Sampling Methods and Practice

From a talk by Richard L. Scheaffer
University of Florida

The topic of Sampling Methods and Practice fits well with that of Categorical Data Analysis. Indeed, most survey questionnaires produce categorical data by asking for Yes/No or Agree/Disagree responses. Typically, the reports on the surveys present proportions and percentages of the responses. In this section, we will consider the topic of Survey Sampling, its important features and appropriate techniques of analysis.

Sample Surveys and Experiments

A sample survey differs from an experiment in several important ways. A sample survey is characterized by

- a clearly specified population
- a sample selected by a random process from that population
- the goal of estimating some population parameters

An experiment is characterized by

- a treatment or treatments of interest
- some form of control, either a control group or another treatment
- randomized assignment of the experimental unit (subject) to a treatment
- the goal of establishing treatment differences, if they exist.

The goals of a sample survey and an experiment are very different. The role of randomization also differs. In both cases, without randomization there can be no inference. Without randomization, the researcher can only describe the observations and cannot generalize the results. In the sample survey, randomization is used to reduce bias and to allow the results of the sample to be generalized to the population from which the sample was drawn. In an experiment, randomization is used to balance the effects of confounding variables.

Some Terminology

Element: An element is an object on which a measurement is made. This could be a voter in a precinct, a product as it comes off the assembly line, or a plant in a field that has either bloomed or not.

Population: A population is a collection of elements about which we wish to make an inference. The population must be clearly defined before the sample is taken.

Sampling Units: Sampling units are nonoverlapping collections of elements from the population that cover the entire population. The sampling units partition the population of interest. The sampling units could be households or individual voters.

Frame: A frame is a list of sampling units.

Sample: A sample is a collection of sampling units drawn from a frame or frames. Data are obtained from the sample and are used to describe characteristics of the population.

Example 1 Suppose we are interested in what students in a particular high school think about the drilling for oil in our national wildlife preserves. The elements are the high school students and the population is the students who attend this high school. The sampling units could be the students as individuals with the frame the alphabetical listing of all students enrolled in the school. The sampling units could be homerooms, since each student has one and only one homeroom, and the frame the class list for homerooms.

Example 2 Suppose we are interested in what voters in a particular precinct think about the drilling for oil in our national wildlife preserves. The elements are the registered voters in the precinct. The population is the collection of registered voters. The sampling units will likely be households in which there may be several registered voters. The frame is a list of households in the precinct.

When the population is the residents of a city, the frame will commonly be the city phone book. However, not everyone in the city has their phone listed in the phone book. In this situation, the frame does not match the population. A survey conducted from the frame of the phone book would likely suffer from undercoverage bias.

Probability Samples

Sample designs that utilize planned randomness are called *probability samples*. The most fundamental probability sample is the *simple random sample*. In a simple random sample, a sample of n sampling units is selected in such a way that each sample of size n has the same chance of being selected. In practice, other more sophisticated probability sampling methods are commonly used, but most of the statistical theory for the introductory course in statistics is based on the simple random sample.

First, we will define a stratified random sample, a systematic sample, and a cluster sample.

Stratified Random Sample: A stratified random sample is one obtained be separating the population elements into non-overlapping groups, called strata, and then selecting a simple random sample from each stratum. (Scheaffer, Mendenhall, and Ott, *Elementary Survey Sampling, 5th edition*, page 125).

Systematic Sample: A systematic sample is obtained by randomly selecting at random one element from the first k elements in the frame and every k^{th} element thereafter. This is known as a 1-in-k systematic sample. (Scheaffer, Mendenhall, and Ott, *Elementary Survey Sampling, 5th edition*, page 252).

Cluster Sample: A cluster sample is a probability sample in which each sampling unit is a collection, or cluster, of elements. (Scheaffer, Mendenhall, and Ott, *Elementary Survey Sampling, 5th edition*, page 289).

Dick Scheaffer, in *Elementary Survey Sampling* (p. 407-408) gives and excellent overview and comparison of the different standard methods of conducting probability samples. We include this discussion with only slight modification.

COMPARISONS AMONG THE DESIGNS AND METHODS

Simple random sampling is the basic building block and point of reference for all other designs discussed in this text. However, few large-scale surveys use only simple random sampling, because other designs often provide greater accuracy or efficiency or both.

Stratified random sampling produces estimators with smaller variance than those from simple random sampling, for the same sample size, when the measurements under study are homogeneous within strata but the stratum means vary among themselves. The ideal situation for stratified random sampling is to have all measurements within any one stratum equal but have differences occurring as we move from stratum to stratum.

