International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Probability Sampling Techniques: A Literature Review

Anmol

Ravenshaw University, PG Department of Psychology, Cuttack, India

Abstract: ***This paper reviews the various sampling methods covered under probability sampling techniques. The paper begins with a formal analysis of the need for sampling procedures. A brief comparison between probability sampling and non-probability sampling techniques has also been made to review the potential advantages and disadvantages present in the given sampling methods. The main section of the paper deals with various forms of probability sampling techniques, which are categorized as random sampling method, stratified sampling, systematic sampling method, cluster sampling, and replicated sampling. Various subdivisions of each of these sampling methods have also been discussed. Moreover, the suitability of each of these methods is discussed. The paper concludes with the comparison of probability sampling methods concerning other available sampling procedures. Based on the present review, it is found that all conditions being equal and feasible; the samples obtained by probability sampling methods are better representatives of populations as compared to other sampling methods.***

Keywords: Sample, Sampling units, Probability Sampling, Non-probability Sampling, True representatives of a population

1. Introduction

Sampling procedures are the most crucial element of a study. Any errors arising during the selection of samples is directly going to affect the interpretation of results [1]. Therefore, researchers must strive to conduct sampling procedures with minimal error. A population is the collection of all the units of a particular study [2]. A sample is a part of the population [3]. A sampling unit is defined as the basic standard of measure of the sample such that the summation of all the sampling units results in a particular category of a sample [2]. Sampling frame consists of all the sampling units from which a sample is taken [4]. As per this logic, if we take a particular city like Los Angeles, then, the population will consist of all the residence of the Los Angeles city. The sampling frame in this situation needs to be defined as per the requirements of the research or the researcher. The most important criteria for the selection of a sample is that it must be a true representative of a population [5]. The requirements for being a genuine representative of a population can vary based on the nature as well as the objective of the research study. However, a set of variables like quality of the sample, sample size, sampling error, and cost of sampling are some of the few factors that are considered for deciding the representatives of a population [2].

The rationale of the study

The existing literature available on sampling procedures are centered more on various sampling techniques and therefore, involve both probability and non-probability sampling procedures. There is a need to differentiate between the probability and non-probability sampling procedures. Further, since probability sampling procedures are more commonly followed in research studies than non-probability sampling procedures, this paper was written with an attempt to review the existing practices related to probability sampling methods. This review will deal with the efficiency of the various probability sampling techniques. It is hypothesized that all conditions being equal, probability sampling methods are better than other sampling methods when it comes to selecting true representatives of a population.

2. Review of Literature

Before we look into the various probability sampling techniques, it is essential to understand the need for the sampling process. An apparent reason why the researcher cannot study the entire population is that of the size of the population and the enormous time that will be involved to analyze every element of the population. Therefore, a representative of the population is used to study the various characteristics of the population [2]. The selection of such representatives of a population, therefore, requires careful and precise techniques that can reduce as much error as possible [5]. Sampling techniques help us to choose a true representative of a population [6].

The criteria for the selection of a sampling technique is contingent upon the following variables [2] -

1) Aim of the study
2) The interest of the researcher
3) Prior information about the population
4) Feasibility of a sampling technique

Sampling techniques are broadly divided as probability sampling techniques and nonprobability sampling techniques [7] . The fundamental difference between the two procedure is based on the theory of probability. The probability sampling follows the theory of probability, whereas non-probability sampling is not necessarily based on the theory of probability [8] . Probability theory is the mathematical study of phenomena characterized by randomness or uncertainty [9]. In other words, chances of selection of each element in a population can be calculated/known in probability sampling techniques. However, the same cannot be said for non-probability sampling procedures [8]. Extending the logic of probability theory to sampling techniques, it can, therefore, be argued that probability sampling techniques are less

biased in their process of selection as compared to non-probability sampling techniques. It is worthwhile to note that the suitability of a particular method depends more on the nature and objective of a research study. Probability sampling is used when we want to generalize the results, and we want a sample to be nearly an accurate representation of the population [2]. The non-probability sampling techniques are used when the principles or the assumptions of probability techniques cannot be applied or are not feasible based on time and cost factors [2].

