FAIRFAX COUNTY DEPARTMENT OF SYSTEMS MANAGEMENT FOR HUMAN SERVICES

April 2003

Informational Brochure

Overview of Sampling Procedures

Table of Contents

Technical References:

American Statistical Association
1429 Duke Street
Alexandria, VA 22314
703-684-1221
http://www.amstat.org

American Association for Public Opinion Research
P.O. Box 14263
Lenexa, KS 66285
913-310-0118
http://www.aapor.org

Council of American Survey Research Organizations
3 Upper Devon
Port Jefferson, NY 11777
631-928-6954
http://www.casro.org

Federal Committee on Statistical Methodology
Statistical Policy Working Papers
http://www.fcsm.gov

The purpose of this brochure is to provide an overview of the sampling procedures available to a researcher. The differences between the various sampling procedures are discussed and examples are provided to illustrate the use of these procedures. The emphasis of this manual is placed on underlying ideas and methods rather than detailed mathematical derivations. For the reader who is interested in pursuing a more thorough approach to the topics discussed, a list of technical references is provided in the box below.

Why Sample

If a researcher desires to obtain information about a population through questioning or testing, he/she has two basic options:

1. Every member of the population can be questioned or tested, a census; or

2. A sample can be conducted; that is, only selected members of the population are questioned or tested.

Contacting, questioning, and obtaining information from a large population, such as the 370,000 households residing in Fairfax County, is extremely expensive, difficult, and time consuming. A properly designed probability sample, however, provides a reliable means of inferring information about a population without examining every member or element.

Often, researchers are working under strict time constraints which make conducting a census unwieldy. For instance, national polling firms frequently must provide information on the public's perceptions of current events or issues. These polling firms tend to limit their national sample sizes to approximately 1,500 respondents. When properly conducted, a probability sample of this size provides reliable information with a very small margin of error for the whole population of the United States, which is nearly 300 million persons.

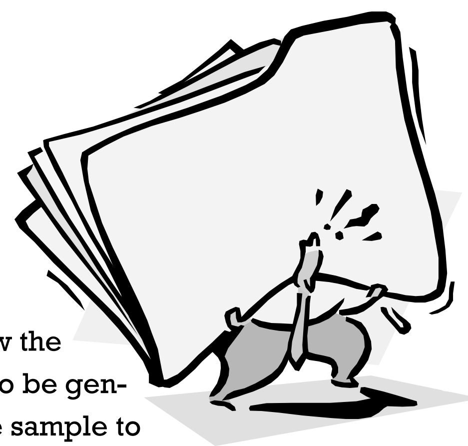
A probability sample frequently is more accurate than a census of the entire population. The smaller sampling operation lends itself to the application of more rigorous controls, thus ensuring better accuracy. These rigorous controls allow the researcher to reduce nonsampling errors such as interviewer bias and mistakes, nonresponse problems, questionnaire design flaws, and data processing and analysis errors. In part, these nonsampling errors are reduced through pretesting which allows careful testing of the survey questionnaire and procedures. Pretesting cannot be done when conducting a census without causing possible contamination of some of the respondents. The detail of information that can be asked in a sample is greater than that in a census due to the cost and time constraints under which most researchers are operating. A relatively long and difficult questionnaire can be administered to a sample more easily than a brief questionnaire can be administered to the entire population. However, not all samples are accurate or the appropriate vehicle for gathering information or testing a hypothesis about a population. The following sections of this brochure will briefly discuss the merits and disadvantages of various sampling procedures.

Sampling Methodologies

Sampling methodologies are classified under two general categories:

1. Probability sampling and
2. Nonprobability sampling.

In the former, the researcher knows the exact possibility of selecting each member of the population; in the latter, the chance of being included in the sample is not known. A probability sample tends to be more difficult and costly to conduct. *However, probability samples are the only type of samples where the results can be generalized from the sample to the population*. In addition, probability samples allow the researcher to calculate the precision of the estimates obtained from the sample and to specify the sampling error.