Systematic sampling is used most often simply as a convenience. It is relatively easy to carry out. But this form of sampling may actually be better than simple random sampling, in terms of bounds on the error of estimation, if the correlation between pairs of elements within the same systematic sample is negative. This situation will occur, for example, in periodic data if the systematic sample hits both the high points and the low points of the periodicities. If, in contrast, the systematic sample hits only the high points, the results are very poor. Populations that have a linear trend in the data or that have a periodic structure that is not completely understood may be better sampled by using a stratified design. Economic time series, for example, can be stratified by quarter or month, with a random sample selected from each stratum. The stratified and the systematic sample both force the sampling to be carried out along the whole set of data, but the stratified design offers more random selection and often produces a smaller bound on the error of estimation.

Cluster sampling is generally employed because of cost effectiveness or because no adequate frame for elements is available. However, cluster sampling may be better than either simple or stratified random sampling if the measurements within clusters are heterogeneous and the cluster means are nearly equal. The ideal situation for cluster sampling is, then, to have each cluster contain measurements as different as possible but to have the cluster means equal. This condition is in contrast to that for stratified random sampling in which strata are to be homogeneous but stratum means are to differ.

Another way to contrast the last three designs is as follows. Suppose a population consists of $N = nk$ elements, which can be thought of as k systematic

samples each of size *n.* The *nk* elements can be thought of as *n* clusters of size k, and the systematic sample merely selects one such cluster. In this case the clusters should be heterogeneous for optimal systematic sampling. By contrast, the *nk* elements can also be thought of as *n* strata of k elements each, and the systematic sample selects one element from each stratum. In this case the strata should be as homogeneous as possible, but the stratum means should differ as much as possible. This design is consistent with the cluster formulation of the problem and once again produces an optimal situation for systematic sampling. So we see that the three sampling designs are different, and yet they are consistent with one another with regard to basic principles.

The Need for Probability Samples

Consider the table shown below of the accuracy in the final Gallup Presidential Polls from 1936 to 1984.

Gallup Poll Accuracy

Year	Gallup Final Survey	Election Result	% Error
1936	55.7% Roosevelt	62.5% Roosevelt	6.8%
1940	52.0% Roosevelt	55.0% Roosevelt	3.0%
1944	51.5% Roosevelt	52.3% Roosevelt	0.8%
1948	44.5% Truman	49.9% Truman	5.4%
1952	51.0% Eisenhower	55.4% Eisenhower	4.4%
1956	59.5% Eisenhower	57.8% Eisenhower	1.7%
1960	51.0% Kennedy	50.1% Kennedy	0.9%
1964	64.0% Johnson	61.3% Johnson	2.7%
1968	43.0% Nixon	43.5% Nixon	0.5%
1972	62.0% Nixon	61.8% Nixon	0.2%
1976	48.0% Carter	50.0% Carter	2.0%
1980	47.0% Reagan	50.8% Reagan	3.8%
1984	59.0% Reagan	59.2% Reagan	0.2%

Source: G. Gallup, Jr. *The Gallup Poll, Public Opinion 1984*. Copyright © 1985, Scholarly Resources Inc., Wilmington, DE.
From Scheaffer, Mendenhall, Ott, *Elementary Survey Sampling, 5th Edition*, Duxbury Press.

Prior to 1948, the Gallup Poll used a quota sampling technique, which is not a probability sample. They had sought to find a representative group that matched the demographics of the country. Although the resulting sample did accurately represent the demographics of the country, it incorrectly predicted that Dewey would beat Truman in the election. Quota sampling failed. The samples taken after 1948 were probability samples. Even though the number of people in the sample was smaller than for polls used prior to 1948, the errors are generally much smaller.

Sources of Errors in Surveys

Statistician Robert Gross of the University of Michigan has categorized the kinds of errors in surveys into errors of non-observation and errors of observation.

Errors of non-observation include sampling error, error in coverage, and errors due to non-response.

- Sampling error is the "natural" error that is a part of any sampling process. If the sampling process were repeated a number of times, the results would differ each time, producing a variation in the estimates of the population parameters.

- Coverage error results when the frame does not match the population. For example, if the frame is the town phone book, then people with unlisted numbers and those without phones will be missing from the frame.

- Non-response error is a result of elements in the frame that have died, moved away, refuse to participate, or otherwise are missing from the sample.

Errors of observation include interviewer error, respondent error, measurement error, and errors in data collection.

- Interviewer error is a result of the interaction between the interviewer and the subject being interviewed. Most people who agree to an interview do not want to appear disagreeable and will tend to side with the view apparently favored by the interviewer, especially on questions for which the respondent does not have a strong opinion. Reading a question with inappropriate emphasis or intonation can force a response in one direction or another. Interviewers of the same gender, racial, and ethnic groups as those being interviewed are, in general, slightly more successful.