Probability sampling techniques

All the methods that are covered under the probability sampling techniques follow the theory of probability. The primary characteristics of the sampling methods used here are -

Inferential statistics can be used with probability sampling techniques [2]. The reason is that the sample selected is a better representation of the population and therefore allows inferential procedures to be operated upon. Another characteristic is that all the samples have chances of selection. However, in nonprobability sampling techniques, it is possible that all the samples do not get chances of selection [8].

Different types of probability sampling techniques includes random sampling method, stratified random sampling method, systematic sampling method, cluster sampling method, and replicated sampling method [2]. We will look at each technique in details in the next section.

a) Simple random sampling

In simple random sampling, all the elements of a population have a fair and equal chance of selection [10]. The selection of the sample is also independent of each other. Generally, the whole population is considered for sampling. Required units are taken via random sampling procedures, for example, lottery method or using a table of random numbers [2]. All elements have equal and independent chances of selection. Equal chances of selection denote that each sample has the same probability of selection as compared to any other sample in the population. For example, out of 300 students in a class, the probability of selection of any student without any form of biasedness will be 1/300. Another thing to note here is that the samples are independent of each other [11]. In other words, the selection of one element does not affect the selection of other elements in the population. The steps that are used in random sampling procedures include [2]-

The first step is the consideration of a population that needs to be analyzed. The second step is the preparation of a list of sampling frames in the population. Finally, the last step involves drawing the required samples by using random procedure methods. The random selection of an element can be carried out by using manual techniques or by the help of calculator or computers. Table of random number can also be used as a reference; however, it is not suitable for larger samples [2]. In such cases, specialized computer programs can be used to select random numbers.

The most apparent advantage of random sampling method is that it is straightforward to apply and easy to understand [11]. Although sampling error still exists here, the samples obtained from the random sampling method are considered to be robust and is truly representative of the population [10]. The random sampling procedure is so crucial that an experimental design which does not include random sampling procedures is not covered under true experimental designs [12]. A particular case when the random sampling method seems to fail is when the population has a lot of extreme values [2]. In such a case, sampling procedures will not give adequate representation in proportion to every element of the population.

b) Stratified random sampling

Strata refer to a group or class. In stratified random sampling, the population is initially divided into suitable groups, and then the appropriate sampling method is followed to select the required samples from each group [13]. The populations are subdivided into homogeneous groups, and from each group, a sample can be drawn randomly. Stratified random sampling is further divided into proportionate stratified random sampling and disproportionate stratified random sampling [13]. In proportionate stratified random sampling, the event of drawing a sample from a group is based on the proportion of the group share in the total population [2]. In other words, elements from each group will be assigned a different probability of selection based on their group's proportion with respect to the entire population. The main advantage here is that proper representations are provided to each group, and statistical efficiency is generally higher [13]. However, one particular disadvantage is that to select a sample from a group based on its proportion, prior knowledge of the composition of the entire population or the group is required [2]. In disproportionate stratified random sampling, no equal or fair representation is given to each group. Therefore, in such a case, the selected sample can either over-represent or under-represent a particular group [2]. In other words, the sample chosen may not be a true representative of the population. This method is mainly used when we want to give more representation to a particular target group.

c) Systematic sampling or Fixed interval method

In systematic sampling, all the elements are first arranged in ascending or descending order and then appropriate random sampling methods are followed [14]. Also, a fixed interval size say 'n,' is decided by dividing the population with the number of sample size that is required. Once this fixed interval 'n' is chosen, the researcher selects any random number between 1 to n. This random number is the first sample, and the subsequent samples are obtained by adding the number n to the selected random number and the samples after that [14]. Some researchers do not consider the systematic sampling method as a probability sampling procedure [2]. The reason is that we find both the qualities of randomness and non-probability traits under the systematic sampling method. Hence systematic random sampling is also addressed by pseudo-random sampling method. Although systematic random sampling is less time-taking, however, all the elements here do not have equal chances of selection as

compared to the random sampling method [2].

