Sample Surveys

1. Learning Objectives

After reviewing this chapter readers should be able to:

- Describe the purpose of conducting a survey and its overall process
- Describe components of total survey error
- Describe considerations and options for designing a sample
- Describe steps in implementing the survey
- Describe considerations in calculating estimates based on survey data and how to report results

2. Introduction

Sample surveys are conducted by selecting units from a population and recording information or data on the units. The units comprise the population and can be individuals or households; businesses or other establishments; students, teachers, or schools; acres of land; financial accounts or items in inventory; or any other element that can be meaningfully thought of as defining a population to be studied. This chapter gives an outline of the steps that are necessary to conduct a scientific sample survey and discusses many of the practical issues involved.

> ***Scientific*** **means that the survey adheres to standards that enable one to make defensible, statistical statements about the population of interest.**

Practical issues arise because of the complexities of gathering information on many real populations. This chapter focuses mainly on surveys of human populations.

References to more specific information on topics covered in this overview of sample surveys are provided at the end of the chapter.

3. Defining Objectives

Surveys are used to collect information that will answer scientific questions. The first consideration in designing a survey is to define the objectives of the survey in the context of these questions.

The objectives should reflect the intent of the scientific questions from the perspective of the survey methodology to be used to collect data for addressing these questions.

Survey objectives should specify:

- The study population (also called a target population) and any subgroups (also called domains) that are of special interest in addressing the scientific questions
- The organization of the population (e.g., housing units are associated with a group of people, students are nested in schools), which often has implications for gathering information on units in the population
- Likely methods for contacting members of the population or gathering data on selected units
- The main variables that will be recorded on the units of the population to address the scientific questions
- Available resources such as limitations on staff funding and time for conducting the survey

Using this information, a design is developed to meet the scientific objectives of the survey, often via an iterative process that leads to modifications of the design and possibly the objectives. Think about a population you would like to study and describe a variable you would like to measure. What are the units of measurements? What would be a high value? What would be a low value? What would be a typical value? What kind question(s) would you ask to obtain data for this variable? Make a note of these, to consider later in this chapter.

Exercise 1: Telephone Survey

Suggested Study: You'd like to conduct a study of middle-aged adults who are unemployed by conducting a telephone survey. You are particularly interested in understanding how unemployment is related to depression, so you have several questions about mood, and interactions with spouses/partners and family members. The study must be conducted within a limited timeframe and requires a large sample size.

Identify which items listed below are positive aspects of conducting a telephone survey and which pose challenges.

Item: Data completeness — Challenge | Benefit

Item: Time to results — Challenge | Benefit

Item: Data accuracy — Challenge | Benefit

Item: Sensitivity of subject matter — Challenge | Benefit

Item: Ability to recruit sample households — Challenge | Benefit

4. Total Survey Error

Total survey error is a conceptual framework used to systematically consider types of survey error during the design process and in describing its quality when completed.

Sampling Error

The most familiar form of error is sampling error. Sampling error represents the uncertainty in estimates that occurs because we observe data on a sample of individuals in the population rather than on every individual. We minimize sampling error by designing samples to provide the most precise estimates given resources. Sampling error is often expressed as standard errors or margins of error (e.g., a 95% confidence interval) for estimates, although these measures can be affected by other sources of error such as measurement error. Other forms of survey error are called nonsampling error, which can lead to bias or reduced precision in estimates.

Coverage Error

A common problem in studies of human populations is coverage error. Coverage error is the bias that occurs when we are unable to select the sample from the whole study population. For example, a telephone survey of households selected from white pages listings will not include households with no phone, households with unlisted land line numbers, or cell phone-only households. It is difficult to quantify coverage error without special studies of the unsampled portion of the population.

Nonresponse Error

Nonresponse error is the bias that occurs when some of the sampled individuals cannot be contacted or decline to provide a response to the survey and they differ in important ways from the respondents. Response rates have traditionally been used to indicate the potential for bias. There is, however, increasing evidence that response rates do not necessarily provide an accurate indicator of bias. In particular, response rates by themselves do not indicate how respondents and nonrespondents are different from each other as groups. Other methods for evaluating the potential effect of nonresponse include comparing sample composition to known population demographics, conducting follow-up studies of nonrespondents, and studying how mean responses change in relation to the effort required to obtain the response.

4. Total Survey Error

Measurement Error

Measurement error is error in the responses for a sampled individual that is due, for example, to failure to accurately recall an event or misrepresentation of an answer due to social pressures or a concern for revealing sensitive behaviors. Measurement error may lead to bias or increased variance, and is difficult to quantify without auxiliary information or special studies that separate sampling error from measurement error or compare respondent information with a more accurate source.