- Respondent error is a result of the differing abilities of the respondents in a sample to answer correctly the questions asked. Most respondent errors are unintentional and are due to either recall bias (the respondent does not remember correctly) or prestige bias (the respondent exaggerates). At times, respondent error may be due to intentional deception (the respondent will not admit breaking a law or has a particular gripe against an agency).

- Measurement error occurs when inaccurate responses are caused by errors of definition in survey questions. For example, what does the term *unemployed* mean? Should the unemployed include those who have given up looking for work, teenagers who cannot find summer jobs, and those who lost part-time jobs? Does *education* include only formal schooling or technical training, on-the-job classes and summer institutes as well? Items to be measured must be precisely defined and be unambiguously measurable.

- Errors in data collection occur in all surveys. The most commonly used methods of data collection in sample surveys are personal interviews and telephone interviews. These methods, with appropriately trained interviewers and carefully planned callbacks, commonly achieve response rates of 60% to 75%. The procedure usually requires the interviewer to ask prepared questions and to record the respondent's answers.

The primary advantage of these interviews is that people will usually respond when confronted in person. However, if the interviewers are not thoroughly trained, they may deviate from the required protocol, thus introducing a bias into the sample data. Any movement, facial expression, or statement by the interviewer can affect the response obtained. Errors in recording the response can also lead to erroneous results. A major problem with telephone surveys is the establishment of a frame that closely corresponds to the population. Telephone directories have many numbers that do not belong to households, and many households have unlisted numbers. A technique that avoids the problem of unlisted numbers is random digit dialing. In this method, a telephone exchange number (the first three digits of the seven-digit number) is selected, and then the last four digits are dialed randomly until a fixed number of households of a specified type are reached.

A mailed questionnaire sent to a specific group of interested persons can achieve good results, but, response rates for this type of data collection are generally so low that all reported results are suspect. Nonresponse can be a problem in any form of data collection, but since we have the least contact with respondents in a mailed questionnaire, we frequently have the lowest rate of response. The low response rate can introduce a bias into the sample because the people who answer questionnaires may not be representative of the population of interest. To eliminate some of this bias, investigators frequently contact the nonrespondents through follow-up letters, telephone interviews, or personal interviews.

Steps in Planning a Survey

(modified from Scheaffer, et al. *Elementary Survey Sampling, 5th Ed*., 1996. p. 68-70)

1. *Statement of objectives.* State the objectives of the survey clearly and concisely and refer to these objectives regularly as the design and the implementation of the survey progress. Keep the objectives simple enough to be understood by those working on the survey and to be met successfully when the survey is completed.

2. *Target population.* Carefully define the population to be sampled. If adults are to be sampled, then define what is meant by *adult* (all those over the age of 18, for example) and state what group of adults are included (all permanent residents of a city, for example). Keep in mind that a sample must be selected from this population and define the population so that sample selection is possible.

3. *The frame*. Select the frame (or frames) so that the list of sampling units and the target population show close agreement. Keep in mind that multiple frames may make the sampling more efficient. For example, residents of a city can be sampled from a list of city blocks coupled with a list of residents within blocks.

4. *Sample design.* Choose the design of the sample, including the number of sample elements, so that the sample provides sufficient information for the objectives of the survey.

5. *Method of measurement.* Decide on the method of measurement, usually one or more of the following methods: personal interviews, telephone interviews, mailed questionnaires, or direct observations.

6. *Measurement instrument.* In conjunction with step *5,* carefully specify how and what measurements are to be obtained. If a questionnaire is to be used, plan the questions so that they minimize nonresponse and incorrect response bias.

7. *Selection and training* of *field-workers.* After the sampling plan is clearly and completely set up, someone must collect the data. Those collecting data, the field-workers, must be carefully taught what measurements to make and how to make them. Training is especially important if interviews, either personal or telephone, are used because the rate of response and the accuracy of responses are affected by the interviewer's personal style and tone of voice.

8. *The pretest.* Select a small sample for a pretest. The pretest is crucial because it allows you to field-test the questionnaire or other measurement device, to screen interviewers, and to check on the management of field operations. The results of the pretest usually suggest that some modifications must be made before a full-scale sampling is undertaken.

9. *Organization* of *fieldwork.* Plan the fieldwork in detail. Any large-scale survey involves numerous people working as interviewers, coordinators, or data managers. The various jobs should be carefully organized and lines of authority clearly established before the survey is begun.

