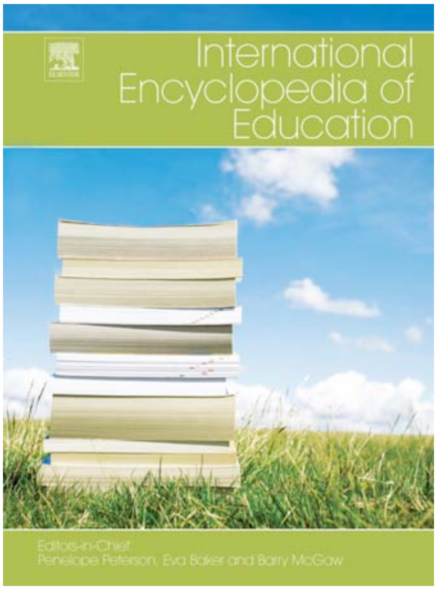

Brown R S (2010), Sampling. In: Penelope Peterson, Eva Baker, Barry McGaw, (Editors), International Encyclopedia of Education. volume 3, pp. 142-146. Oxford: Elsevier.

Sampling

R S Brown, University of Southern California, Los Angeles, CA, USA

Introduction

Throughout the course of our everyday existence, we make judgments, inferences, and decisions based on less than the full collection of potential data. We rate restaurants and hotels based on a few meals or several nights of occupancy. We judge athletes based on samples of athletic performance, appraise students based on relatively few assignments, assessments, or tests, and form impressions of new acquaintances based on limited interactions. Sampling is a pervasive and important part of how we interact with and form judgments about the world in which we live. Sampling occurs both informally – as in the instances stated above – and formally as in virtually all governmental surveys, market research polls, and research studies. This article attempts to provide an overview of sampling for research purposes, including terms and definitions, along with various sampling designs.

The basic premise behind sampling is that we seek to gather information from a subset of a larger group of potential respondents in order to make judgments or inferences back to the larger group. As Pedhazur and Schmelkin (1991) state, "Formal sampling is a process aimed at obtaining a representative portion of some whole, thereby affording valid inferences and generalizations to it" (p. 318). Thus, the goal is to identify a group that can adequately represent the whole – so that inferences and generalizations can be made from their responses to survey items or other measures back to the larger population, with some known, or estimable, amount of imprecision. Sampling theory seeks to develop sample-selection strategies and estimation methods that generate sufficiently precise estimates at minimum cost, thereby making sampling more efficient.

There are a number of advantages to sampling compared to collecting data from an entire population. These include reduced cost, greater speed, enhanced feasibility, and greater accuracy. Clearly, cost can be reduced by limiting the amount of data collected to a sample rather than the full population of potential respondents. Likewise, collecting less data would require less time. This consideration for reducing the amount of time necessary for data collection becomes important in situations in which the outcome measure being collected may be influenced by changing events. In volatile social circumstances, early respondents' reactions may differ from later respondents' reactions as influences on the outcomes change over an extended data-collection period. Reducing the timeframe for data collection by employing a sampling strategy can minimize this impact. It is also possible to enhance the feasibility of a study by sampling. In some cases, a complete census is simply impractical and without sampling the study could not be accomplished. In other cases, a census approach would be infeasible for other reasons – such as lack of knowledge of the consequences of a treatment or intervention. It would not be advisable or appropriate to administer a new drug to the entire population, or to compel all students to undertake a given curriculum without first knowing whether the drug is safe and effective or that the curriculum generates enhanced knowledge acquisition (Pedhazur and Schmelkin, 1991). Sampling can also enhance data accuracy by providing more control over the training and supervision of the collection and recording of data, as well as following up on nonresponses. In addition, sampling frees up valuable resources that can be used in the instrument-development and -refinement stages of an investigation.

