e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

SAMPLING THEORY

Binu George[*1]

[*1]Mahatma Gandhi University, Statistics, Christ College, Kattappana, Kerala, India.

ABSTRACT

Sampling Theory in statistics is as common and important as salt in the food. In this part of the paper, we will discuss a few details regarding the sampling theory. It is written to enhance students, knowledge and also convince them the importance and uses of sampling theory in modern world.

Keywords: Sampling, Population, Sample, Random.

I. INTRODUCTION

Often we interested in making conclusions about a large group. Instead of examining the entire group, we may examine only a small part of group. Our objective is to draw valid inferences about certain facts for the population based on sample; a process known as statistical inference. The process of obtaining samples is called sampling and theory concerning the sampling is called sampling theory. Sampling Theory is one of the statistical technique used for analysis. It provides the tools and techniques for data collection. Statistical sampling theory provides a powerful theoretical framework for generalizing from samples to corresponding populations. Two advantages of sampling are lower cost and faster data collection than measuring the total population.

II. SAMPLING THEORY

Statistical Survey

A statistical survey is any structured inquiry designed to obtain aggregated data, which may be qualitative or quantitative. The purpose of a survey is the collection of information to satisfy a definite need. The need to collect data arises in all walks of life. The data we need may be about

- The population (total, sex, age, migration, rate of growth, literacy, religion …)
- Labour (no. of employees, hrs. of work, wages, strikes, unemployment …)
- Agriculture (area under diff. crops, forests, agriculture income, manures, cultivation practices …)
- Industry (turn over, production, capital investment, water consumption, pollution …)
- Trade (wholesale/ retail prices, demand, profit/ loss …) etc.

A statistical survey is a sort of investigation carried out by an agency or individual to study the nature of the unknown characteristics of a population. We undertake a survey for a variety of purposes. However in most cases our interest may be concentrated on four important unknown values of the population under study such as the population total, the population mean, the population proportion, the population ratio. Proportion: whole and part (proportion of smokers, males, defectives, distinctions …) Ratio : part and part (sex ratio, import/ export ratio, birth/ death ratio …)

Population: Aggregate of all objects about which we want to collect information (houses in an area, students in a class, fishes in a lake, viewers of a specific T. V. programme, normal population with mean 50 and SD 5 …) Characteristic: Any aspect of the population about which we want to collect information (color, height, life length, yield, political affiliation, income, employment …)

There are two ways of collecting data, they are Census method and Sampling method.

Census Method

This method is also known as complete enumeration method. In census, we are collecting data from each unit. Census method is most representative method. This method is applicable when population is finite.

Merits of census method

- The results are more representative, accurate and reliable
- The results are free from sampling errors
- A census data may be used as a basis for various other surveys However, despite of these advantages, the census method is not popularly used in practice.
- Effort, money, time required for completing census is very large.

- There is no way of checking the error in the data except through a re- survey or sample checks.
- Census is practically impossible for a researcher or a small organization.
- If the population is infinite or the enumeration is destructive in nature, census cannot be used.

Sampling Method

Sampling method is a method of collecting data from a representative part of the population only. When universe is infinite or hypothetical, we can adopt sampling method.

Merits of sampling include

- Reduced cost, time and labour
- Greater scope (need only less number of trained investigators, less administrative cost, less number of equipment …)
- Greater accuracy
- If the population is hypothetical or infinite, only sampling is possible
- It is always possible to determine the extent of sampling error Demerits of sampling method are
- A proper choice of the sampling method is not made, the results may be misleading
- The chances of sampling errors are great in sampling
- When the population is small, we can't use sampling
- When the information is needed from each and every unit in the population (Voters list preparation, income-tax assessment, college admissions …), sampling cannot be used.
- Neither sampling nor census admit universal application.
- Census and sampling will produce identical conclusions when the population is perfectly homogenous.
- It is a curious fact that the results from a carefully planned, well executed sample survey are expected to be more accurate than those from a census survey.
- The aim of sampling theory is to make sampling more effective so that the answer to a particular question is given in a quick, valid, efficient and economical way.

Errors in Surveys

Two major types of errors can arise when a survey is conducted to make observations on a characteristic defined over the population. They are Sampling errors and non-sampling errors.

