Significance and methods of Sampling

When carrying out a survey, it would be impractical to study a whole population. Sampling is a method that allows researchers to infer information about a population based on results from a subset of the population. It is important to ensure that the individuals selected are representative of the whole population.

There are several different sampling techniques available that can be grouped into two categories as probability sampling, and non-probability sampling. The difference between the two techniques is whether the sample is selected based on randomization or not.

In probability sampling, alternatively knows as random sampling, you start with a complete sample frame of all eligible individuals that have an equal chance to be part of the selected sample. The selection must occur in a 'random' way, meaning that they do not differ in any significant way from observations not sampled. It is typically assumed that statistical tests contain data that has been obtained through random sampling. For example, exit polls from voters that aim to predict the likely results of an election.

- **Types of random sampling:**

There are 4 types of random sampling techniques:

1. Simple random sampling:

Simple random sampling is the most straightforward approach to getting a random sample. It involves picking the desired sample size and selecting observations from a population in such a way that each observation has an equal chance of selection until the desired sample size is achieved. For example, a random selection of 20 students from a class of 50 students gives a probability of selection being 1/50.

2. Stratified random sampling:

This technique divides the elements of the population into key subgroups or strata. The elements are randomly selected from each of these strata. For example, males under 30, females under 30, males 30 or over, and females 30 or over. Say you want to achieve a sample size of 200, then you can pick samples of 50 from each stratum. The required sample size for each stratum will be designed either to match the known population proportions or to over-represent key subgroups of interest. We need to have prior information about the population to create subgroups. The main benefit of stratified sampling over simple random sampling is making sure that you have good sample sizes in key subgroups

3. Cluster sampling:

Cluster sampling starts by dividing a population into groups, or **clusters.** What makes this different that stratified sampling is that each cluster must be representative of the population. Then, you randomly selecting entire clusters to sample. For example, if an elementary school had five different grade eight classes, cluster random sampling might be used and only one class would be chosen as a sample, for example.

4. Systematic Random Sampling:

Systematic random sampling is a very common technique in which you sample every k'th element. For example, if you were conducting surveys at a mall, you might survey every 100th person that walks in, for example. If you have a sampling frame then you would divide the size of the frame, N, by the desired sample size, n, to get the index number, k. You would then choose every k'th element in the frame to create your sample.

- **Significance of random sampling:**

Probability sampling is also known as 'random sampling' or 'chance sampling'. Under this sampling design, every item of the universe has an equal chance of inclusion in the sample. It is, so to say, a lottery method in which individual units are picked up from the whole group not deliberately but by some mechanical process. Here it is blind chance alone that determines whether one item or the other is selected. The results obtained from probability or random sampling can be assured in terms of probability i.e., we can measure the errors of estimation or the significance of results obtained from a random sample, and this fact brings out the superiority of random sampling design over the deliberate sampling design. Random sampling ensures the law of Statistical Regularity which states that if on an average the sample chosen is a random one, the sample will have the same composition and characteristics as the universe. This is the reason why random sampling is considered as the best technique of selecting a representative sample.

Random sampling from a finite population refers to that method of sample selection which gives each possible sample combination an equal probability of being picked up and each item in the entire population to have an equal chance of being included in the sample. This applies to sampling without replacement i.e., once an item is selected for the sample, it cannot appear in the sample again (Sampling with replacement is used less frequently in which procedure the element selected for the sample is returned to the population before the next element is selected. In such a situation the same element could appear twice in the same sample before the second element is chosen). In brief, the implications of random sampling (or simple random sampling) are:

(a) It gives each element in the population an equal probability of getting into the sample; and all choices are independent of one another.

(b) It gives each possible sample combination an equal probability of being chosen.

Keeping this in view we can define a simple random sample (or simply a random sample) from a finite population as a sample which is chosen in such a way that each of the *NCn* possible samples has the same probability, 1/*NCn*, of being selected. To make it more clear we take a certain finite population consisting of six elements (say *a*, *b*, *c*, *d*, *e*, *f*) i.e., *N* = 6. Suppose that we want to take a sample of size *n* = 3 from it. Then there are 6*C*3 = 20 possible distinct samples of the required size, and they consist of the elements *abc*, *abd*, *abe*, *abf*, *acd*, *ace*, *acf*, *ade*, *adf*, *aef*, *bcd*, *bce*, *bcf*, *bde*, *bdf*, *bef*, *cde*, *cdf*, *cef*, and *def*. If we choose one of these samples in such a way that each has the probability 1/20 of being chosen, we will then call this a random sample.

- **Sampling error:**

Sampling errors are the random variations in the sample estimates around the true population parameters. Since they occur randomly and are equally likely to be in either direction, their nature happens to be of compensatory type and the expected value of such errors happens to be equal to zero. Sampling error decreases with the increase in the size of the sample, and it happens to be of a smaller magnitude in case of homogeneous population.

Sampling error can be measured for a given sample design and size. The measurement of sampling error is usually called the 'precision of the sampling plan'. If we increase the sample size, the precision can be improved. But increasing the size of the sample has its own limitations viz., a large sized sample increases the cost of collecting data and also enhances the systematic bias. Thus, the effective way to increase precision is usually to select a better sampling design which has a smaller sampling error for a given sample size at a given cost. In practice, however, people prefer a less precise design because it is easier to adopt the same and also because of the fact that systematic bias can be controlled in a better way in such a design.