Specification Error and Processing Error

Two other sources of error receive less attention, but are critical aspects of total survey error. Specification errors arise from misstating the objectives of the survey and concepts to be studied, as well as asking questions or selecting data elements that fail to address scientific objectives. Processing errors can also be generated by problems with procedures and computer systems that support operational aspects of the survey such as selecting the sample, collecting the data, and performing statistical calculations.

Survey design involves careful consideration of each type of survey error, and how best to minimize errors in addressing survey objectives with available resources.

It is important to identify where possible sources of error could be in various designs. Based on the information provided in this section, consider the population you would like to study and think about:

- **What kinds of survey error are involved in the contact methods you considered**
- **What kinds of survey error are involved in the variable you proposed**

5. Designing a Sample

Scientific surveys should be based on **probability sampling** from clearly specified study populations. Probability sampling provides the statistical framework for making credible inferences about the study population.

> **In a probability sample, all units in the population have a positive probability of being selected. That is, no part of the population is left out of the sampling process. In addition, the probability of selecting a sampled unit is known, which is necessary for making valid estimates from survey data.**

Although many people think of random sampling as haphazard, it involves well-defined rules for selecting individuals in the population. The random aspect of probability samples has to do with the units that are chosen to be included in the sample.

In contrast, other sampling methods involve personal judgment or a tendency to recruit readily available units. Selection biases associated with non-probabilty approaches lead to biased estimates, and there is no valid method for calculating standard errors. Designs that should be avoided include:

- Purposive samples, in which the investigator personally selects specific sample units without a chance mechanism;
- Convenience samples such as intercept surveys, in which persons walking by are asked to complete a survey;
- Quota samples, in which data collection within a group is halted once the number of responses for the group reaches a specific quota level; and
- Other non-probability sampling methods such as respondent-driven sampling that do not rely on a formal probability sampling mechanism.

5. Designing a Sample

The first step in selecting a probability sample is to **identify a sampling frame that lists or includes all units in the population** to the extent possible. For most behavioral surveys of human populations, it is necessary to sample a telephone number or address to obtain responses from a population of individuals. Telephone survey frames include listed telephone numbers, randomly digit dial (RDD) or randomly generated telephone numbers (which include unlisted numbers), and post office delivery sequence files matched to telephone numbers. Cell phone listings are beginning to emerge for national samples, but currently provide no geographic links for regional samples. For mail surveys, white page listings and post office delivery sequence files are commonly used. Quite often, sampling frames fail to contain a segment of the population, which can lead to coverage bias. For example, because persons without land lines (i.e., only have a cell phone or have no phone) tend to have lower incomes, a survey of economic well-being that uses white page telephone listings will probably underestimate the percent of the population who have economic difficulties.

In some cases, **two or more frames can be used to cover a population** (e.g., cell phone and RDD telephone numbers will cover persons with cells, but not those without any phone) or to obtain data to estimate coverage bias. Alternatively, an area sampling frame may be used to randomly select a sample of small geographic units for which a complete sampling frame is developed (e.g., sampling tracts from a list of all tracts in a region, and then listing all housing units in the sampled tracts).

5. Designing a Sample

Probability samples are designed to achieve goals such as:

- **Obtaining a representative sample;**
- **Ensuring precision levels for specific subpopulations; and/or**
- **Reducing data collection costs.**

A variety of sampling approaches may be combined to achieve one or more of these goals simultaneously. While most researchers are familiar with simple random sampling, in which units are selected completely at random, this is rarely the best approach to sampling populations. For example, to improve the representation of the population, a stratified sample design is often used, which involves dividing the sampling frame into distinct subgroups called strata (e.g., gender and age categories, states or regions) and selecting an independent sample from each group. To improve precision, strata should reflect differences in the population with respect to the study topic or important analysis domains. Systematic sampling with a sampling frame sorted by a characteristic correlated with the study topic can also be effective in improving representation and precision. Systematic sampling involves selecting a random start and then using a fixed interval to select sample units from the frame.

Study objectives often require **oversampling** of certain groups in a population, such as specific ethnic or racial groups. This can be achieved by:

- Increasing sampling rates in strata that contain a high proportion of target populations. For example, we might stratify by region and select proportionately large samples in geographic regions with high concentrations of ethnic minorities.
- Sampling units with probability proportional to a size, or PPS. In PPS, an importance (or size) measure that is correlated with the target group is used to increase the likelihood of including a targeted unit in the sample. For example, to increase the number of low income households in our sample relative to the general population, we might select an area sample of counties using a size measure that is proportional to the county poverty rate.

