Sampling Design and Techniques in Research

Dr. Fulwinder Pal Singh
Associate Professor of Economics cum
Principal SBJS Khalsa College,
Satlani Sahib, Amritsar-143107

Abstract

In business research, companies must often generate samples of customers, clients, employees, and so forth to gather their opinions. Sample design is also a critical component of marketing research and employee research for many organizations. During sample design, firms must answer questions such as: - What is the relevant population, sampling frame, and sampling unit? - What is the appropriate margin of error that should be achieved? - How should sampling error and non-sampling error be assessed and balanced? Sampling is a process of selecting samples from a group or population to become the foundation for estimating and predicting the outcome of the population as well as to detect the unknown piece of information. A sample is the sub-unit of the population involved in your research work. There are a few advantages and disadvantages associated with the sampling process. Among the advantages are that sampling can save cost and human resources during the process of research work. In ICT, sampling does not cause much constraint such as heavy use of tools and technology in predicting the research output.

Introduction

A sample design is made up of two elements.

- **Sampling method**. Sampling method refers to the rules and procedures by which some elements of the population are included in the sample. Some common sampling methods are simple random sampling , stratified sampling , and cluster sampling .
- **Estimator**. The estimation process for calculating sample statistics is called the estimator. Different sampling methods may use different estimators. For example, the formula for computing a mean score with a simple random sample is different from the formula for computing a mean score with a stratified sample. Similarly,

the formula for the standard error may vary from one sampling method to the next.

The "best" sample design depends on survey objectives and on survey resources. For example, a researcher might select the most economical design that provides a desired level of precision. Or, if the budget is limited, a researcher might choose the design that provides the greatest precision without going over budget.

Although one could employ a census to measure the entire population, it is more common to take a sample of the population. A properly designed probability sample can be used to make estimates for not only the sample itself, but also for the underlying population from which it was selected. A probability sample is one in which each element of the (underlying) population has a known and non-zero chance of being selected. That is, every person has a chance to be included in the study and have his or her characteristics, opinions, etc., become part of the data. It should be noted that everyone does not have to have an equal chance of being selected – just a known non-zero chance of being selected. Probability samples have several desirable characteristics. They enable us to put a margin of error or confidence interval on our estimates – essentially a measure of how accurate the estimate is compared to the same estimate calculated on the full population. Probability samples make it possible not only to compare the sample to the population, but also to compare a sample from one population to a sample from another population, as occurs in **multinational, multicultural, or multiregional** surveys, which we refer to as **“3MC”** surveys. An optimal sample design is one that maximizes the amount of information obtained per monetary unit spent within the allotted time and meets the specified level of precision.

One important prerequisite for 3MC surveys is that all samples are full probability samples from comparable target populations. Different nations have different sampling resources and conditions. For a multinational survey, this means that the optimal sample design for one country may not be the optimal design for another. Therefore, allowing each participating country some flexibility in its choice of sample design is highly recommended, so long as all sample designs use probability methods at each stage of selection .It encourages survey organizers to allow sample designs to differ among participating countries while, at the same time, ensuring standardization on the principles of probability sampling. Although flexibility will usually be necessary in sample design and implementation are only directly applicable if a probability framework is followed. Probability designs can be expensive. To reduce costs, some survey organizations select non probability samples, including convenience samples and quota samples. While not all non probability samples are biased, the risk of bias is extremely high and, most importantly, cannot be measured—a survey that uses non probability sampling cannot estimate the error in the sample estimates.

SAMPLE SIZE AND SELECTION

Most of the new researchers always wonder about the sample size that needs to be selected. You must remember that the larger the sample for your research, the better outcome you can evaluate at the end of the research process. The larger the sample, the more likely the sample mean and standard deviation will become a representation of the population mean and standard deviation. For instance, in IT survey, the sample size required depends on the statistical outcome needed for the findings. The objective of selecting a sample is to achieve maximum accuracy in your estimation within a given sample size and to avoid bias in the selection of the sample. This is important as bias can attack the integrity of facts and jeopardize your research outcome.

A perfect sample would be like a scaled-down version of the population, mirroring every characteristic of the whole population. Of course, no such perfect sample can exist for complicated populations (even if it did exist, we would not know it was a perfect sample without measuring the whole population). But a good sample will reproduce the characteristics of interest in the population, as closely as possible. It will be representative in the sense that each sampled unit will represent the characteristics of a known number of units in the population.

Some definitions are needed to make the notion of a good sample more precise.

Observation unit An object on which a measurement is taken. This is the basic unit of observation, sometimes called an element. In studying human populations,observation units are often individuals.

Target population The complete collection of observations we want to study. Defining the target population is an important and often difficult part of the study. For example, in a political poll, should the target population be all adults eligible to vote? All registered voters? All persons who voted in the last election? The choice of target population will profoundly affect the statistics that result.

