations of the survey process using the incomplete frame will tend to err in the same direction.

Some error sources produce errors that are primarily variable error. For example, keying error is typically variable error, since errors data keyers make tend to be haphazard and omnidirectional. Rarely will a group of keyers make errors that tend to either increase or decrease the value of an estimate. Rather, for the most part, these types of errors will tend to cancel one another out. Another example used earlier is respondent estimation, as in estimating the number of trips to the grocery store. Some respondents may guess high while others may guess low, so that, on balance, the average of the guesses may be very near zero.

Other error sources produce both variable error and systematic error. For example, the errors committed by interviewers in carrying out their assignments may be a combination of variable and systematic errors. Consider the income question again. Some interviewers, by the mannerisms, dress, comments made earlier in the interview, and so on, may have a tendency to elicit responses that are higher than the true incomes, while other interviewers may tend to have just the opposite influence. However, particularly at lower income levels, the general tendency in the population may be to overstate actual income as a result of the respondents wanting to appear better off than they really are. Thus, although the overall tendency (over many repetitions of the survey process) is to overreport income, interviewers also provide a variable error component so that the income data are both biased and unreliable.

In the next section we discuss a method for summarizing the combined effects of variable errors and systematic errors on a survey estimate using the mean squared error of the estimate.



(*a*) **Large variance and small bias** (*b*) **Large bias and small variance**

Figure 2.5 Systematic and variable error expressed as targets. If these targets represented the error in two survey designs, which survey design would you choose? The survey design in part (*a*) produces estimates having a large variance and a small bias, while the one in part (*b*) produces estimates having a small variance and a large bias.

2.5 MEAN SQUARED ERROR

Analogy of the Marksman and the Target

To better understand how total survey error can be decomposed into components for systematic error and variable error—or, equivalently, bias and variance—consider the picture in Figure 2.5. In this figure we use the process of shooting at a target to illustrate the process of using a survey to estimate some parameter of the population. The bull's-eye on the target represents the population parameter we wish to estimate with the survey data. Of course, in practice, there may be many parameters that we wish to estimate with a survey, but for now we concentrate on a particular parameter, such as the average income for the population. Conducting a survey to estimate the parameter is analogous to a marksman taking aim and shooting at the bull's-eye on the target. If the marksman's aim is accurate, he or she scores a bull's-eye; otherwise, he or she misses the bull's-eye by some distance. The distance between the point where the marksman hits the target and the bull's-eye is the total error in the marksman's aim. Similarly, if we conduct a survey, the goal is to estimate the population parameter exactly, but we miss the parameter because of survey error. The "distance" between the estimate from the survey and the population parameter is the total error in the estimate and is analogous to the total error in the marksman's aim.

Now suppose the marksman shoots repeatedly at the target, each time aiming at the bull's-eye. That is analogous to repeating the survey process a number of times under the same conditions, each time attempting to estimate the population parameter. In practice, we would implement the survey process only once to estimate the population parameter. However, if we could repeat the survey many times, the variation we observed in the survey estimates would tell us something about the total error in the survey process, just as the pattern of hits on the target tells us something about the error in the marksman's aim.

Now the error in a survey estimate, like the marksman's aim, consists of systematic and variable error components. For example, if the marksman's sights are not properly adjusted, he or she will probably miss the bull's-eye. Further, if his or her aim is consistent, the distance between each hit and the bull's-eye will be roughly the same, due to this sight misalignment. The marksman's sight misalignment is analogous to a biased survey process. The survey process may produce very consistent results each time it is implemented; however, the estimates all differ from the parameter value by roughly the same amount and in the same direction.

In addition to sight misalignment, the marksman may miss the target for a number of other reasons. For example, the marksman's aim may not be steady, and therefore each time he or she shoots, the bull's-eye will be missed by an unpredictable, random amount. Sometimes the hit is to the left of the bull's-eye, and other times the hit may veer to the right, and above or below the bull's-eye. Other factors, such as the wind or weather, the shape of the projectiles being fired, and the weapon itself may also have unpredictable, random effects on the accuracy of each shot. These factors are analogous to the variable errors in a survey. Each time the survey process is repeated, random variation due to a whole host of factors may affect the accuracy of the estimate and add to the total error (the distance between the hit and the bull's-eye).