5. Designing a Sample

Cluster sampling is used to reduce data collection costs or to sample populations for which a complete sampling frame is unavailable. Clusters are groups of related units (e.g., through geographic proximity or family relations). Instead of sampling population units directly, clusters are selected first, after which a sample of units may be selected from each sampled cluster. If necessary, a sampling frame is developed for the sampled cluster as part of this process. For example, in most behavioral surveys, a landline telephone number or postal address represents a cluster of persons who live in a household. To sample individuals in a household, a list of eligible residents is made, and one is randomly selected. Travel costs for in-person surveys can be greatly reduced if samples are geographically clustered rather than evenly distributed throughout a region. In an area sample, clusters such as counties or tracts are selected, and then households are selected from each cluster. Because units in clusters tend to be positively correlated, the reduction in costs for cluster samples are often associated with a reduction in precision relative to designs that spread the sample across the population but are more expensive to implement.

Complex probability samples used to conduct surveys are simply samples drawn from a design that combines several of these sampling methods to address research and operational objectives. For example, a common complex sample design in national surveys is to stratify the U.S. into geographic regions, then select clusters such as counties or metropolitan areas within regions using a selection probability proportional to population size (or a demographic measure), then randomly select households within these clusters (maybe using more stratification or clustering), and then select an adult within a household (another cluster) using simple random sampling.

Often, stratified, clustered or systematic samples are designed so that each individual in the population has the same chance of being selected, which may simplify some analyses. However, if some groups in the population are deliberately oversampled, then survey weights are needed to correct for variation in sampling rates when calculating population estimates (see Section 8).

5. Designing a Sample

Many other forms of sampling exist to address special situations such as studying rare or less common populations. For example, adaptive and network sample designs involve selecting a probability sample of population units (e.g., persons) and collecting data about the presence of a relatively uncommon characteristic (e.g., whether the respondent has a rare disease, whether the respondent engages in drug use). If the uncommon characteristic is detected, then a list of additional units who might also have this trait is elicited from the respondent, and the same data are collected on all or a sample of those units.

This is another example of oversampling part of the population, and the procedures should be designed to provide information to calculate the selection probability and survey weight for each unit. Methods such as respondent-driven sampling often do not provide this kind of information and should be avoided.

Exercise 2: Appropriateness of Stratification

Read the statement below and select the appropriate response to the question asked. Once you answer this question, you will be asked a series of related questions.

Statement:
Suppose you need to conduct an inventory survey for a large company that has its products stored in five warehouses. The company generally puts products of similar sizes into warehouses together. In this case, would stratification be appropriate?

+ Yes — No

Question 1 of 3: Which sampling goals can stratification address?

a) Writing a question about household income	+ Yes	— No
b) Obtaining a minimum sample size for a subpopulation	+ Yes	— No
c) Changing sample sizes after the survey has started	+ Yes	— No
d) Increasing the spread of the sample across the range of characteristics in the population	+ Yes	— No
e) Reducing travel costs	+ Yes	— No
f) Reducing the margin of error for estimates	+ Yes	— No

Question 2 of 3: Which sampling design tends to be associated with higher standard errors relative to simple random sampling?

a) Systematic sampling	+ Yes	— No
b) Cluster sampling	+ Yes	— No
c) Stratified random sampling	+ Yes	— No
d) Sampling with probability proportion to size	+ Yes	— No

Question 3 of 3: What sampling designs tend to be associated with lower standard errors relative to simple random sampling?

a) Systematic sampling	+ Yes	— No
b) Cluster sampling	+ Yes	— No
c) Stratified random sampling	+ Yes	— No
d) Sampling with probability proportion to size	+ Yes	— No

6. Developing a Survey Instrument

A survey instrument is a tool for consistently implementing a scientific protocol for obtaining data from respondents. For most social and behavioral surveys, the instrument involves a questionnaire that provides a script for presenting a standard set of questions and response options. The survey instrument includes questions that address specific study objectives and may also be used to collect demographic information for calculating survey weights. In some surveys, questionnaire responses are augmented by other kinds of measurements derived from instruments, such as lab samples or physical assessments.

A difficult task in creating a questionnaire is translating a researcher's questions into items that are sufficiently succinct and simple for the respondent to understand and provide accurate answers.

In general, survey questions should:

- **Contain only one idea or question**
- **Define the scope to consider, such as the time period or activities that are relevant to the question**
- **Be written with neutral language to avoid leading the respondent to a specific answer**
- **Use language that enables less educated persons to easily understand the question.**
- **Contain response options that are simple, clear, consistent, and include the full range of responses that might occur**
- **For categorical responses, be mutually exclusive and exhaustive so that a respondent can pick one and only one option**
- **For numeric responses, guide the respondent to provide the response in a consistent format and units**

Methodologies used to formulate good questions rely on cognitive psychology principles that express how humans process information. For example, because cognitive errors can be made recalling past events (recall error is a form of measurement error), dietary surveys ask each respondent to describe dietary intake for a recent, short time period such as the prior day. For some health and social science topics, standardized question sets are available that provide consistency across studies for specific measures. For example, a standard question set has been developed for evaluating a household's risk of food insecurity or hunger.