d) Cluster sampling

In cluster sampling, a population is divided into clusters. The clusters are selected randomly, and then from each selected cluster, the required samples are selected using an appropriate sampling technique [15]. Therefore cluster sampling can include single or multiple stages. In single-stage sampling, the clusters are selected randomly, and all the elements in the selected clusters are studied uniformly [14]. In two-stage sampling, the clusters are first chosen randomly, and then in the second stage, samples from each selected clusters are also drawn randomly. When two-stage sampling is extended to more than two stages, it is called multistage sampling [15].

The primary advantage of cluster sampling is that it is more economical and time-saving as compared to other random sampling procedures [2]. Also, the cluster sampling method is more convenient for the broader population. However, the disadvantage is that the chances of sampling error are higher and further due to the random selection of clusters, selected samples may not be adequate representatives of the population [15]. The cluster sampling consists of several subtypes that include area sampling method, probability proportionate size sampling method, multiphase sampling method, two phase sampling multistage sampling method and subsampling method [15]. Each of these methods listed is based on the principles of cluster sampling in general.

Area sampling

Area sampling is a particular case of cluster sampling where the sampling units are specific geographical areas like states, cities, towns, and so on [16]. Area sampling can simultaneously involve multistage sampling methods.

Probability proportionate size

Probability proportionate size is a form of cluster sampling where the clusters are assigned proportions with respect to their size that affects the chances of selection [15]. Here a list of clusters is drawn along with their sizes and then each cluster is assigned a proportion based on the appropriate statistical method. Once proportions are assigned to each cluster, the clusters are selected randomly based on a multistage sampling process [14]. The main advantage is that the cluster of various sizes gets proportional representation, and therefore, this method can be more precise than the general process of cluster sampling. However, this method cannot be used if cluster sizes are not known [2].

Multiphase sampling

In multiphase sampling, sampling is carried out in stages [17]. The main population is considered as the sampling unit of the first stage. The sampling units selected in the early stage then serve as the basis for the second stage sampling process and so on [2]. However, the sampling units remains the same at all levels. Multiphase sampling is an extension of the two-phase sampling method.

Two-phase sampling method

In two-phase sampling, first, a larger sample say 'L,' consisting of all the samples of interest is selected using an appropriate sampling method. Then from this pre-selected sample 'L,' the required number of samples are extracted using a particular and appropriate sampling method [2]. It is worthwhile to note that multiphase sampling is not similar to multistage sampling. In multistage sampling, sampling units differ at each stage of the sampling process [15].

Multistage Sampling

In multistage sampling, sampling units are first drawn from the main population. From the selected sampling units, another set of sampling units are drawn. This continues until the researcher arrives at his desired number of samples [2]. The main difference between multistage and multiphase sampling is that the sampling units in multistage sampling units remain the same at each stage of the sampling process, but the same is not true for multistage sampling methods [14]. Multistage sampling is suitable for a vast population or large geographical areas [17]. However, the process of multistage sampling is difficult to carry out as compared to simple processes of stratified random sampling. Subsampling is a subtype of the multistage sampling method.

Subsampling

In the multistage sampling process, sampling in the second and subsequent stages is called subsampling [2]. Cluster sampling reduces cost and time, but the chances of sampling error are high [15]. Subsampling method attempts for reducing the sampling error at each stage. Therefore, at each step of the multistage sampling, various techniques can be employed that usually produces lower sampling errors.

e) Replicated or interpenetrating sampling method

Instead of selecting one larger sample to say 'S,' we choose different like sub-samples like S1, S2, S3 and so on such that the summation of all these sub-samples is equal to the larger sample 'S' [15] . Each subsample is selected as per a particular sampling method. Each subsample is an independent sample. The main advantage of replicated sampling method is that it saves time as different samples can be worked upon simultaneously [14]. The main disadvantage is that once a particular example is divided into subsamples, the subsamples can lose the essential properties of the larger sample [2].