The two targets in Figure 2.5 could correspond to two different marksmen with two different weapons. Note that the pattern of hits on the target on the left suggests that systematic error may be a problem for that marksman; that is, there is something inherently wrong with either the weapon or some other aspect of the shooting process that affects all the shots at the target in the same way. The pattern of hits for the left target suggests that the systematic error is smaller, but variable error is a problem. That is, the cumulative effect of many factors associated with shooting at the target causes the marksman to miss the target in seemingly random ways.

These targets can be used not only to help us understand systematic error (bias) and variable error (variance) in a survey process, but also how to measure them. Suppose that we were to repeat the same survey process many times under the very same conditions. That is, we use the same sampling procedure (but a different sample of respondents at each replication), the same questionnaire, the same process for hiring and training interviewers, the same data collection procedures, and so on. Each replication of the survey will yield one estimate of the population parameter represented by a hit on the target. The target on the right corresponds to one type of survey process for estimating the parameter, and the target on the left represents another survey process for estimating the same parameter. Thus, the two survey processes produce a different mixture of systematic and variable error components. The survey corresponding to the left target has a considerable systematic error; however, the variance (i.e., variable error) for that process is relatively small. The survey corresponding to the right target has a large variable error, but the systematic error is small.

Now suppose that both survey processes have approximately the same cost. Which survey process should be chosen for estimating the population parameter: the survey corresponding to the left target or the one corresponding to the right target? In making this decision, it would be very helpful if there were some way of quantifying the accuracy of the two surveys by combining the effects of systematic and variable error into a single dimension; in other words, a way of summarizing the total error into a single number. Then the survey process producing the smaller level of total survey error would be the preferred survey process.

Summarizing the Total Error of an Estimator Using the Mean Squared Error

There are many ways to summarize error in processes that produced the patterns or more generally, the total error in a survey process, depicted in Figure 2.1. However, one that is used in the statistical literature universally because of its favorable statistical properties is the *mean squared error*, a measure of the average closeness of the hits to the bull's-eye, where *closeness* is defined as the squared distance between a hit and the bull's-eye. To compute the mean squared error, we measure the distance between each hit and the bull's-eye, square that distance, and average these squared distances across all the hits.

As an example, consider the situation where each survey process, process A and process B, has been repeated 10 times, each time under identical conditions. For these 20 implementations, the error in the estimates for survey process A (distances for the left target) and survey process B (the right target) is shown in Figure 2.5. The average of the squared distances of the 10 hits for the right target is 0.15 and the average for the left target is 4.5. Therefore, by the mean squared error criterion, the survey process represented by the left target is preferred because the mean squared error of the estimate is smaller.

> The *mean squared error* (MSE) of the marksman's aim is the average squared distance between the hits on the target and the bull's-eye. The MSE of a survey estimate is the average squared difference between the estimates produced by many hypothetical repetitions of the survey process (corresponding to the hits on the target) and the population parameter value (corresponding to the bull's-eye).

To put this illustration more concretely into a survey context, suppose that the right target represents a survey process where the data are collected from respondents by interviewers (i.e., interviewer-assisted mode) and the left target represents a survey process where the same data are collected by respondents recording their responses directly on the questionnaire (self-administered mode). Let us assume that both survey processes are based on the same sample size so that any differences in variance are due strictly to nonsampling variable error. Suppose that the interviewer-assisted mode has a larger bias, due to the influencing effect of the interviewer on the respondent; however, the variance is smaller as a result of reduction of respondent comprehension due to the assistance provided by the interviewer. Suppose further that the self-administered mode eliminates the bias resulting from interviewer influence but the process introduces larger variable error as a result of respondents interpreting the questions in different ways. Thus, the error in the interviewer-assisted survey resembles the target on the right in Figure 2.5, and the error in the self-administered mode resembles the left target. This example illustrates how two survey processes based on different modes but aimed at collecting the same information can have very different bias and variance characteristics.

A key aspect of the definition of the mean squared error given above is based on the idea that the same survey process is repeated many times for the same population and under the same survey conditions. The mean squared error is then the average squared difference between the estimate from each replication of the process and the population parameter value. There are two problems with this definition. First, it is usually impossible to repeat the survey process under identical conditions each time. Our world is dynamic and ever changing and the survey conditions may also change considerably from one implementation to the next. Many factors (the weather, politics, etc.) influence the outcomes of surveys, and these factors vary over time. The population parameter value itself may change over time. Further, it would not be cost-effective to repeat the survey process multiple times, or even twice, for the purpose of estimating the MSE. There are more efficient and effective ways of estimating the components of the MSE, and many of these are discussed in Chapter 8.