10. *Organization of data management.* Outline how each piece of datum is to be handled for all stages of the survey. Large surveys generate huge amounts of data. Hence, a well-prepared data management plan is of the utmost importance. This plan should include the steps for processing data from the time a measurement is taken in the field until the final analysis is completed. A quality control scheme should also be included in the plan in order to check for agreement between processed data and data gathered in the field.

11. *Data analysis.* Outline the analyses that are to be completed. Closely related to step 10, this step involves the detailed specification of what analyses are to be performed. It may also list the topics to be included in the final report.

12. *Final Report.* The final report should match the stated objectives in step 1. Considering the final report before the survey is conducted may be helpful in determining what items are to be measured in the survey.

13. *Recapitulation.* After the final report is completed, you should consider what changes should be made if/when the survey is repeated. Most surveys are conducted periodically. It is important to keep track of what went well and what difficulties occurred.

Stratification methods for the Gallop Poll and New York Times are presented below (quoted from Scheaffer, et al, *Elementary Survey Sampling, 5th Edition*, page 50-51):

The Gallup Poll

Although most Gallup poll findings are based on telephone interviews, a significant proportion is based on interviews conducted in person in the home. The majority of the findings reported in Gallup Poll surveys is based on samples consisting of a minimum of 1,000 interviews. The total number, however, may exceed 1,000, or even 1,500, interviews, where the survey specifications call for reporting the responses of low-incident population groups such as young public-school parents or Hispanics.

Design of the Sample for Telephone Surveys

The findings from the telephone surveys are based on Gallup's standard national telephone samples, consisting of unclustered directory-assisted, random-digit telephone samples utilizing a proportionate, stratified sampling design. The random-digit aspect of the sample is used to avoid "listing" bias. Numerous studies have shown that households with unlisted telephone numbers are different from listed households. "Unlistedness" is due to household mobility or to customer requests to prevent publication of the telephone number. To avoid this source of bias, a random-digit procedure designed to provide representation of both listed and unlisted (including not-yet-listed) numbers is used.

Telephone numbers for the continental United States are stratified into four regions of the country and, within each region, further arranged into three size-of-community strata. The sample of telephone numbers produced by the described method is representative of all telephone households within the continental United States.

Only working banks of telephone numbers are selected. Eliminating nonworking banks from the sample increases the likelihood that any sampled telephone number will be associated with a residence.

Within each contacted household, an interview is sought with the youngest man 18 years of age or older who is at home. If no man is home, an interview is sought with the oldest woman at home. This method of respondent selection within households produces an age distribution by sex that closely approximates the age distribution by sex of the total population.

Up to three calls are made to each selected telephone number to complete an interview. The time of day and the day of the week for callbacks are varied to maximize the chances of finding a respondent at home. All interviews are conducted on weekends or weekday evenings in order to contact potential respondents among the working population.

The final sample is weighted so that the distribution of the sample matches current estimates derived from the U.S. Census Bureau's Current Population Survey (CPS) for the adult population living in telephone households in the continental United States.

Design of the Sample for Personal Surveys

The design of the sample for personal (face-to-face) surveys is that of a replicated area of probability sample down to the block level in the case of urban areas and to segments of townships in the case of rural areas.

After stratifying the nation geographically and by size of community according to information derived from the most recent census, over 350 different sampling locations are selected on a mathematically random basis from within cities, towns, and counties that, in turn, have been selected on a mathematically random basis.

The interviewers are given no leeway in selecting the areas in which they are to conduct their interviews. Each interviewer is given a map on which a specific starting point is marked and is instructed to contact households according to a predetermined travel pattern. At each occupied dwelling unit, the interviewer selects respondents by following a systematic procedure that is repeated until the assigned number of interviews has been completed.

The New York Times

The latest New York Times/CBS News Poll is based on telephone interviews conducted from Sept. 8 to 11 with 1,161 adults around the country, excluding Alaska and Hawaii.

The sample of telephone exchanges called was selected by a computer from a complete list of exchanges in the United States. The exchanges were chosen to assure that each region of the country was represented in proportion to its population. For each exchange, the telephone numbers were formed by random digits, thus permitting access to both listed and unlisted numbers. Within each household, one adult was designated by a random procedure to be the respondent for the survey.

The results have been weighted to take account of the household size and the number of telephone lines into the residence, and to adjust for variations in the sample relating to region, race, sex, age and education.

In theory, in 19 cases out of 20 the results based on such samples will differ by no more than three percentage points in either direction from what would have been obtained by seeking out all American adults. For smaller subgroups the potential sampling error is larger. For example, for blacks it is plus or minus 10 percentage points.

In addition to sampling error, the practical difficulties of conducting any survey of public opinion may introduce other sources of error into the poll. Variations in question wording or the order of questions, for example, can lead to somewhat different results.