Sampling error refers to the error arising due to drawing inferences about the population on the basis of few observations taken from it. This error is inherent and unavoidable in any sample survey. It can be decreased by increasing the sample size. Sampling error is inversely proportional to the square root of the sample size. Sampling errors are absent in census surveys. Few reasons for sampling errors are, faulty selection of the sample (purposive or judgment sampling, use of inappropriate sampling scheme like srs for heterogeneous populations …), substitution (when difficulties arise, investigator may substitute a convenient member of the population), faulty identification of the sampling units (high in area surveys or crop surveys) …

Non-sampling errors are more serious and are due to mistakes made in the acquisition of data. This is present in both sample surveys and census surveys. It can occur at any stage of its planning, execution and analysis. Few reasons for non-sampling errors are – faulty planning or definitions(faulty objectives, faulty questionnaire, errors in measurements, lack of trained investigators …), errors due to non-response (not at homes, unable to answer, refuses to answer the questions …), response errors (respondent may misunderstand a question and may furnish false data, prestige bias, investigator bias …), errors in coverage(inclusion/exclusion of units which are to be excluded/included in a survey …), compiling errors(errors in coding, editing, tabulation …), publication errors (errors in printing, presentation …) …

THE LANGUAGE OF SAMPLING

Population

The set of all objects about which we want to collect information. The population may be living or non-living. Population can be finite and infinite. The finite population is also known as a countable population in which the population can be counted. In finite population we can accurately measure size of the units. For example, number of students in a college.

The infinite population is also known as an uncountable population in which the counting of units in the population is not possible. In other words, in infinite population there are infinite number of units. For example, The number of stars in the sky.

Sample

Any subset of population is called sample. It is a small group of members selected from a population to represent the population. A sampling unit is a single member of the sample. The number of observations taken from a population through which statistical inferences for the whole population are made is called sample size. The sample size may large or small. If the sample is more than 30 then it is known as large sample otherwise it is small sample

Estimation

A process in which we obtain the values of unknown parameters with the help of sample data.

Estimator

It is a rule, formula or function that tells how to calculate an estimate. An estimator is a method of estimating a population parameter. It is generally expressed as a function of sample variates.

Estimate

An estimate is the numerical value of the estimator. Generally, few units are selected from a population and then observations are taken on these selected units. The constant for a characteristic is calculated from these sample observations. The constant, so obtained, is known as an estimate and stands for population parameter.

CLASSIFICATION OF SAMPLING TECHNIQUES

Probability sampling – Scientific method of selecting samples from the population. In this procedure, each unit in the population has a definite pre assigned non zero probability of being selected into the sample.

Non probability sampling - Method of selecting samples in which the choice of selection of units into the sample depends entirely on the judgment of the sampler (investigator).

Probability sampling is based on the fact that every member of a population has a known and equal chance of being selected. This method is based on the theory of probability. Probability sampling is based on the fact that every member of population has known and equal chance of being selected. Non-probability sampling involves non-random selection based on convenience.

III. PROBABILITY SAMPLNG

There are four types of probability sampling. They are,

1. Simple Random Sampling (SRS)

Simple random sampling is a technique in which every member of a study population has an equal chance of being selected. In this selection of items completely depends on chance and randomness. This technique is also known as method of chance. It is a fundamental sampling method and can easily be a component of a more complex sampling method. The main attribute of this sampling method is that every sample has the same probability of being chosen.

In the procedure of selecting a simple random sample, initially define the population and select a suitable sampling frame. Each element is assigned a number from 1 to N, then generate n different random numbers between 1 to N. The numbers generated denote the elements that should be included in the sample. Lottery method and random number table are the two procedures available for selecting simple random sample.

Lottery Method

This is the most popular and simplest method. In this method all the items of the population are numbered on separate slips of same size, shape and colour. They are folded and stored in a container. Shuffle them thoroughly. Slips are then drawn one by one till the require number or units are selected into the sample.

Random Number Tables

As the lottery method cannot be used, when the population is large, the alternative method is that of using the table of random numbers. There are several standard tables of random numbers. They are, Tippet's table, Fisher and Yates' table, Kendell and Smith' table, etc

2. Stratified Random Sampling

This sampling method is appropriate when the population has mixed characteristics, and you want to ensure that every characteristic is proportionally represented in the sample. That is, when the population is heterogeneous in nature the stratified random sampling is used. Here, you divide the population into subgroups called stratum or strata, based on the relevant characteristics from each stratum, random sample is drawn. Strata should be mutually exclusive and collectively exhaustive so that every population element should be assigned to one and only one strata and no population elements should be omitted. Elements are selected from each stratum by a random sample procedure, usually srs. The pool them together to get the stratified sample.