Sample A subset of a population.

Sampled population The collection of all possible observation units that might have been chosen in a sample; the population from which the sample was taken.

Sampling unit The unit we actually sample. We may want to study individuals but do not have a list of all individuals in the target population. Instead, households serve as the sampling units, and the observation units are the individuals living in the households.

Sampling frame The list of sampling units. For telephone surveys, the sampling frame might be a list of all residential telephone numbers in the city; for personal interviews, a list of all street addresses; for an agricultural survey, a list of all farms or a map of areas containing farms.

In an ideal survey, the sampled population will be identical to the target population, but this ideal is rarely met exactly. In surveys of people, the sampled population is usually smaller than the target population.

Sample Design. A **sample design** is made up of two elements. **Sampling** method. **Sampling** method refers to the rules and procedures by which some elements of the population are included in the **sample**. Some common **sampling** methods are simple random **sampling** , stratified **sampling** , and cluster **sampling** .

A good sample will be as free from selection bias as possible. Selection bias occurs when some part of the target population is not in the sampled population. If a survey designed to study household income omits transient persons, the estimates from the survey of the average or median household income are likely to be too large. A sample of convenience is often biased, since the units that are easiest to select or that are most likely to respond are usually not representative of the harder-to-select or non responding units.

Sampling Error

Error can occur during the sampling process. Sampling error can include both systematic sampling error and random sampling error. Systematic sampling error is the fault of the investigation, but random sampling error is not. When errors are systematic, they bias the sample in one direction. Under these circumstances, the sample does not truly represent the population of interest. Systematic error occurs when the sample is not drawn properly, as in the poll conducted by Literary Digest magazine. It can also occur if names are dropped from the sample list because some individuals were difficult to locate or uncooperative. Individuals dropped from the sample could be different from those retained. Those remaining could quite possibly produce a biased sample. Political polls often have special problems that make prediction difficult. Random sampling error, as contrasted to systematic sampling error, is often referred to as chance error. Purely by chance, samples drawn from the same population will rarely provide identical estimates of the population parameter of interest. These estimates will vary from sample to sample. For example, if you were to flip 100 unbiased coins, you would not be surprised if you obtained 55 heads on one trial, 49 on another, 52 on a third, and so on. Thus, some samples will, by chance, provide better estimates of the parameter than others. In any given sample, some attributes of interest may be over represented and some under represented. However, this type of error is random. Moreover, it is possible to describe this error statistically and take it into account when drawing inferences. Thus, in sampling voters prior to an election, we can make claims of the following sort: "There is a 95% probability that the proportion of voters who will cast their ballot for Candidate A will fall within the interval from 43% to 47%." Before selecting a sample, researchers usually decide the amount of error they are willing to tolerate and the level of confidence they want to have. This margin of error is expressed in terms of a confidence interval—

for example, 95% of the time the sample will correctly reflect the population values with a margin of error of plus or minus 4 percentage points.Sampling error can affect inferences based on sampling in two important situations. In one situation, we may wish to generalize from the sample to a particular population. With a small sampling error, we feel reasonably comfortable about generalizing from the sample to the population. Survey research is most concerned about this kind of sampling error. The second situation in which sampling error plays a role is when we wish to determine whether two or more samples were drawn from the same or different populations. In this case, we are asking if two or more samples Can feel more confident that our sample is representative of the population. Thus, we can are sufficiently different to rule out factors due to chance. An example of this situation is when we ask the question “Did the group that received the experimental treatment really differ from the group that did not receive the treatment other than on the basis of Chance.

Sampling Techniques

Simple Random Sampling

Researchers use two major sampling techniques: probability sampling and non probability sampling. With probability sampling, a researcher can specify the probability of an element’s (participant’s) being included in the sample. With non probability sampling, there is no way of estimating the probability of an element’s being included in a sample. If the researcher’s interest is in generalizing the findings derived from the sample to the general population, then probability sampling is far more useful and precise. Unfortunately, it is also much more difficult and expensive than non probability sampling. Probability sampling is also referred to as random sampling or representative sampling. The word random describes the procedure used to select elements (participants, cars, test items) from a population. When random sampling is used, each element in the population has an equal chance of being selected (simple random sampling) or a known probability of being selected (stratified random sampling). The sample is referred to as representative because the characteristics of a properly drawn sample represent the parent population in all ways. One caution before we begin our description of simple random sampling: Random sampling is different from random assignment. Random assignment describes the process of placing participants into different experimental groups.