Another difficulty encountered in trying to assess the MSE of a survey estimate is determining the true value of a population parameter so that the error in the estimate can be quantified. Simply conducting another survey to estimate the parameter is not likely to be sufficient unless special procedures are put in place to ensure that the estimate is highly accurate. Other methods for assessing the true values of the characteristic for the sample members could involve the use of very accurate administrative records or the use of more expensive and elaborate measurement devices that eliminate most of the error inherent in the original survey process. Usually, these methods are conducted after the original survey data have been collected. In Chapter 8 we discuss these methods in some detail.

Estimation of the MSE is usually a complex and costly process, and therefore the total MSE is seldom estimated in practice. When it is estimated, the result is often only a rough approximation of the actual MSE. In addition, quite often only a few of the most important components of the MSE are estimated. For example, a typical approach to computing the MSE involves estimating several bias and/or variance components associated with the major sources of systematic and variable errors and then combining this error information to produce the mean squared error. The computational formula described above does not lend itself readily to computing the mean squared error in this manner. In the next section we describe an alternative method for computing the mean squared error that addresses this shortcoming.

Decomposing the Mean Squared Error into Bias and Variance Components

Let us revisit the targets in Figure 2.5 to discuss another way to view the total error in a survey estimate. Note that for each target, the hits form a cluster of points on the target. Consider the target on the right first and locate the point that is approximately the center (or centroid) of the hits or estimates for this target. The point in the center of the cluster represents the mean or average error of the survey process. The distance between the center of the cluster of hits and the bull's-eye is the systematic error or bias in the survey process. Further, the distances between the individual hits and the cluster center represent the variable errors in the survey process.

Thus, the bias in a survey estimate can be computed by computing the average of the estimates produced by repeating the survey process many times under the same conditions, averaging these estimates, and then subtracting from this average the value of the true population parameter. The variance of the estimator from the survey can be computed as the average squared distance between each survey estimate from the repetitions and the average survey estimate.

This discussion suggests that there are two ways to compute the mean squared error of a survey estimator:

METHOD 1

MSE = average squared distance between the hits and the bull's-eye, or, in survey terms,
= average squared difference between the survey estimates from repeating the survey many times and the true population parameter value

METHOD 2

MSE = squared distance between the center of the hits and the bull's-eye + average squared distance between the hits and the center of the hits, or, in survey terms,
= squared distance between the average value of the estimator over replications of the survey process and the true population parameter value + average squared difference between the estimates from the replications and the average value of the replications

Both methods of computing the mean squared error will produce the same value; however, method 2 is often preferred because it decomposes the mean squared error into two components: squared bias and variance. The *squared bias* is the squared distance between the average value of the estimator over replications of the survey process and the true population parameter value. The *variance* is the average squared difference between the estimates from the replications and the average value of the replications. Therefore, another way of writing the formula for method 2 is

$$\text{MSE} = \text{Squared bias} + \text{Variance} \tag{2.1}$$

This formula says that the total MSE for an estimate is equal to the bias squared plus the variance. If we know that the bias of an estimate is 0, the MSE is simply the variance of the estimate. However, when the bias is not zero, the bias must be estimated in order to compute an estimate for the MSE. Computation of the bias requires knowledge of the true parameter value; however, the variance can be computed without knowing the true parameter value.

To see this, consider Figure 2.6, where we show the hits on the targets in Figure 2.5 without the targets. Notice that we can still compute the variance, which is the squared distance from each hit to the center or average of the hits. However, we cannot compute the bias since there is no bull's-eye or true parameter value. This figure also illustrates the fallacy of choosing between two survey processes on the basis of variance alone. Note that the survey process on the right would always be chosen over the survey process on the left, regardless of the bias in that design, since the variance is all that is known.