The merits of stratified sampling are the following

- Stratified sample will be more representative of the population as it is sure that units from each of the sub population will be present in the sample. For example, in a study of educational institutions the population is the set of schools and colleges. If we take a simple random sample, it is possible that the sample may contain only primary schools. But if we divide the educational institutions into three strata viz, primary schools, secondary schools and colleges and take a stratified sample we are assured that the sample will contain primary schools and secondary schools and colleges.
- In many situations stratified sampling may be administratively more convenient. If we want to take a sample of housed in the state, we may take the districts as strata and take simple random sample of houses from each district. This evidently is administratively more convenient.
- Stratified sample provides estimates of strata means and their standard errors.
- The variability of the estimate obtained by stratified sampling is much less than of the corresponding simple random sampling.

3. Cluster Sampling

A sampling procedure assume the division of the population into a finite number of distinct and identify the units called sampling units. The simplest unit into which the population can be divided is called the ultimate stage unit or elementary unit or simply elements of the population and groups of elements are called clusters. Sometimes, it is not possible to have a list of all the elementary units belonging to the population so that selecting a srs is not feasible. The method of cluster or area sampling is available in such cases.

In cluster sampling the population is first divided into a number of non-overlapping clusters. A cluster is a collection of elementary units or ultimate stage units of study. When the sampling units are clusters the procedure is called cluster sampling. If the area containing the population under study is divided into non-overlapping smaller segments the procedure is sometime called area sampling.

It is a probability sampling technique that is commonly employed to study large populations that are geographically dispersed. In a cluster sample technique, multiple clusters of people from the target population are created. Clusters created must have a similar distribution of characteristics as the distribution of the population as a whole. In this method, first define target your target population. Next, divide your sample into clusters then randomly select cluster as your samples. Collect data from your chosen samples.

Example: Population - Students in a College

Clusters - Classes/Divisions

Elements - Students

The main problem we face in cluster sampling is how to divide the population into appropriate clusters or how to specify the clusters. Cluster sampling involves less cost of travelling than srs with same number of elementary units. On the other hand, clusters are generally made up of neighbouring elements and therefore the elements within a cluster are on an average homogeneous in nature. So a cluster sampling is often less precise than a comparable srs.

4. Systematic Sampling

According to this method, units of the population are numerically, geographically and alphabetically arranged. Every n^{th} item of the numbered item is selected as a sample item. This method is also known as Quasi-random sampling. In this sampling scheme first unit is selected with the help of random numbers and remaining units are automatically selected according to a predetermined pattern.

Suppose N units in population are numbered 1 to N in some order. Suppose N = nk. Select a random number between 1 to k. Suppose it is r. Then the elements with the numbers r, r+k, r+2k, r+3k, r+4k ,..., r+(n-1)k will

constitute the systematic random sample. The merits of this method are, this is a simple method and sample is easily determined. In this method there is no personal bias. It is operationally more convenient than SRS. In this method much less training is needed for surveys to collect units.

Certain points we have to remember before using systematic sampling scheme. First, efficiency of systematic sampling depends on the order of arrangement of the units in the population. Second, If the units in the population show an increasing or decreasing trend along with the systematic sample means will also show the same tendency like rank lists, salary lists etc. Third, if the population is almost periodic or cyclic in nature, then the efficiency of systematic sampling depends on the value of k, the sampling interval.

The sample mean is taken as the estimate of the population mean. If N is a multiple of n, it is an unbiased estimate. If not, it will be approximately unbiased. The following are some advantages of systematic sampling.

- It is much easier and quicker to take a systematic sample and even a layman can draw a sample.
- Large continuous parts of the population will not fail to be represented.
- It will give precise estimates than simple random sampling.

5. Multistage Sampling

In this method sampling is done in stages. Sampling at each stage starts from the larger units, intermediate units, finally reaches the ultimate units of selection. To enable proper representation at each stage in the whole sampling process, eg: in a selection of a given number of households, the selection may be first made of states and then of districts, villages etc all on a random basis before the selection of the households, which are the ultimate units made.