6. Developing a Survey Instrument

The mode of data collection describes the way in which the respondent's data are obtained. Modes may involve interviewer-mediated data collection, such as interviews conducted over the telephone or in-person, or may involve the respondent completing a self-administered instruments sent via the mail or offered on the web.

The choice of modes depends on several factors, including:

- **Available funds and staff for conducting the survey;**
- **Desired data quality;**
- **Available frame information for contacting respondents; and**
- **The topic of the survey.**

Telephone or in-person interviews provide the best data quality for most surveys, although for sensitive topics such as substance use or sexual behaviors, self-administered methods encourage more accurate reporting. When budgets are small, mail or web surveys are often preferred. If a reliable source of telephone numbers or email addresses for the population is not available, a mail survey is usually conducted.

6. Developing a Survey Instrument

Multiple-mode surveys are increasingly common as a means to improve the chances of contacting and recruiting respondents, reduce study costs, or collect different kinds of information. For example, a longitudinal survey in which a household is contacted several times over a time period may begin as an in-person survey for the first contact, but once rapport and familiarity with the survey is established, subsequent contacts are made via phone to save costs. Alternatively, to reduce field costs, a survey may involve an initial telephone interview and then an in-person follow-up phase to obtain physical measurements on an individual who has completed the telephone interview. A telephone survey may also have a mail follow-up phase to obtain more sensitive data in a self-administered mode or as an alternative contact strategy for respondents that are hard to reach during normal calling hours. Figure 1 illustrates 3 possible options for initial contact and follow-up.

The mode of data collection affects the way in which the survey instrument is designed.

Complicated skip patterns are difficult to implement in a mail survey. Similarly, if a respondent listens to a pre-recorded voice asking questions and then enters answers by pressing buttons on a phone, then the questions and choices probably need to be shorter and simpler than with a live telephone interviewer.

Think of a population you would like to study and a question you would like to ask. What data collection mode do you think would be best for your survey?

Exercise 3: Modes of Administration

Question: Suppose you are conducting a survey about trash disposal behaviors and are particularly interested in specific recycling or littering behavior. Which mode, or modes, of administration would be most appropriate?

Mail

Interviewer Administered

Phone

Web

Question: Suppose you are conducting a study which is interested in capturing how much people exercise. Which question below is likely to receive the most accurate answer?

On how many days did you exercise at least 20 minutes last week?

On how many days did you exercise at least 20 minutes last year?

Do you exercise a lot?

7. Collecting Data

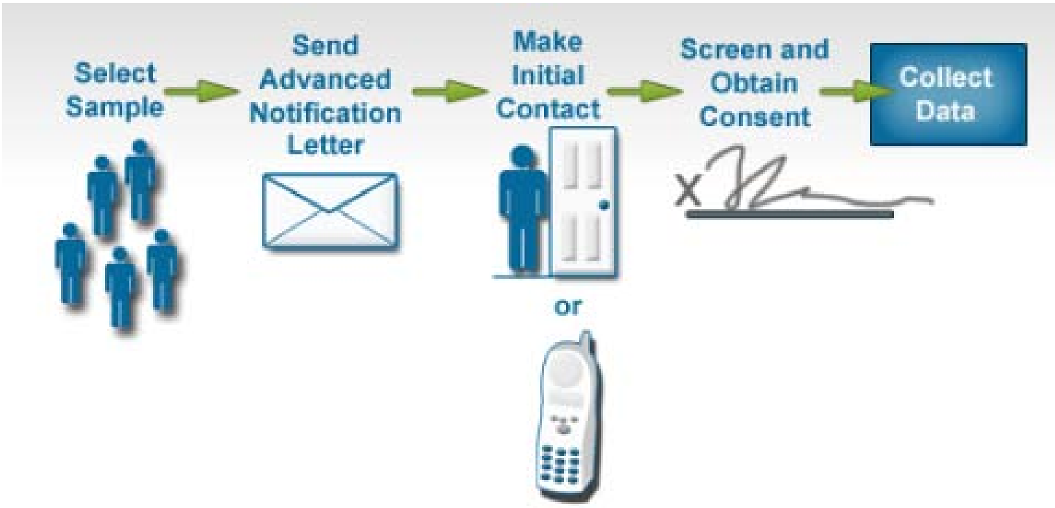
When designing a scientific survey, methods for each step of the data collection process must be carefully specified. The **data collection process**, which involves contacting and recruiting a respondent and then collecting the data, differs somewhat for interviewer- and self-administered surveys.