To illustrate that both methods for computing the MSE will yield the same results, consider the data in Tables 2.3 and 2.4. In Table 2.3 we compute the MSE for the two survey designs in Figure 2.5 using method 1. We measured the distance from each hit in Figure 2.5 to its corresponding bull's-eye and then squared these distances. The average squared distance for each target is the MSE for the corresponding process. For the survey process on the left, this

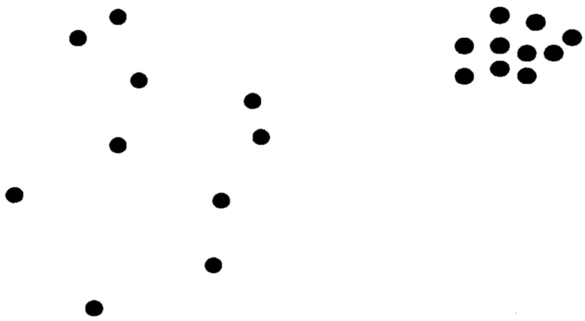


Figure 2.6 Truth is unknown. Determining the better survey process when the true value of the population parameter is unknown is like trying to judge the accuracy of two marksmen without having a bull's-eye. All that is known about the sets of survey estimates is that the survey design in part (*a*) produces a large variance and the survey design in part (*b*) produces a small variance.

Table 2.3 Computation of the Mean Squared Error Using Method 1[a]

Left Target			Right Target		
Hit	Distance from Hit to Bull's-eye	Squared Distance to Bull's-eye	Hit	Distance from Hit to Bull's-eye	Squared Distance to Bull's-eye
1	2.2	4.8	1	3.1	9.6
2	−3.6	13.0	2	3.7	13.7
3	−4.5	20.3	3	5.3	28.1
4	6.8	46.2	4	4.9	24.0
5	5.1	26.0	5	6.1	37.2
6	−7.2	51.8	6	4.4	19.4
7	−3.9	15.2	7	2.8	7.8
8	5.3	28.1	8	6.1	37.2
9	−1.8	3.2	9	4.5	20.3
10	3.1	9.6	10	4.1	16.8
Avg or center	0.15	21.8 (= MSE)	Avg or center	4.5	21.4 (= MSE)

[a] The mean squared error is the average squared distance from the hit (or estimate) to the bull's-eye (or parameter value). For method 1 we compute the distance (or error) from the hit to the bull's-eye for each hit, square the result, and then average over all 10 hits.

average is 21.8, and for the process on the right it is 21.4 (see the last row in Table 2.3).

In Table 2.4 the MSE is computed using method 2. For this method we locate the center of the hits and measure the distance between each hit and the center of the hits and square the result. The average of these squared distances (i.e., 21.81 for the left target and 1.17 for the right target) is the

Table 2.4 Computation of the Mean Squared Error Using Method 2[a]

Left Target			Right Target		
Hit	Distance from Hit to the Center of the Hits	Squared Distance from Center	Hit	Distance from Hit to the Center of the Hits	Squared Distance from Center
1	(2.2 – 0.15) = 2.05	4.20	1	(3.1 – 4.5) = –1.40	1.96
2	(–3.6 – 0.15) = –3.75	14.06	2	(3.7 – 4.5) = –0.80	0.64
3	–4.65	21.62	3	0.80	0.64
4	6.65	44.22	4	0.40	0.16
5	4.95	24.50	5	1.60	2.56
6	–7.35	54.02	6	–0.10	0.01
7	–4.05	16.40	7	–1.70	2.89
8	5.15	26.52	8	1.60	2.56
9	–1.95	3.80	9	0.00	0.00
10	2.95	8.70	10	–0.40	0.16
Avg	0.0	21.81 (= Variance)	Avg	0.0	1.17 (= Variance)
Bias = (0.15 – 0.0) = 0.15		Bias^2 = 0.023	Bias = (4.5 – 0.0) = 4.5		Bias^2 = 20.25
MSE = Bias^2 + Variance = 21.8			MSE = Bias^2 + Variance = 21.4		

[a] The mean squared error can be computed in two stages. First, measure the distance from each hit to the center of the hits and square the result. Average these square distances over all 10 hits. Finally, compute the distance between the center of the hits to the bull's-eye and square the result. The MSE is the sum of these two quantities. Note that these results agree with the results from Table 2.3, demonstrating that the two methods yield the same results.

variance for the estimate. Note that the variance for the left target is many times larger than the variance for the right target. Then the distance from the center of the hits and the bull's-eye is measured to produce 0.15 for the left target and 4.5 for the right target. As we have already observed, the right target shows a much larger bias than the left target. Putting the variance and bias components together, we have MSE (left target) = $(0.15)^2 + 21.81 = 21.83$ and MSE (right target) = $(4.5)^2 + 1.17 = 21.42$. This agrees with the MSEs we computed in Table 2.3. Thus, we have demonstrated that the two methods for computing the MSE are equivalent.