Nonprobability samples, in contrast, do not allow the study's findings to be generalized from the sample to the population. When discussing the results of a nonprobability sample, the researcher must limit his/her findings to the persons or elements sampled. This procedure also does not allow the researcher to calculate sampling statistics that provide information about the precision of the results. The advantage of nonprobability sampling is the ease in which it can be administered. Nonprobability samples tend to be less complicated and less time consuming than probability samples. If the researcher has no intention of generalizing beyond the sample, one of the nonprobability sampling methodologies will provide the desired information.

Nonprobability Samples

The three common types of nonprobability samples are convenience sampling, quota sampling, and judgmental sampling.

A. Convenience Sampling

As the name implies, convenience sampling involves choosing respondents at the convenience of the researcher. Examples of convenience samples include people-in-the-street interviews—the sampling of people to which the researcher has easy access, such as a class of students; and studies that use people who have volunteered to be questioned as a result of an advertisement or another type of promotion. A drawback to this methodology is the lack of sampling accuracy. Because the probability of inclusion in the sample is unknown for each respondent, none of the reliability or sampling precision statistics can be calculated. Convenience samples, however, are employed by researchers because the time and cost of collecting information can be reduced.

B. Quota Sampling

Quota sampling is often confused with stratified and cluster sampling—two probability sampling methodologies. All of these methodologies sample a population that has been subdivided into classes or categories. The primary differences between the methodologies is that with stratified and cluster sampling the classes are mutually exclusive and are isolated prior to sampling. Thus, the probability of being selected is known, and members of the population selected to be sampled are not arbitrarily disqualified from being included in the results. In quota sampling, the classes cannot be isolated prior to sampling and respondents are categorized into the classes as the survey proceeds. As each class fills or reaches its quota, additional respondents that would have fallen into these classes are rejected or excluded from the results.

An example of a quota sample would be a survey in which the researcher desires to obtain a certain number of respondents from various income categories. Generally, researchers do not know the incomes of the persons they are sampling until they ask about income. Therefore, the researcher is unable to subdivide the population from which the sample is drawn into mutually exclusive income categories prior to drawing the sample. Bias can be introduced into this type of sample when the respondents who are rejected, because the class to which they belong has reached its quota, differ from those who are used.

C. Judgmental Sampling

In judgmental or purposive sampling, the researcher employs his or her own "expert" judgment about who to include in the sample frame. Prior knowledge and research skill are used in selecting the respondents or elements to be sampled.

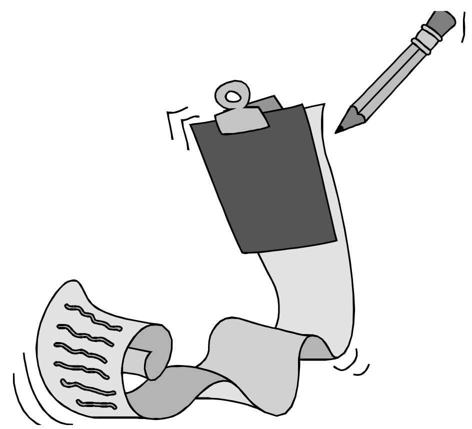
An example of this type of sample would be a study of potential users of a new recreational facility that is limited to those persons who live within two miles of the new facility. Expert judgment, based on past experience, indicates that most of the use of this type of facility comes from persons living within two miles. However, by limiting the sample to only this group, usage projections may not be reliable if the usage characteristics of the new facility vary from those previously experienced. As with all nonprobability sampling methods, the degree and direction of error introduced by the researcher cannot be measured and statistics that measure the precision of the estimates cannot be calculated.

Probability Samples

Four basic types of methodologies are most commonly used for conducting probability samples; these are simple random, stratified, cluster, and systematic sampling. Simple random sampling provides the base from which the other more complex sampling methodologies are derived.

A. Simple Random Sampling

To conduct a simple random sample, the researcher must first prepare an exhaustive list (sampling frame) of all members of the population of interest. From this list, the sample is drawn so that each person or item has an equal chance of being drawn during each selection round. Samples may be drawn with or without replacement. In practice, however, most simple random sampling for survey research is done without replacement; that is, a person or item selected for sampling is removed from the population for all subsequent selections. At any draw, the process for a simple random sample without replacement must provide an equal chance of inclusion to any member of the population not already drawn. To draw a simple random sample without introducing researcher bias, computerized sampling programs and random numbers tables are used to impartially select the members of the population to



be sampled.