For **interview-based surveys**

- The first step often begins with advanced notification via a letter describing the purpose of the survey and respondent rights.
- The next step is to contact the **sampled** unit (e.g., household) to explain the survey and respondent rights again and request participation in the survey. The contact process generally involves repeated contact attempts (often 12 or more) on different days and at different times of day, as well as ad hoc searches for updated telephone numbers or addresses.
- Depending on the type of survey, verbal or written **consent** may be needed from adult respondents or by parents on behalf of child respondents. If the respondent refuses initially, then a later attempt to recruit the respondent may be made.
- If the respondent agrees to participate, there may be **screening process** to determine eligibility and possibly a selection process such as randomly sampling an adult from the household.

Once these steps are completed, the interview takes place using the questionnaire.

Figure 1: Data Collection Process



7. Collecting Data

For **mail or web surveys**, the data collection process is often simpler. Often no pre-notification is given. Instead, the initial contact via mail or email explains the survey and respondent rights and provides either the paper materials for completing a mail survey or information for accessing the web survey. When mailings are returned by the post office or computer system without reaching the sampled person, attempts are made to identify the correct postal or email addresses. Also, if there is no response to the first contact, a standard procedure is to re-contact the sampled person one or more times. In mail surveys, this usually involves a postcard reminder 10 days after the initial mailing, and then a new copy of the questionnaire if no response is received from the postcard.

With multi-mode surveys, these approaches are combined to create a data collection protocol that maximizes response rates to the degree possible given available resources.

In addition to repeated contacts and refusal conversions, incentives may be used to improve response rates. Many incentives are cash, but they may also take the form of gift cards, novelty items, or entry into a drawing for a larger prize. These may be given prior to conducting the survey or after the survey is completed. Pre-survey incentives are generally viewed as more effective, but it may not be possible to use this approach if a postal address is not available or if only a small fraction of sampled respondents will be deemed as eligible to participate based on the screening process.

At the end of the survey, each case is labeled with a disposition describing the final state of the contact process. Dispositions for telephone surveys include:

- Non-working or non-residential number;
- Max calls (full set of attempts made without reaching a person);
- Unable to complete survey (due to, for example, language or health issues);
- Refusal to participate;
- Unable to reach specific respondent;
- Screened out of survey (respondent is not eligible given the requirements of the survey); and
- Response (partial or full).

8. Preparing for Analysis

Once data are collected, it is important to edit or ‘clean’ the data.

One should begin by checking that the **values recorded for each variable are within the allowable range or set of possible values**. If data are typed into a computer by the respondent, an interviewer, or someone else working on the survey, then there is a possibility of data entry errors. Ages less than 18 years for a survey of adults, weights and heights of unrealistic sizes, and category indicators that do not correspond to a category used in the survey need to be identified and, if possible, corrected. If it is not feasible to correct errors, then one should delete the errant response so that it is not analyzed with the legitimate data. Ideally a record will be kept of the edits that are made to the data set.

Another step in the editing process is to **address the issue of missing values in the data set**. Some variables will not have values recorded for some respondents because the variables do not pertain to them. For example, it does not make sense to ask nonsmokers about smoking habits or individuals of one gender about health issues that affect only members of the other gender. In the survey instrument, questions that are not relevant to a respondent should be skipped and a “not applicable” code indicating that it was legitimate to skip that question for the respondent should be recorded for such questions. Other values are missing due to accidental skips or a refusal of the respondent to answer. Codes indicating that the response was refused or is unintentionally missing need to be recorded.

At this point one may wish to impute or fill in plausible values for missing values. Imputation should be handled with care. Simply filling in average values from the responding units will create distortions in the distribution of values and has the potential to dramatically affect statistical analyses. Imputations can be created through statistical modeling and estimation, such as from a linear regression model that predicts missing values based on an estimated linear regression equation. Alternatively, imputations can be created by randomly selecting donors with complete data to provide data values from respondents to donate to the cases with missing values. Methods that use auxillary variables to match potential donors to cases with missing values can be used to select good donors. Donors that match the cases with missing values closely are likely to create reasonable imputations. It is always good practice to keep an indicator variable that indicates which values are imputed in the data set so that this information is available when assessing conclusions based on statistical analyses.

A step that is necessary when preparing data for analysis is to create a **clearly documented list of variables and their meanings**. For example, if sex is recorded with values of 0 and 1 (or 1 and 2), one needs to know which value means female and which means male. One also needs to know how legitimate skips, missing values, do-not-know responses or refusals, and imputations are indicated. Other information that could be of use to the data analyst should also be recorded. This might include how the respondent was contacted, when the interview took place, who conducted the interview, who the respondent was (e.g., was it the target respondent or a proxy report by someone else).