Sampling Concepts and Terminology

Population

The population refers to the "aggregate from which the sample is chosen" (Cochran, p. 5). It generally refers to the collection or larger grouping to which one wishes to infer or generalize. In some cases, this can be quite straightforward. In a factory setting, the population may be the total collection of widgets produced in a given month or year. A sample of those widgets may be useful for providing sample statistics that estimate population parameters of interest, such as size, weight, or density. In other cases, the population must be defined explicitly in terms of content, units, extent, and time. An example of a given population could be all children attending preschool in California during the 2009–10 academic years.

Sample Units

The sample units are the elements of the population to be sampled. In aggregate, they define the population to which inferences are to be drawn. The sample units may be individuals, classrooms, schools, or a variety of possible elements, but they are the units from which data are collected and serve as the unit of analysis. Sample units may include single elements from the population or clusters of elements from the population, such as when a dwelling may be the sample unit and individual persons would be the elements from the population. In the example-population given above, the sample units would be the individual children attending preschool in California during the 2009–10 academic years.

Sampling Frame

The sampling frame is the list of sample units from which the sample is drawn. A perfect frame is one in which each element in the population is separately listed once, only once, and no other population irrelevant or extraneous elements are listed. However, not all sampling frames are perfect, and it takes effort and attention to review potential sampling frames to ensure that they are free from error or that the errors in the frames can be addressed. Kish (1965) identifies four primary problems with sampling frames. The first problem is that some sampling frames are incomplete; thus, they do not include all available elements or sample units from the population. Another problem involves clustering of elements within a single listing, which violates the notion that each element be listed separately. A third problem deals with blanks listing or foreign listings in a sampling frame. This violates the rule that each listing should contain a single element. The fourth primary problem with sampling frames is that occasionally duplicate elements appear in the listing. This violates the edict that each element should be listed only once in a sampling frame.

Selection Process

Following the establishment of an appropriate sampling frame, a selection process must be identified, of which there are numerous considerations depending on the study purpose, resources, timeframe, and sample structure. The process depends upon the list from the sampling frame. As Kish states, "The sampling frame or list is the keystone around which the selection process must be designed" (Kish, p. 53). The selection process may involve probability or nonprobability approaches, which are described in more detail below.

Estimation Process

Samples provide data from which sample statistics can be calculated. These sample statistics are used to estimate population values. A common sample statistic of interest is the mean value for the sample. This mean value is a random variable dependent upon the sample of n units from which it is derived. Another sample of size n from the population may result in a different sample mean. The distribution of sample means (the sampling distribution of the mean) allows for the calculation of the population mean and variance (or standard deviation). In practice, not all samples are drawn, so not all sample means are available, thus a sample design is used. A sample design is unbiased if the expected value of the sample mean (i.e., the mean value of the sample means) is equal to the population mean. Estimates of the population variance can be obtained from the sample statistics and used to develop standard errors of the mean and confidence intervals around the population mean estimates. The magnitude of the sampling error is reflected in the standard errors of the mean and the confidence intervals provide a range of scores that can be interpreted to provide confidence in the population mean estimate. A much more detailed treatment of estimation formulas and estimates is presented in Cochran (1977).

Probability Sampling

There are a number of probability-sampling approaches. All of these approaches require that each element in the population has a known, nonzero probability of being selected into the sample. Probability-sampling approaches rely heavily on random selection. In fact, Kish states emphatically, "Probability selection demands randomized selection. When randomization is both simple and important, disregarding it amounts to carelessness or ignorance" (Kish, p. 28). Utilization of random-selection procedures such as random-number generators and tables of random numbers is important because human beings have shown they are incapable of identifying a random sample independent of such protocols. As Kish states, "Personal judgment has been shown inadequate for selecting random samples of integers, or stones from a pile, or plants from a field, or people on streets or in homes" (Kish p. 29).