Major Components of the MSE

Each source of error in Table 2.2 can contribute both bias and variance components to the total MSE; however, some error sources pose a greater risk for bias, some for variance, and some error sources can contribute substantially to both bias and variance. Table 2.5 lists each major error source along with an assessment of the risk of variable error and systematic error for each. These risks will depend on the specifics of the survey design and the population to be surveyed, and there are no hard-and-fast rules regarding which sources of error are more problematic for systematic error or variable error. Nevertheless, Table 2.5 provides an indication of the risk for a typical survey.

Table 2.5 Risk of Variable Errors and Systematic Errors by Major Error Source

MSE Component	Risk of Variable Error	Risk of Systematic Error
Specification error	Low	High
Frame error	Low	High
Nonresponse error	Low	High
Measurement error	High	High
Data processing error	High	High
Sampling error	High	Low

For example, in a typical survey using acceptable random sampling methods, the risk of bias resulting from sampling error is quite small, and sampling variance is inevitable. Conversely, for specification error, the error in the estimate is primarily bias, since errors in the specification of the survey question is more likely to lead to systematic error than variable error. Nonresponse error also poses a greater risk to nonsampling bias than to variance; although some nonresponse adjustment methods can contribute substantially to the variance when the nonresponse rate is quite high. Frame error, particularly error due to population members missing from the frame, is viewed primarily as a biasing source of error. However, as we see later in this book, measurement error and data processing error can pose a risk for both bias and variance in the survey estimates.

Using Table 2.5 as a guide, we can write an expanded version of the MSE equation in (2.1). The squared-bias component can be expanded to include bias components for all the sources of error in the table that have a high risk of systematic error (i.e., specification bias, B_{SPEC}; nonresponse bias, B_{NR}; frame bias, B_{FR}; measurement bias, B_{MEAS}; and data processing bias, B_{DP}). Note that these components of bias sum to produce the total bias component, called *bias*. Similarly, the variance component is the sum of the components for the major sources of variance in the table (i.e., sampling variance, Var_{SAMP}; measurement variance, Var_{MEAS}; and data processing variance, Var_{DP}. Thus, the expanded version of the MSE formula showing components for all the major sources of bias and variance is as follows:

$$
\begin{aligned}
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&= (B_{SPEC} + B_{NR} + B_{FR} + B_{MEAS} + B_{DP})^2 \\
&\quad + \text{Var}_{SAMP} + \text{Var}_{MEAS} + \text{Var}_{DP}
\end{aligned} \tag{2.2}
$$

In practice, one method of assessing the MSE of an estimate is to estimate the eight components shown in (2.2). Similarly, one method of reducing the MSE is to develop a survey design that minimizes the contributions of each of the eight components in (2.2) to the total MSE. This approach to survey design is the basis of the fundamental principles of survey design discussed throughout this book. The illustration discussed in Section 2.6 shows how this approach can be used to identify the best survey design for a given survey budget among several alternative designs.

2.6 ILLUSTRATION OF THE CONCEPTS

To illustrate the usefulness of the mean squared error as a survey quality measure, consider a scenario in which a survey planner must choose among three alternative survey designs, labeled A, B, and C, for collecting data to meet the same set of research objectives. Design A specifies that the data will be collected by face-to-face interviewing, design B specifies telephone interviewing, and design C specifies data collection by mail, which is a self-administered mode. Since the costs associated with each mode of administration differ, the designs have been adjusted so that the total data collection costs are the same for each. For example, since face-to-face interviewing is more expensive than telephone or mail data collection, the sample size under design A must necessarily be smaller than that for the other two designs to meet the same total cost. Similarly, the per interview cost of design B is higher than the per interview cost of design C, and thus to cost the same, the sample size for design B must be smaller than that for design C.

In this illustration we assume that we have fairly complete information regarding the biases and variances associated with various sources of survey error for each design. In particular, we know roughly the biases associated with nonresponse, frame coverage, and measurement error. Since design A is expected to achieve the highest response rate, we estimate that the bias due to nonresponse will be the lowest for this design. Design C is expected to have the lowest response rate so its nonresponse bias will be highest.

