Representative Sampling

Jessica Choe, Lisa K. Lashley, Charles J. Golden

Nova Southeastern University

The research process begins with an initial observation that scientists or researchers of various backgrounds want to engage in and understand, which prompts them to formulate theories and hypotheses, collect data, and use statistical procedures to organize, summarize, and interpret gathered data. Research conducted through questionnaires or surveys are often utilized to determine the nature of a population or the interests of members in a particular group.

Because it is not practical to survey every individual member of a population, the use of sampling procedures allow scientists to obtain data from a set of observations drawn from a larger population, which may closely represent the larger population. In other words, representative sampling can estimate or infer the nature of a population from the characteristics of a sample taken from the population.

The two main sampling categories of research are probability samples and nonprobability samples. Nonprobability samples are not random; thus, an individual lacks an equal and independent chance of being selected. In contrast to nonprobability samples, probability samples (i.e., representative sampling) are most desirable to researchers because it may provide an adequate representation of the characteristics of the population.

Although not required, a randomization of the samples within probability samples may help to reduce bias and provide more reliable results. Randomization also means that each member within a population of interest has an equal and independent chance of being selected for the sample to represent the larger population. Thus, a random sampling of the population may control for bias. Earlier research showed that representative sampling can be utilized through large-scale surveys if it is not feasible to collect information from every individual within a very large

population (e.g., urban population) in the United States. Therefore, a method of representative sampling was used to collect a random sample of 800,000 households in 84 cities arising from the geographic distribution of the urban population in the US, which reflected the overall population.

Two types of representative sampling, including simple random samples and stratified random samples, are considered here. Simple random samples select members of a group through a method where every individual of a population has an equal and independent chance of being represented. This type of sample can best approximate the value of the population when the sample is larger. The second type of representative sampling known as stratified random sampling involves individuals who are selected in a cross section or subdivided into an equal number of groups relative to one or more stratifying variables that mirror the population.

A cross section of the population means that a sample of individuals in which the groups are defined by some classification are represented in proportion to the number of individuals in the larger population. A stratified sample or subpopulation then accurately represents and characterizes a population. For instance, if a researcher is aware that the population contains 60 percent females and 60 percent males, a gender-stratified random sample should contain six females for every four males. In the case of U.S. national public opinion polls, such as the *Fortune* polls, specific stratifying variables has been used to form a sample representative of populations with respect to age, sex, color, and economic status for each population subgroup according to the areas chosen to represent major geographical districts or cities.

The concept of representative sampling dates to 1895, where Kiaer proposed the idea of representative sampling to describe the importance of the relationship in comparing sample results with population. Kiaer considered the representative sampling as miniature representation of the population. Since then, ideas concerning representative sampling have been argued, tested, and

modified by other statistician advocates, including Wilbur (1896), Bowley (1903), and Neyman (1935), to constitute major advances in the understanding of representative sampling over time. To date, representative sampling has been utilized in many community-based studies, public opinion polls, large-scale surveys, and other relevant research.

Wilks, a mathematical statistician, posited that poll results obtained from representative samples can be relied on the opinions of the whole population when the assumptions and rules of the representative sampling have been followed. The assumptions of the representative sampling are that the method which have been used to subdivide the population group is a "relevant" method such that it yields subgroups within relatively more consistent opinions than a process that would involve a multitude of opinions thrown together. Furthermore, the assumption of the sampling is representative with respect to the subgroup and is random within each subgroup. It has been suggested that the larger the sample, the more likely the poll results will reflect an accurate and true representative sample that closely reflects the population. Increasing the sample size can also bring the result closer to a higher probability level within confidence limits, such that the probability level of 0.99 indicates a higher chance of the sample more adequately representing the population.

In sum, representative sampling is a practical statistical procedure that can provide information and a more reliable and valid estimation regarding the nature of a population based on the characteristics of a sample taken from a target population.

Further Reading

Huck, S. W. (2012). *Reading statistics and research (6th ed.).*New York:Pearson Education.

Kruskal, W., & Mosteller, F. (1980). Representative sampling, IV: The history of the concept in statistics, 1895-1939. *International Statistical Review/Revue Internationale de Statistique*, 169-195.

Stephan, F. F. (1939). Representative sampling in large-scale surveys. *Journal of the American Statistical Association*, *34*(206), 343-352.

Wilks, S. S. (1940). Representative sampling and poll reliability. *Public Opinion Quarterly*, *4*(2), 261-269.