Sampling also, in this type of sampling design, is carried out through stages. Firstly, only a member of first stage unit is selected. For each of the selected first stage sampling units, a number of second stage sampling units is selected. The process is carried out until we select the ultimate sampling units. As an example of multi stage sampling, in order to find the extent of unemployment in India, we may take state, district, Police station and household as the first stage, second stage, third stage and ultimate sampling units respectively. It also saves computational labour and is cost effective. It adds flexibility into the sampling process which is lacking in other sampling schemes. However, compared to stratified sampling, multistage sampling is likely to be less accurate.

6. Probability Proportional to Size Sampling (PPS Sampling)

In srs scheme, the selection probabilities were equal for all the units in population. But when the units in the population vary in their size, obviously srs is not an appropriate procedure for selecting samples because no importance is given to the size of the units. In survey of industries the number of workers may be taken as the size. In agricultural survey, area under crop may be taken as the size etc. In such cases auxiliary or supplementary information about the size of the units can be utilized to select the sample. So that they can suggest more efficient estimators of the population parameters. One such method is to assign unequal probabilities of selection to different units in the population depending on their sizes.

When the units in a population vary in their size and the study variate highly correlate with the size of the units. Probability of selection may be done in proportion to in size of the units. The sampling procedure where the probability of selection is proportional to the size of the units in the population is known as probability proportional to size sampling.

There is a basic difference between srs and PPS sampling. In srs, the probability of selecting any specified unit at any specified draw is the same. But in PPS sampling it differs from draw to draw. The theory of PPS sampling is consequently more complex than that of srs.

IV. NON-PROBABILITY SAMPLNG

The important non-probability sampling is

1. Purposive Sampling

Any type of sampling in which the sample selected depends on personal discretion or judgement of the investigator is called a subjective or judgement sampling. This type of sampling is used with a definite purpose in view and as such is not used for general purpose. In this sampling method, the choice of sample items depends exclusively on the judgement of investigator. For example, an investigator may conduct a survey for getting opinions and views on certain issues like family welfare programmes, everyday business problems,

liking and disliking of T. V. programmes etc. by selecting a sample of person by his own judgement. Thus, the success of this sampling method depends on the proper judgement of the investigator in choosing a representative sample, otherwise the application of this method would bring in personal bias. This is the biggest limitation of this sampling method.

2. Quota Sampling

In this method each person engaged in the primary collection of data is assigned a quota of investigation. Although certain criteria are prescribed for the selection of respondents, the actual choice of all such respondents or part of the whole quota, who are for any reason not approachable is left to the investigator who is permitted to substitute other to fulfil the quota assigned to him. The method is often adopted in marketing research studies where it is not possible to sick to it with out delay and expenditure. This method also cause bias.

3. Convenience sampling

It is one of the non-probability sampling method. Convenience sampling is not based on any rule or procedure. This method is entirely based on convenience. Convenience sampling is one in which a sample is obtained by selecting such units of the population which may be conveniently located and contacted. In this method, the samples are biased and may not much represent the population.

4. Snowball sampling

It is a non-probability method where existing subjects recruit future subjects among their acquaintances. Snowball sampling used when it is to reach the relevant sample. In this method, the selection of individuals by using an initial set of members as informants. The selected cases are dependent on the choice of the informants. Thus, samples are biased and unrepresentative. For example, reaching a particular group in a society, persons with specific preferences or traits (sex, religion, ideology etc) in an organization.

5. Haphazard sampling

Haphazard sampling is one in which the investigator selects samples without following a structured technique, but avoiding any conscious bias or predictability. The results obtained cannot be replied upon to arrive at the conclusion about the population.

V. CONCLUSION

Sampling theory is one of the statistical techniques which helps us to drawing inferences about the population based on the samples taken from the population. Using this technique in research saves mainly on money and time. Many sampling methods are available. We must choose appropriate sampling method for our study.

VI. REFERENCES

[1] Hansen M. H., Hurwitz and W. G. Madow. (1953), “Sample survey methods and theory”, John Wiley & Sons, Vol. 1, N. Y. 638 p.

[2] Gupta S. C., Fundamentals of statistics, Himalaya Publishing House.

[3] Robert M. Groves (2009), “Survey Methodology”

[4] Lohr Sharon L, “Sampling: Design and Analysis”

[5] L. R. Potti, “Quantitative Techniques”, Yamuna Publications.

[6] Parimal Mukhopadhyay, “Theory of sample surveys with applications”, Prentice Hall of India.