9. Calculating Survey Weights

Once the data are collected and prepared for analysis, one must consider whether it is necessary to compute survey weights.

Survey weights are numbers associated with the respondents that specify the weight or influence the various observations should have in analysis. Weights are used to adjust for unequal probabilities of selection in the survey sampling design, unequal response rates between subsets of the intended sample, differences between the population covered by the survey frame or list and the target population, and differences between sample estimates and desired numerical controls. The final survey weight can usually be thought of as a measure of the number of population units represented by the response.

9. Calculating Survey Weights

Survey weights are often computed in a series of steps.

1. For each unit in the sample, **the initial weight is typically computed as the inverse of the probability of selection**. Units with high probabilities of selection get low weights, whereas units with low probabilities of selection get high weights. Sampling rates can be unequal across strata to increase efficiency of estimation (i.e., it is a good idea to have higher sampling rates when the distribution of values in a stratum is more variable, when it is less expensive to sample in the stratum, and when the stratum is larger) and to ensure an adequate sample size for a subpopulation of interest.

2. **If response rates vary among groups in the sample, then the weights in the groups are usually adjusted to compensate**. For example, if a survey of students is being conducted and the year of study (first year student, sophomore, junior, or senior) is known for all sampled students, but seniors respond half as much as the others, then one would double the sampling weights on all seniors.

3. Subsequent steps are used to **further adjust sampling weights so that estimates of some quantities match population totals or sizes and desired numerical controls**. Suppose a survey is designed based on information about student enrollment in schools for the previous academic year. A sample is taken and the first two weighting steps are implemented. Now suppose the enrollment numbers for the current academic year are available for all school districts in the state, and enrollment has shifted relative to the prior year in response to changing populations in some districts. One could adjust weights by increasing them proportional to relative population changes in areas that gained student numbers, and decreasing them proportionally in other areas where student counts declined. Essentially, we use the new population counts as 'control totals' so that the weights reflect what we know about the population. Control totals are numbers either at the population level or estimated based on samples (possibly in the past) that one wants to be able to reproduce or match in estimation using the current sample. If there is only one factor, a ratio adjustment can be made. If adjustments are desired for two or more control factors, a procedure called 'raking' is often used to iteratively adjust the weight for the totals from each control factor.

4. Finally, one should **examine the distribution of sampling weights produced by these steps**. If the distribution of weights is highly variable, then estimates of population parameters may have large standard errors. Highly variable weights for parts of the sample corresponding to an important subpopulation (e.g., a racial/ethnic group, an at-risk group) can be of particular concern. One might consider trimming (i.e., restricting the range of) survey weights to reduce the weight maximum or shorten the range of weights produced by the computations directly. Modifications of survey weights to reduce their variability aim to reduce variability of estimates of population characteristics, but may cause statistical bias that was addressed by previous weighting steps.

9. Calculating Survey Weights

There are other approaches to weighting that go beyond these basic steps. For example, auxiliary information that is correlated with response variables can be incorporated via regression weighting, with the goal of increasing the precision of estimates.

Regardless of the process, survey weighting seeks to compensate for several factors, including:

- Sampling rates;
- Response rates;
- Population sizes and other control characteristics known from other sources; and
- Precision of estimates.

Exercise 4: Survey Weights

Question 1 of 2:
If you use a simple random sample design, then you do not need to use survey weights.

+ True | – False

Question 2 of 2:
What are some reasons survey weights are used? Identify which reasons for using survey weights below are true and which are false.

True | False

Reasons	
Weights are unimportant except in large-scale government surveys.	
To adjust for nonresponse, especially if it differs across subpopulations.	
To adjust for unequal selection probabilities so that estimates are unbiased.	
To adjust for the differences in the respondents.	
To incorporate known information about the population expressed as control totals so that estimates will be more accurate.	