Stratified Random Sampling

Where the population is heterogeneous in nature the procedure used is stratified random sampling which is also a form of probability sampling. To stratify means to classify or to separate people into groups according to some characteristics, such as position, rank, income, education, sex, or ethnic background. These separate groupings are referred to as subsets or subgroups. For a stratified random sample, the population is divided into groups or strata. A random sample is selected from each stratum based upon the percentage that each subgroup represents in the population. Stratified random samples are generally more accurate in representing the population than are simple random samples. They also require more effort, and there is a practical limit to the number of strata used. Because participants are to be chosen randomly from each stratum, a complete list of the population within each stratum must be constructed. Stratified sampling is generally used in two different ways. In one, primary interest is in the representativeness of the sample for purposes of commenting on the population. In the other, the focus of interest is comparison between and among the strata.

Convenience Sampling

Convenience sampling is used because it is quick, in expensive, and convenient. Convenience samples are useful for certain purposes, and they require very little planning. Researchers simply use participants who are available at the moment. The procedure is casual and easy, relative to random sampling. Contrast using any available participants with random sampling, where you must (1) have a well-defined population, (2) construct a list of members of the population if one is not available, (3) sample randomly from the list, and (4) contact and use as many individuals from the list as possible. Convenience sampling requires far less effort. However, such convenience comes with potential problems, which we will describe. Convenience samples are non probability samples. Therefore, it is not possible to specify the probability of any population element's being selected for the sample. Indeed, it is not possible to specify the population from which the sample was drawn.

Quota Sampling

In many large-scale applications of sampling procedures, it is not always possible or desirable to list all members of the population and randomly select elements from that list. The reasons for using any

alternative procedures include cost, timeliness, and convenience. One alternative procedure is quota sampling. This technique is often used by market researchers and those taking political polls. Usually, when this technique is used, the population of interest is large and there are no ready-made lists of names available from which to sample randomly. The Gallup Poll is one of the best known and well conducted polls to use quota sampling. This poll frequently reports on major public issues and on presidential elections. The results of the poll are syndicated for a fee that supports it. In this quota sampling procedure, localities are selected and interviewers are assigned a starting point, a specified direction, and a goal of trying to meet quotas for subsets (ethnic origins, political affiliations, and so on) selected from the population. Although some notable exceptions have occurred, predictions of national elections over the past few years have been relatively accurate—certainly, much more so than guesswork. With the quota sampling procedure, we first decide which subgroups of the population interest us. This, in turn, is dictated by the nature of the problem being investigated (the question being asked). For issues of national interest (such as abortion, drug use, or political preference), frequently used subsets are age, race, sex, socioeconomic level, and religion. The intent is to select a sample whose frequency distribution of characteristics reflects that of the population of interest. Obviously, it is necessary to know the percentage of in dividuals making up each subset of the population if we are to match these percentages in the sample. For example, if you were interested in ethnic groups such as Italians, Germans, Russians, and so on, and knew their population percentages, you would select your sample so as to obtain these percentages.

General Summary

When we conduct research, we are generally interested in drawing some conclusion about a population of individuals that have some common characteristic. However, populations are typically too large to allow observations on all individuals, and we resort to selecting a sample. In order to make inferences about the population, the sample must be representative. Thus, the manner in which the sample is drawn is critical. Probability sampling uses random sampling in which each element in the population (or a subgroup of the population with stratified random sampling) has an equal chance of being selected for the sample. This technique is considered to be the best means of obtaining a representative sample. When probability sampling is not possible, non probability sampling must be used. Convenience sampling involves using participants who are readily available' It is the easiest technique but the poorest from a methodological Stand point. Quota sampling is essentially convenience sampling in which there is an effort to better represent the population by sampling a certain percentage of participants from subgroups that correspond to the prevalence of those subgroups in the population. By their very nature, samples do not perfectly match the population from which they are drawn. There is always some degree of sampling error, and the degree of error is inversely related to the size of the sample. Larger samples are more likely to accurately represent characteristics of the population, and smaller samples are less likely to accurately represent characteristics of the population. Therefore, researchers strive for samples that are large enough to reduce sampling error to an acceptable level. Even when samples are large enough, it is important to evaluate the specific method by which the sample was drawn. We are increasingly exposed to information obtained from self-selected samples that represent only a very narrow subgroup of individuals. Much of such information is meaningless because the subgroup is difficult to identify.

****

References

1. www.stats.gla.ac.uk/steps/glossary/sampling.html

2. http://www.socialresearchmethods.net/kb/sampling.php

3. http://uca.edu/psychology/files/2013/08/Ch7-Sampling-Techniques.pdf

4. Salant, Priscilla, I. Dillman, and A. Don. How to conduct your own survey. No. 300.723 S3.. 1994.

5. Hansen, Morris H., William N. Hurwitz, and William G. Madow. "Sample Survey Methods and Theory." (1953).