Similarly, for design A, an area frame sampling approach will be used which implies that all housing units in a sample of areas will be listed and sampled. This intensive listing and sampling process will ensure complete coverage of the target population and therefore frame bias will be zero. By contrast, design B will use a random-digit-dial (RDD) telephone frame. For this frame, 10-digit telephone numbers are generated randomly so that all housing units that have a telephone have a chance of being selected. However, nontelephone housing units have no chance of being selected, and consequently, a small frame bias is expected as a result of the noncoverage of nontelephone housing units. Design C specifies the use of a telephone directory-type listing for obtaining the addresses of target population members, and thus both nontelephone and telephone households with unlisted numbers will be missed by this frame. Consequently, frame bias is expected to be the highest for this design.

The last source of bias for which information is available for all three designs is measurement error. As discussed previously, measurement bias arises from many sources, including the questionnaire, the respondent, the interviewer (designs A and B only), the setting, and the mode of administra-

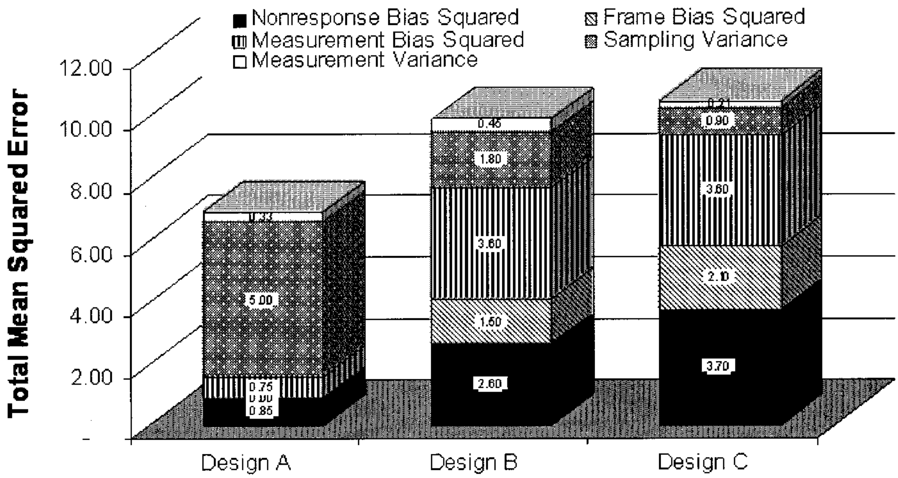


Figure 2.7 Comparison of the total mean squared error for three survey designs. Design A is preferred since it has the smallest total error, even though sampling error is largest for this design.

tion. This combination of systematic error sources is expected to be smaller for face-to-face interviewing for the primary objectives of the survey. Designs B and C are expected to cause larger measurement errors in these data; however, the net effect of measurement error bias will not differ appreciably for the two designs.

Next, with regard to measurement variance, interviewer error variance is expected to be particularly problematic for the primary contents of this survey. Interviewer variance is related to the influencing effects of interviewers on responses as a result of their expectations regarding respondent reactions to questions, their feedback to respondents, any inconsistent probing for "acceptable" responses, and many other biasing behaviors. The effect is expected to be much worse for face-to-face interviewing than for telephone interviewing and, of course, nonexistent for the mail survey. Thus, measurement variance is highest for design A and lowest for design C.

Finally, the sampling variance will be approximately proportional to the sample sizes for each design. Thus, design C, with the largest sample size, has a very small sampling variance; design A has the largest sampling variance, corresponding to its relatively small sample size.

Given this information, how does one proceed to choose the best or *optimal* design? Since the sample sizes for the designs are such that the total data collection costs for each are the same, and assuming that the time required to complete the data collection under each design is not an important criterion, the optimal design is the one that achieves the smallest mean squared error.

To compute the mean squared error, we simply sum up the squared bias and variance components as in Figure 2.7. As we see from this figure, the optimal design by this criterion is design A, the design with the smallest sample size. Note that had our criterion been to choose the design that minimizes the sampling error without regard to the nonsampling error components, design C would have been optimal. However, when nonsampling and sampling error components are considered jointly, this design ranks last.

Figure 2.7 leads to the following final lessons for this chapter:

- The mean squared error is the sum of the total bias squared plus the variance components for all the various sources of error in the survey design.
- Costs, timeliness, and other quality dimensions being equal for competing designs, the optimal design is the one achieving the smallest mean squared error.
- The contributions of nonsampling error components to total survey error can be many times larger than the sampling error contribution.
- Choosing a survey design on the basis of sampling error or variance alone may lead to a suboptimal design with respect to total data quality.