10. Calculating Estimates

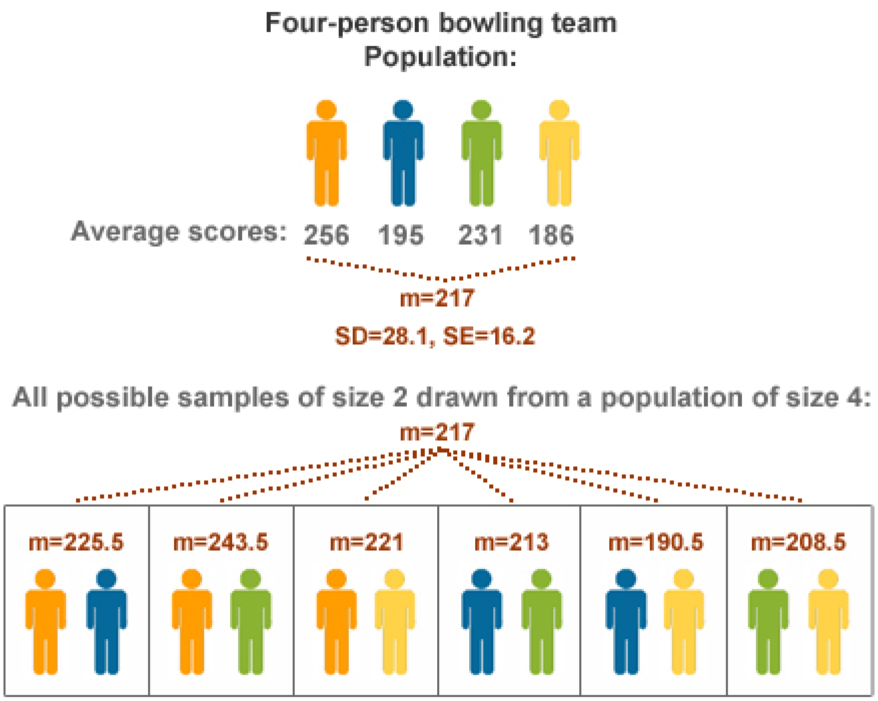
Once the data are collected and survey sampling weights are computed, one uses the data and weights to estimate characteristics of the population. If the survey data are collected based on a simple random sample and there are no complications, then reporting the average of the sample values produces a statistically unbiased estimate of the population average.

Standard deviation (SD) is used to measure the variability (or dispersion or spread) of observations while the standard error (SE) is the standard deviation of a statistic.

When using sample means as a method of estimating the population mean, the standard error of the mean is the standard deviation of the sample means calculated over all possible samples of a given size drawn from the population. It is frequently estimated by taking the standard deviation of a sample and dividing by the square root of the number of observations in the sample.

The unbiasedness of the sample mean in this case is a property of the sampling design and the estimator (procedure used to turn the data values into an estimate). It means that when estimates from every possible sample selected using the specified sample design are averaged, the average is equal to the population parameter that is estimated by the estimator. This is illustrated in Figure 2.

Figure 2: Calculating Sample Mean



10. Calculating Estimates

When a design other than simple random sampling is used, or if it is of interest to estimate population parameters other than the mean (e.g., the total number of population units with a specific characteristic), the principle of design-based estimation is commonly applied. For example, suppose that a survey design calls for unequal probabilities of selection for population units and survey weights for the sample are computed as inverse probabilities of selection. If there are no complications such as nonresponse, then the survey weight is the inverse of the selection probability, and an unbiased estimator of the population mean is the weighted average of the observations. Under the survey design, the average of estimated means from all possible samples equals the population mean, Design unbiased estimates, or approximately design unbiased estimates, are available for population parameters other than the population mean as well.

If complications occur, then it is more difficult to believe that the estimator is unbiased or approximately unbiased. Bias is a particular concern with:

- Missing data (nonresponse error);
- Systematic differences between the frame used to select the sample and the target population (coverage error); and
- Variables that are measured with error (measurement error).

Weighting adjustments and imputation measurement error models can be used to minimize the error associated with these factors.

10. Calculating Estimates

Along with the estimate of a population characteristic one should also report a **standard error or margin of error**. When sampling from a finite population, for any particular sample design, one can generate different samples. Each sample can produce a different estimate.

The standard error is the standard deviation of estimates across the various samples; that is, it measures the variability of the sample estimates under the design used to select the sample.

A small standard error is desirable. In the case of design-based estimation, there often is a formula to compute an estimate of the standard error, or approximate standard error, of the estimator. A margin of error is another way to express the uncertainty in the estimate.

The margin of error equals the standard error times a reference value that is related to the confidence that the interval for the margin of error actually includes the unknown population parameter value.

For a 95% confidence level, this reference value is usually about 2. Margins of error are used to form confidence intervals for population characteristics, i.e., it is the numerical interval that includes the value of the characteristic in the population with a given level of confidence. For example, suppose a poll with a sample size of 1000 produces an estimate that 50% of the voting population supports a specific candidate, with a margin of error of 3%. This margin of error is based on a 95% confidence level, and the confidence interval for this estimate is 50% ± 3% or (47%, 53%).

