

SAMPLING AND TYPES OF SAMPLING TECHNIQUES

Sampling refers to a statistical analysis in which a small part of the population is considered to perform a data analysis on them. Also, sampling is important as it should hold all the features of the population, more like the samples which is a small portion of the population should represent the whole population.

The sampling techniques are:

1. Simple Random Sampling
2. Systematic Sampling
3. Stratified Random Sampling
4. Cluster Sampling

These are the probability sampling techniques. There do exist some non-probability sampling techniques as well!

In probability sampling techniques, all the members of the population have an equal probability of being selected as a sample.

In **simple random sampling**, the sample is randomly picked from a larger population. Hence all the individual datapoints have an equal probability to be selected as sample data. Ex: employees survey in a company. No sample bias and the sample would be random without being partial (balanced sample). No extra domain knowledge is needed. Cons: population size should be high. It can't represent the population well sometimes.

In **systematic sampling**, the samples are picked from the population at regular intervals. This type of sampling is carried out if the population is homogenous and the data points are uniformly distributed. Ex: every 10th member from a population of 10,000. Pros: quick, less bias, even distribution of data. Cons: data manipulation risk, requires randomness in data, population shouldn't have patterns.

In **stratified random sampling**, the population is sub-divided into smaller groups called Strata. Samples are then obtained randomly from these strata. Pros: finds important characteristics in the population, High precision can be obtained if the differences in the strata are very high. Cons: Can't be performed on populations which can't be divided into groups, overlapping of datapoints.

Cluster sampling is carried out on populations that have inherent groups. This population is subdivided into clusters and then random clusters are taken as samples. Pros: requires only fewer resources, reduced variability. Cons: Can't be performed on populations without natural groups, overlapping datapoints, Can't provide a general insight for the entire population.