

SAMPLING METHODS

The theory of sampling has been developed recently but this is not new. In our everyday life we have been using sampling theory as we have discussed in introduction. In all those cases we believe that the samples give a correct idea about the population. Most of our decisions are based on the examination of a few items that is sample studies

Sample:

Statisticians use the word **sample** to describe a portion chosen from the population. A finite subset of statistical individuals defined in a population is called a sample. The number of units in a sample is called the **sample size**.

Sampling unit:

The constituents of a population which are individuals to be sampled from the population and cannot be further subdivided for the purpose of the sampling at a time are called sampling units. For example, to know the average income per family, the head of the family is a sampling unit. To know the average yield of rice, each farm owner's yield of rice is a sampling unit.

Sampling frame:

For adopting any sampling procedure it is essential to have a list identifying each sampling unit by a number. Such a list or map is called sampling frame. A list of voters, a list of house holders, a list of villages in a district, a list of farmers etc. are a few examples of sampling frame.

Reasons for selecting a sample:

Sampling is inevitable in the following situations:

1. Complete enumerations are practically impossible when the population is infinite.
2. When the results are required in a short time.
3. When the area of survey is wide.
4. When resources for survey are limited particularly in respect of money and trained persons.
5. When the item or unit is destroyed under investigation.

Parameters and statistics:

We can describe samples and populations by using measures such as the mean, median, mode and standard deviation. When these terms describe the characteristics of a population, they are called **parameters**. When they describe the characteristics of a sample, they are called **statistics**. A parameter is a characteristic of a population and a statistic is a characteristic of a sample. Since samples are subsets of population statistics provide estimates of the parameters. That is, when the parameters are unknown, they are estimated from the values of the statistics.

Principles of Sampling:

Samples have to provide good estimates. The following principle tell us that the sample methods provide such good estimates

1. **Principle of statistical regularity:** A moderately large number of units chosen at random from a large group are almost sure on the average to possess the characteristics of the large group.
2. **Principle of Inertia of large numbers:** Other things being equal, as the sample size increases, the results tend to be more accurate and reliable.
3. **Principle of Validity:** This states that the sampling methods provide valid estimates about the population units (parameters).
4. **Principle of Optimisation:** This principle takes into account the desirability of obtaining a sampling design which gives optimum results. This minimizes the risk or loss of the sampling design. The foremost purpose of sampling is to gather maximum information about the population under consideration at minimum cost, time and human power. This is best achieved when the sample contains all the properties of the population.

Sampling errors and non-sampling errors:

The two types of errors in a sample survey are sampling errors and non - sampling errors.

Sampling errors: Although a sample is a part of population, it cannot be expected generally to supply full information about population. So there may be in most cases difference between statistics and parameters. The discrepancy between a parameter and its estimate due to sampling process is known as **sampling error**.

Non-sampling errors: In all surveys some errors may occur during collection of actual information. These errors are called Non-sampling errors.

Advantages and Limitation of Sampling:

There are many advantages of sampling methods over census method. They are as follows:

1. Sampling saves time and labour.
2. It results in reduction of cost in terms of money and man hour.
3. Sampling ends up with greater accuracy of results.
4. It has greater scope.
5. It has greater adaptability.
6. If the population is too large, or hypothetical or destroyable sampling is the only method to be used.

The limitations of sampling are given below:

1. Sampling is to be done by qualified and experienced persons. Otherwise, the information will be unbelievable.
2. Sample method may give the extreme values sometimes instead of the mixed values.
3. There is the possibility of sampling errors. Census survey is free from sampling error.

Types of Sampling:

The technique of selecting a sample is of fundamental importance in sampling theory and it depends upon the nature of investigation. The sampling procedures which are commonly used may be classified as

1. Probability sampling.
2. Non-probability sampling.
3. Mixed sampling.

Probability sampling (Random sampling): A probability sample is one where the selection of units from the population is made according to known probabilities

Non-Probability sampling: It is the one where discretion is used to select ‘representative’ units from the population (or) to infer that a sample is ‘representative’ of the population. This method is called **purposive** sampling. This method is mainly used for opinion surveys; A common type of purposive sample used in surveys is quota sample. This method is not used in general because of prejudice and bias of the enumerator. However if the enumerator is experienced and expert, this method may yield valuable results.

Mixed Sampling: Here samples are selected partly according to some probability and partly according to a fixed sampling rule; they are termed as mixed samples and the technique of selecting such samples is known as **mixed sampling**.

Methods of selection of samples:

Here we shall consider the following three methods:

1. Simple random sampling.
2. Stratified random sampling.
3. Systematic random sampling.

Simple random sampling:

A simple random sample from finite population is a sample selected such that each possible sample combination has equal probability of being chosen. It is also called unrestricted random sampling.

Simple random sampling without replacement: In this method the population elements can enter the sample only once (ie) the units once selected is not returned to the population before the next draw.

Simple random sampling with replacement: In this method the population units may enter the sample more than once. Simple random sampling may be with or without replacement.

Methods of selection of a simple random sampling:

The following are some methods of selection of a simple random sampling.

a) 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 paper of same size, shape and colour. They are folded and mixed up in a container. The required numbers of slips are selected at random for the desired sample size. For example, if we want to select 5 students, out of 50 students, then we must write their names or their roll numbers of all the 50 students on slips and mix them. Then we make a random selection of 5 students. This method is mostly used in lottery draws. If the universe is infinite this method is inapplicable.

b) Table of Random numbers:

As the lottery method cannot be used, when the population is infinite, the alternative method is that of using the table of random numbers. There are several standard tables of random numbers.