Simple Random Sampling

Simple random sampling is a method of selecting a sample size, n, from the population of N elements such that each of the ${}_{N}C_{n}$ distinct samples has an equal probability of being selected. This is usually done by selecting at the unit level, with all units being numbered from 1 to N. A table of random numbers or a computer-generated list of random numbers can then be applied to each unit. A total of

$$N!/[(N-n)!n!]$$

distinct sample of size n can be drawn. For example, with a population of size 10, and sample size of 3, a total of 120 unique, nonoverlapping samples can be drawn. From a practical standpoint, simple random sampling is not frequently utilized in large-scale studies. In addition, many studies are interested in more than just the effects on the general populations, for instance many studies are concerned with the effects on subpopulations. Thus, additional sampling strategies have been developed.

Systematic Sampling

Systematic sampling is similar to simple random sampling in that all N elements of the population are numbered and ordered from 1 to N. However, unlike in simple random sampling in which each of n elements is then randomly chosen, in systematic sampling a element is chosen from the first k elements (k is often defined as N/n), then every kth element is chosen until n elements are selected. For instance, if k is 11 and the first element selected is 7, then elements 18, 29, 40, and 51 would be chosen to complete the sample of $n = 5$. Systematic sampling has the advantage of being easier to execute in some instances than simple random sampling, and may be more precise than both simple random sampling and stratified random sampling.

Stratified Random Sampling

Stratified random sampling utilizes known information about the population elements to separate the sample units into nonoverlapping groups, or strata, from which they are then randomly selected. For example, a population of fourth-grade school children may be stratified into various geographic regions, or types of schools attended. Therefore, the full population, N, is subdivided into k strata, such that

$$N = N_1 + N_2 + N_3 + \ldots N_k.$$

Once the strata have been determined, a simple random sample is selected from within each stratum. There are a number of advantages to stratified random sampling. One such advantage is that it provides more precise estimates of the subpopulation parameters than does simple random sampling. Stratified sampling can reduce sampling variability by creating relatively homogeneous subsets of sample units with respect to the outcome variable when the stratifying variable is correlated with the outcome measure, although the gains in precision are often minimal. Another potential advantage to stratified sampling is that sampling frames for naturally occurring strata may be readily available, such as student rosters within schools as opposed to all second-grade students in a given country.

There are disadvantages to stratified sampling procedures as well as potential advantages. Stratified sampling designs are more complex both in the selection process and in the estimation process. In addition, some investigations are concerned with multiple outcomes measures, not all of which will necessarily be well correlated with the stratifying variable. In this case, using a stratified sampling design might be beneficial for one outcome measure, but less beneficial for others. Another consideration is the determination of which variables should be used for establishing the strata, and how many strata are optimal. Kish (1965) suggests that the initial advantage obtained by stratification diminishes with increasing numbers of strata, and that from three to ten strata should be sufficient for any single outcome variable.

Proportionate Sampling

In proportionate sampling, strata sample sizes are held proportional to their existence in the population. This is often used to show the representativeness of a sample relative to the population's characteristics. The sampling fraction for each stratum is held constant and equivalent to the sampling fraction for the overall population. For example, assume that a study seeks a sample of 200 from a population of 1000 (thus the sampling fraction, f, for the population is 200/1000 or 1/5). The population is divided into three strata of differing sizes: $N_1 = 500$, $N_2 = 200$, and $N_3 = 300$. The overall sampling fraction, $f = 0.2$, is applied to each stratum to yield samples of $n_1 = 100$, $n_2 = 40$, and $n_3 = 60$. As a result, the proportions of selected sample units in each stratum match the proportions of the strata units relative to the population (e.g., $N_1 = 500/1000 = 0.5 = n_1/n = 100/200 = 0.5$). The advantage of this approach is that it generates a self-weighting sample and the population mean can be estimated from the simple mean of the sample elements.