An example of a simple random sample would be a survey of County employees. An exhaustive list of all County employees as of a certain date could be obtained from the Department of Human Resources. If 100 names were selected from this list using a random number table or a computerized sampling program, then a simple random sample would be created. Such a random sampling procedure has the advantage of reducing bias and enables the researcher to estimate sampling errors and the precision of the estimates derived through statistical calculations.

B. Stratified Random Sampling

Stratified random sampling involves categorizing the members of the population into mutually exclusive and collectively exhaustive groups. An independent simple random sample is then drawn from each group. Stratified sampling techniques can provide more precise estimates if the population being surveyed is more heterogeneous than the categorized groups, can enable the researcher to determine desired levels of sampling precision for each group, and can provide administrative efficiency.

An example of a stratified sample would be a sample conducted to determine the average income earned by families in the United States. To obtain more precise estimates of income, the researcher may want to stratify the sample by geographic region (northeast, mid-Atlantic, et cetera) and/or stratify the sample by urban, suburban, and rural groupings. If the differences in income among the regions or groupings are greater than the income differences within the regions or groupings, precision of the estimates is improved. In addition, if the research organization has branch offices located in these regions, the administration of the survey can be decentralized and perhaps conducted in a more cost-efficient manner.

C. Cluster Sampling

Cluster sampling is similar to stratified sampling because the population to be sampled is subdivided into mutually exclusive groups. However, in cluster sampling the groups are defined so as to maintain the heterogeneity of the population. It is the researcher's goal to establish clusters that are representative of the population as a whole, although in practice this may be difficult to achieve. After the clusters are established, a simple random sample of the clusters is drawn and the members of the chosen clusters are sampled. If all of the elements (members) of the clusters selected are sampled, then the sampling procedure is defined as one-stage cluster sampling. If a random sample of the elements of each selected cluster is drawn, then the sampling procedure is defined as two-stage cluster sampling.

Cluster sampling is frequently employed when the researcher is unable to compile a comprehensive list of all the elements in the population of interest. A cluster sample might be used by a researcher attempting to measure the age distribution of persons residing in Fairfax County. It would be much more difficult for the researcher to compile a list of every person residing in Fairfax County than to compile a list of residential addresses. In this example, each address would represent a cluster of elements (persons) to be sampled. If the elements contained in the clusters are as heterogeneous as the population, then estimates derived from cluster sampling are as precise as those from simple random sampling. However, if the heterogeneity of the clusters is less than that of the population, the estimates will be less precise.

D. Systematic Sampling

Systematic sampling, a form of one-stage cluster sampling, is often used in place of simple random sampling. In systematic sampling, the researcher selects every nth member after randomly selecting the first through nth element as the starting point. For example, if the researcher decides to sample every 20th member of the population, a 5 percent sample, the starting point for the sample is randomly selected from the first 20 members. A systematic sample is a type of cluster sample because each of the first 20 members of the sampling frame defines a cluster that contains 5 percent of the population.

A researcher may choose to conduct a systematic sample instead of a simple random sample for several reasons. Systematic samples tend to be easier to draw and execute. The researcher does not have to jump backward and forward through the sampling frame to draw the members to be sampled. A systematic sample may spread the members selected for measurement more evenly across the entire population than simple random sampling. Therefore, in some cases, systematic sampling may be more representative of the population and more precise.

One of the most attractive aspects of systematic sampling is that this method can allow the researcher to draw a probability sample without complete prior knowledge of the sampling frame. For example, a survey of visitors to the County's publications desk could be conducted by sampling every 10th visitor after randomly selecting the first through 10th visitor as the starting point. By conducting the sample in this manner, it would not be necessary for the researcher to obtain a comprehensive list of visitors prior to drawing the sample.