1. Tippett's table
2. Fisher and Yates' table
3. Kendall and Smith's table are the three tables among them.

A random number table is so constructed that all digits 0 to 9 appear independent of each other with equal frequency. If we have to select a sample from population of size $N = 100$, then the numbers can be combined three by three to give the numbers from 001 to 100. Units of the population from which a sample is required are assigned with equal number of digits. When the size of the population is less than thousand, three digit number 000, 001, 002, 999 are assigned. We may start at any place and may go on in any direction such as column wise or row-wise in a random number table. But consecutive numbers are to be used. On the basis of the size of the population and the random number table available with us, we proceed according to our convenience. If any random number is greater than the population size N , then N can be subtracted from the random number drawn. This can be repeatedly until the number is less than N or equal to N .

Merits:

1. Personal bias is eliminated as a selection depends solely on chance .
2. A random sample is in general a representative sample for a homogenous population.
3. There is no need for the thorough knowledge of the units of the population.
4. The accuracy of a sample can be tested by examining another sample from the same universe when the universe is unknown.
5. This method is also used in other methods of sampling.

Limitations:

1. Preparing lots or using random number tables is tedious when the population is large.
2. When there is large difference between the units of population, the simple random sampling may not be a representative sample.
3. The size of the sample required under this method is more than that required by stratified random sampling.
4. It is generally seen that the units of a simple random sample lie apart geographically. The cost and time of collection of data are more.

Stratified Random Sampling:

Of all the methods of sampling the procedure commonly used in surveys is stratified sampling. This technique is mainly used to reduce the population heterogeneity and to increase the efficiency of the estimates. Stratification means division into groups. In this method the population is divided into a number of subgroups or strata. The strata should be so formed that each stratum is homogeneous as far as possible. Then from each stratum a simple random sample may be selected and these are combined together to form the required sample from the population.

What is Stratification?

The process of dividing a population of elements into distinct subpopulations called strata. Strata are formed so that each population element is assigned to only one stratum.

Types of Stratified Sampling:

There are two types of stratified sampling. They are proportional allocation and optimum allocation.

The population is divided into strata so that each population element is a member of only one stratum. We use the letter H to represent the number of strata that are formed and N_h to denote the number of population elements which fall in the h -th stratum. The total number of elements in the population is therefore

$$N = N_1 + N_2 + \dots + N_H = \sum_{h=1}^H N_h$$

A sample size is chosen for each stratum. We denote the sample size in the h -th stratum by the symbol, n_h . The total sample size over all strata is then

$$n = n_1 + n_2 + \dots + n_H = \sum_{h=1}^H n_h$$

In the proportional sampling equal and proportionate representation is given to subgroups or strata. If the number of items is large, the sample will have a higher size and vice versa. That is in this case we select the sample satisfying the condition

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_h}{N_h} = \frac{n}{N}$$

Or

$$n_i = \frac{n}{N} N_i, i = 1, 2, \dots, h.$$

In optimum allocation we select samples by considering the cost of obtaining information from each strata's.

Merits and limitations of stratified sampling:

Merits:

1. It is more representative.
2. It ensures greater accuracy
3. It is easy to administer as the universe is sub - divided.
4. Greater geographical concentration reduces time and expenses.
5. When the original population is badly skewed, this method is appropriate.
6. For non – homogeneous population, it may field good results.

Limitations:

1. To divide the population into homogeneous strata, it requires more money, time and statistical experience which is a difficult one.
2. Improper stratification leads to bias, if the different strata
3. overlap such a sample will not be a representative one.

Systematic Sampling:

This method is widely employed because of its ease and convenience. A frequently used method of sampling when a complete list of the population is available is **systematic sampling**. It is also called **Quasi-random sampling**.

Selection procedure:

The whole sample selection is based on just a random start. The first unit is selected with the help of random numbers and the rest get selected automatically according to some pre designed pattern is known as **systematic sampling**. With systematic random sampling every K^{th} element in the frame is selected for the sample, with the starting point among the first K elements determined at random.

For example, if we want to select a sample of 50 students from 500 students under this method K^{th} item is picked up from the sampling frame and K is called the **sampling interval**.

$$K = \left[\frac{N}{n} \right] = \left[\frac{\text{Population size}}{\text{Sample Size}} \right] = \left[\frac{500}{50} \right] = 10$$

Now select a number between 1 and 10 this number is called the random start. Let it be 4 then our sample consist of the units numbered 4, 14, 24, 34, That is in general if r is a random start then our sample consist of r^{th} , $(r+k)^{\text{th}}$, ... $(r+(n-1)k)^{\text{th}}$ units.

The technique will generate K systematic samples with equal probability.

Merits :

1. This method is simple and convenient.
2. Time and work is reduced much.
3. If proper care is taken result will be accurate.
4. It can be used in infinite population.

Limitations:

1. Systematic sampling may not represent the whole population.
2. There is a chance of personal bias of the investigators.

Systematic sampling is preferably used when the information is to be collected from trees in a forest, house in blocks, entries in a register which are in a serial order etc.