10. Calculating Estimates

In many circumstances, one does not rely on design-based estimators. Instead, one considers estimators based on statistical models. Sometimes the statistical models suggest functions of the data that can be evaluated within the random sampling design-based context. That means that models help determine how one wants to use the data, but the evaluation of the resulting estimators returns to a focus on the design and its properties. The most common situation is the use of regression models, in which an outcome is predicted via a linear equation from one or more covariables. The use of statistical models in survey sampling estimation is an area of active research. Generally it is acknowledged that models have a role to play when there are missing data and when one wants to use auxiliary information to improve estimation. Careful consideration must be used when deciding exactly how to use survey weights and the design factors in conjunction with statistical models. In general, one should include variables related to the design, such as cluster or strata indicators, and to the computation of survey weights in the models. For model based estimators, one should also report measures of uncertainty, or mean square errors, which are analogous to standard errors.

11. Reporting Results

Reports based on survey data should give the reader enough information to appreciate most of the important elements, as described below.

Reports should clearly describe the scientific methods that were applied and any potential biases associated with these methods, such as:

- A clear definition of the target population and the variables being measured;
- A clear statement of how the sample was selected, how respondents were contacted, and parts of the population that are not included in the sampling frame or the responding units;
- Disposition rates such as contact, cooperation, and response rates for the sample and possibly for important subpopulations;
- Response rates to important questions, such as questions related to income or negative behaviors such as drug use, particularly if missingness rates for the question is higher than unit nonresponse from the survey; and
- Any other substantial problems with the survey.

In terms of numerical estimates, one should report estimates along with standard errors (or margins of errors, confidence intervals, or mean square errors) and sample sizes. If one reports hypothesis tests, then ideally estimates, standard errors, and p-values are also reported. Simply reporting a p-value does not effectively communicate much useful information. Professional statistical standards should be followed when forming tables and figures. And, of course, it is critical to interpret the resulting estimates in the context of the scientific problem being studied.

12. Resources

Further Reading

Biemer, P. P. and L. E. Lyberg. (2003). *Introduction to Survey Quality*. John Wiley & Sons, Inc., New York. 424 pp.

Cochran, W. G. (1977). *Sampling Techniques, Third edition*. John Wiley & Sons, Inc., New York. 428pp.

Dillman, D. A. (1978). *Mail and Telephone Surveys: The Total Design Method*. John Wiley & Sons, Inc., New York. 375 pp.

Dillman, D. A., Smyth, J. D., Christian, L. M. (2008). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons, Inc., New York. 500 pp.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., Tourangeau, R. (2004). *Survey Methodology*. John Wiley & Sons, Inc., New York. 448 pp.

Kish, L. (1965). *Survey Sampling*. John Wiley & Sons, Inc., New York. 643pp.

Lohr, S.L. (1999). *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA. 494pp.

Salant, P., and D. A. Dillman. (1994). *How to Conduct Your Own Survey*. John Wiley & Sons, Inc., New York, NY. 232 pp.

Tourangeau, R., Rips, L. J., Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press: New York. 401 pp.

Selected Web Resources

Proceedings of the Survey Research Methods Section, American Statistical Association. From the Joint Statistical Meetings and selected papers from the Annual Meeting of the American Association for Public Opinion Research. Available at: http://www.amstat.org/sections/srms/Proceedings [last accessed June 13, 2008]

Standards & Ethics resources (including methods for outcome rates), American Association for Public Opinion Research. Available at: http://www.aapor.org/standardsethics [last accessed June 13, 2008]

What is a Survey? *Survey Research Methods Section, American Statistical Association*. Available at: http://www.whatisasurvey.info [last accessed June 13, 2008]

13. References

There are no references for this chapter; however, see Section 12 for a list of further reading.

14. Author Biography

Sarah Nusser, PhD is interested in survey design and estimation, computer-assisted survey methods, and statistical applications in the biological sciences. She is part of a team of ISU statisticians and survey professionals that collaborate with the USDA Natural Resources Conservation Service to develop and implement statistical and survey methods for the National Resources Inventory (NRI). The NRI is a large agro-environmental survey designed to monitor the nation's natural resources on nonfederal lands. She currently has research projects in using geospatial data in computer-assisted survey data collection, measurement error models for physical activity data, and evaluating the effect of household financial stress on childhood obesity.

Mike Larsen, PhD received his PhD in Statistics from Harvard University. He was a senior lecturer at The University of Chicago and an assistant and then associate professor at Iowa State University. Currently he is an associate professor at The George Washington University with appointments in Statistics and the Biostatistics Center. His sample survey experience includes ISU's Center for Survey Statistics and Methodology, the U.S. Census Bureau, the Gallup Organization, Inc., and NORC. His research interests include sample surveys, missing data, record linkage and administrative records, confidentiality, small area estimation, statistical modeling, and hierarchical modeling.