3. Conclusion

This paper was about the review of various probability sampling methods. The paper started with a comparison of probability sampling and non-probability sampling methods. We then reviewed the various procedures covered under probability sampling methods. The paper reviewed various probability sampling methods starting from simple random sampling to cluster random sampling followed by systematic sampling and replicating sampling method. Based on the review, it is concluded that samples obtained by using probability sampling methods are better representatives of the population [18]. Though probability sampling procedures are not free from errors and shortcomings, yet, they are the first choice of researchers while conducting a study. It is to be noted that we are not saying probability sampling methods are superior to all other methods. As already mentioned, the suitability of a sampling procedure depends

more upon the need and the objectives of the research or the researcher [2]. However, when conditions are suitable to carry out either the probability sampling procedures or non-probability sampling procedure, then the sample obtained by the former procedures are a better representation of the population as compared to the samples obtained by the latter method [19]. It is, therefore, concluded that the probability sampling procedures are better suited to select representatives of a population as compared to other sampling methods.

References

[1] D. B. Chelton, "Effects of sampling errors in statistical estimation," *Deep Sea Res. Part A, Oceanogr. Res. Pap.*, 1983.
[2] O. R. Krishnaswami and M. Ranganatham, *Methodology of Research in Social Sciences*, 2nd REVISE. Mumbai: Himalaya Publication House, 2011.
[3] "What Is a Sample? Review Your Psychology Concepts." [Online]. Available: https://www.verywellmind.com/what-is-a-sample-2795877. [Accessed: 28-Feb-2018].
[4] "Sampling Frame / Sample Frame Definition," Dec-2014. [Online]. Available: https://www.statisticshowto.datasciencecentral.com/sampling-frame/. [Accessed: 25-Jul-2019].
[5] S. S. Bollard and J. S. Browder, "Sampling," in *Practical Handbook of Spectroscopy*, 2017.
[6] I. Etikan, "Sampling and Sampling Methods," *Biometrics Biostat. Int. J.*, 2017.
[7] G. Sharma, "Pros and cons of different sampling techniques," *Int. J. Appl. Res.*, 2017.
[8] M. Doherty, "Probability versus non-probability sampling in sample surveys," *New Zeal. Stat. Rev.*, 1994.
[9] "Probability theory." [Online]. Available: https://www.sciencedaily.com/terms/probability_theory.htm. [Accessed: 25-Jul-2019].
[10] "Simple Random Sampling," in *Sampling of Populations: Methods and Applications: Fourth Edition*, 2011.
[11] M. L. Patten, M. Newhart, M. L. Patten, and M. Newhart, "Simple Random and Systematic Sampling," in *Understanding Research Methods*, 2018.
[12] A. K. Singh, *Tests, Measurement and Research Methods in Behavioural Sciences*, FIFTH. Patna: Bharati Bhawan (P & D), 2017.
[13] M. L. Patten, M. Newhart, M. L. Patten, and M. Newhart, "Stratified Random Sampling," in *Understanding Research Methods*, 2018.
[14] A. S. Acharya, A. Prakash, P. Saxena, and A. Nigam, "Sampling: why and how of it?," *Indian J. Med. Spec.*, 2013.
[15] P. Sedgwick, "Cluster sampling," *BMJ (Online)*. 2014.
[16] "Sampling methods," *Dev. Water Sci.*, 1985.
[17] S. Ahmed, "Methods in sample surveys: Cluster sampling," 2009.
[18] S. Surbhi, "Difference Between Probability and Non-Probability Sampling," *keydifferences.com*, 2016. .
[19] A. Deshmukh, "Sampling - Probability Vs Non-Probability." 2013.

Author Profile

Anmol completed his B.A. and M.A. degree in Psychology from PG Department of Psychology, Ravenshaw University. He has been awarded JRF by UGC and is currently a PhD scholar at Ravenshaw University. He was awarded the gold medal by the University at the end of his post-graduation course in 2018.