Cluster Sampling

Cluster sampling is a probability-sampling design that capitalizes on naturally occurring groups, or clusters, in the population. Examples of such naturally occurring groups are students within a classroom or school, residents of a city block, or patients at a given medical facility. In cluster sampling, the groups or clusters are first randomly selected; then, all members of that group are included in the study. Thus, if the clusters are schools and school A is chosen, all the students (or all sixth-graders, depending upon the sampling frame and study intent) in school A would participate. Cluster sampling has the advantage of reducing cost and time associated with sampling and data collection. However, the selected clusters need to represent the population of clusters. When the clusters are more heterogeneously composed internally, this is less problematic. But when the clusters are more internally homogeneous with respect to the outcome measures, a considerable loss of precision in estimation can occur. This leads Kalton (1983) to conclude that unless the economy of using a cluster-sampling design allows for a sufficient increase in sample size to override the associated loss in estimator precision, cluster-sampling designs should be avoided. In addition to the heterogeneity within a cluster, the variability of the size of clusters makes a difference in estimation using cluster sampling. Since most naturally occurring clusters are of unequal size, this variation in cluster size should be taken into account. Large variability of cluster sizes leads to less precise estimation, unless adequately accounted for. However, additional sampling designs, such as selection with probabilities proportionate to size provides approximate control over the influence of unequal cluster sizes.

Multistage Sampling

Multistage sampling is an extension of cluster sampling in that, first, clusters are randomly selected and, second, sample units within the selected clusters are randomly selected. In this design, random selection occurs at both the cluster or group level and at the sample unit level. Multistage sampling also may be useful when naturally occurring cluster sizes are rather large, resulting in reduced precision when compared to the stratified random-sampling approach. In this event, smaller clusters can be created and sampled. For example, rather than sampling all sixth graders at a selected school, a secondary cluster, such as classes, could be utilized and sampled. In this case, only students in sampled classes at sampled schools would be included. This has the added advantage of reducing the cluster size, thereby enhancing estimator precision.

Double Sampling

Double sampling refers to the practice of selecting an initial sample in order to obtain information on a potential stratifying variable, then using that information to develop a stratified random sampling plan to obtain information on the outcome measure of interest. This is frequently done when the distributional characteristics of the stratifying variable, x, are unknown in the population. The advantage of this approach is that it provides information that may be used to create strata for the sampling design for the collection of data on the outcome variable, y, of primary interest. The disadvantage is that resources that might otherwise be utilized in data collection and analysis for information on y must be spent to collect information on x. Thus, it is important that the increase in precision of estimation due to stratification offsets the loss of precision due to reduced sample size in the data collection for variable y. This technique is also, sometimes, referred to as two-stage sampling.

Nonprobability Sampling

There are a number of nonprobability sampling techniques that have been identified in the sampling literature. While these approaches may yield subjects from whom data may be collected, these design do not benefit from the main advantage of probability approaches, namely, that probability designs allow for the development of statistical theory to examine the properties of sample estimators (Kalton, 1983). Nevertheless, these approaches are used widely in social science research and they include convenience sampling, purposive sampling, or quota sampling. In convenience sampling, a sample is selected based on ready availability – such as students in a given classroom, or passers-by on a street corner or in front of a busy market. Other convenience samples may include respondents to an advertisement in a magazine or dial-in number for a reality television program. Purposive sampling differs from convenience sampling in that certain characteristics of the sample are sought out *a priori*; that is, a sample that possesses certain characteristics, often to be seen as representative of some larger population is sought. One example of a purposive sample may involve a researcher selecting communities from across a state to ensure geographic diversity.

Another nonprobability sampling approach is quota sampling. In quota sampling, the researcher attempts to collect data on a specified number of respondents in each of a number of groups of potential respondents. For example, a researcher may seek responses from 20 third-grade teachers, 20 fourth-grade teachers, 20 eighth-grade teachers, and 20 tenth-grade teachers. One advantage of quota sampling is that it can reduce data-collection time and associated costs. It is somewhat similar to stratified sampling in that it seeks to obtain responses from more homogeneous subsets of the total populations. However, the major and pivotal distinction is that with stratified random sampling, the sample units within the strata are randomly selected, whereas in quota sampling, the sample units within the quota groups are not. Further, while some may argue that quota sampling reduces nonresponse, the reality is that quota sampling merely replaces nonrespondents with other respondents in the quota group, thereby underestimating the responses of hard-to-find or unwilling sample units.