As with other types of cluster sampling, systematic sampling is as precise as simple random sampling if the members contained in the clusters are as heterogeneous as the population. If this assumption is not valid, then systematic sampling will be less precise than simple random sampling. In conducting systematic sampling, it is also essential that the researcher does not introduce bias into the sample by selecting an inappropriate sampling interval. For instance, when conducting a sample of financial records, or other items that follow a calendar schedule, the researcher would not want to select "7" as the sampling interval because the sample would then be comprised of observations that were all on the same day of the week. Day-of-the-week influences may cause contamination of the sample, giving the researcher biased results.

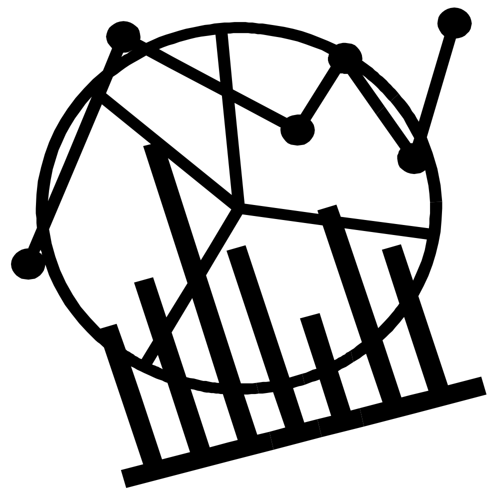
Summary

Sampling can be a powerful tool for accurately measuring opinions and characteristics of a population. However, there is a genuine potential for misuse of this tool by researchers who do not understand the limitations of various sampling procedures. The differences between nonprobability and probability sampling procedures are often difficult to discern but are extremely important for determining how the results of the research can be used. Nonprobability sampling techniques can provide valuable information but the results cannot be generalized to a larger population nor can statistics indicating the reliability of the results be calculated. Well conducted probability samples provide the researcher with the ability to gather information from a relatively small number of members of a large population and accurately generalize the results to the entire population. In addition, probability samples enable the researcher to calculate statistics that indicate the precision of the data.

Glossary

- *Bias (error):* Distorted or unreliable survey results. All surveys contain some bias. Bias is increased when the respondents (persons answering the survey) are not representative of the population being questioned, when questions are poorly written or misunderstood, and when the researcher uses inappropriate techniques to analyze the data.
- *Census*: A study using all available elements (members) of a population.
- *Data*: The collection of observations and information resulting from the survey process.
- *Element (member)*: The basic unit about which survey information is sought (i.e., person, business, household, car, dog, et cetera).
- *Heterogeneous*: A population whose elements have dissimilar characteristics. Heterogeneity is the state of being dissimilar.
- *Homogeneous*: A population whose elements have similar characteristics. Homogeneity as the state of being similar.
- *Instrument*: The tool or device used for survey measurement; usually a questionnaire.
- *Nonresponse*: Unit nonresponse refers to the refusal of persons selected to be sampled to participate in a survey (i.e., person does not return the mail questionnaire). Item nonresponse refers to selected questions left unanswered by the person surveyed.

- *Population*: The universe or collection of all elements (persons, businesses, et cetera) being described or measured by a sample.
- *Pretest*: An initial evaluation of the survey design by using a small, subsample of the intended population for preliminary information.
- *Questionnaire*: A measuring device used to query a population/sample in order to obtain information for analysis.
- *Respondent*: An element or member of the population selected to be sampled.
- *Sample*: Any portion of the population, less than the total.
- *Sampling Frame*: An exhaustive list of all members of the population from which a sample can be drawn.
- *Survey*: A process of inquiry for the purpose of data collection and analysis using observation, polls, questionnaires, and/or interviews.
- *Statistics*: Descriptive measures based upon a probability sample.



Fairfax County Department of Systems Management for Human Services

Economic and Demographic Research
12011 Government Center Parkway
Suite 222
Fairfax, Virginia 22035

Phone: 703-324-4519
TTY: 711 (Virginia Relay)
Fax: 703-803-8598

Building Upon the Strengths of a Diverse Community

We're on the Web!
www.fairfaxcounty.gov/aboutfairfax