The fundamental problem with nonprobability sampling designs is that they are potentially biased in their sample estimators, and the magnitude of this bias is unknown. What is known is that the concern with regard to the bias in sample estimators increases with sample size, since probability-sampling designs become more precise with larger samples. Thus, whereas bias in sample estimators may be of less concern when deploying nonprobability sampling approaches for small-scale studies, more care should be taken to avoid nonprobability designs with larger research efforts. These sampling approaches may be utilized widely, but as Pedhazur and Schmelkin state, "..., the incontrovertible fact is that, in nonprobability sampling, it is not possible to estimate sampling errors. Therefore, validity of inferences to a population cannot be ascertained" (p. 321).

Sample Size and Power

One of the most common questions that surfaces when dealing with discussions on sampling deals with how large a sample is required for a given purpose. Samples should be large enough to provide the representativeness desired, but the precision of estimation is also a critical component. Since sample size is a primary component in determining estimation precision, once an acceptable level of precision is identified, the optimum samples for such precision can be readily calculated (see Cochran, 1977, for a more thorough treatment of optimal sample sizes for various sampling designs). The determination of what constitutes an acceptable amount of precision is not easily made, and is determined, in part, by the purpose of the research, practical and economic constraints, and the estimated effect size (ES) for a given treatment or intervention. Sample size also affects the likelihood of identifying statistically significant effects or differences among groups.

Of significant importance is the requisite sample size to provide a research investigation with sufficient power to validly and statistically explore the research questions of interest. Sociobehavioral research is full of analytical overviews of research indicating that, in many disciplines, much of the published research fails to establish high levels of statistical power – primarily due to inadequate sample sizes. The statistical power of an analysis is determined by three components. These include: the ES, the probability of rejecting the null hypothesis when it should not be rejected (α), and sample size (N). The power of a statistical test is defined as $1 - \beta$, or 1 minus the probability of failing to reject the null hypothesis when it should be rejected (β). Thus, for a given or expected ES with an *a priori* determined α-level – given statistical analysis, and desired power level – the appropriate sample size can be calculated. Tables of such values can be found in Cohen (1988). In addition, computer programs are widely available (e.g., G*Power 3, Power and Precision, Power Analysis and Sample Size (PASS) are but a few of the many programs available) to assist the researcher in estimating necessary sample sizes given a desired power level, expected effect size, and α-level.

Conclusion

Sampling is a necessary and important component of much of the data-collection efforts we undertake both in informal and formal settings. Understanding the importance of sampling, the various approaches to collecting adequate samples, the advantages and disadvantages of different sampling procedures, and the relationship of sample size to statistical power enables researchers to make more valid inferences and generalizations from sample data to the population of interest, and thereby increases our knowledge of the domain in which they operate.

Bibliography

Cochran, W. G. (1977). *Sampling Techniques*, 3rd edn. New York: Wiley.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.

Kalton, G. (1983). *Introduction to Survey Sampling*. Newbury Park, CA: Sage.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Pedhazur, E. J. and Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Erlbaum.

Further Reading

Babbie, E. (1990). *Survey Research Methods*, 2nd edn. Belmont, CA: Wadsworth.

Fink, A. (1995). *How to Sample in Surveys*. Thousand Oaks, CA: Sage.

Fowler, F. J. (2002). *Survey Research Methods*, 3rd edn. Newbury Park, CA: Sage.

Henry, G. T. (1990). *Practical Sampling*. Newbury Park, CA: Sage.

Jaeger, R. M. (1984). *Sampling in Education and the Social Sciences*. New York: Longman.

Konjin, H. S. (1973). *Statistical Theory of Sample Survey Design and Analysis*. New York: Elsevier.

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