

Sampling

Sampling is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. The methodology used to sample from a larger population depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling.

Sample

A sample is a smaller, manageable version of a larger group. It is a subset containing the characteristics of a larger population. Samples are used in statistical testing when population sizes are too large for the test to include all possible members or observations. A sample should represent the whole population and not reflect bias toward a specific attribute.

Concept of Population and sampling

First, you need to understand the difference between a population and a sample, and identify the target population of your research.

- The **population** is the entire group that you want to draw conclusions about.
- The **sample** is the specific group of individuals that you will collect data from.

The population can be defined in terms of geographical location, age, income, and many other characteristics.

It can be very broad or quite narrow: maybe you want to make inferences about the whole adult population of your country; maybe your research focuses on customers of a certain company, patients with a specific health condition, or students in a single school.

It is important to carefully define your target population according to the purpose and practicalities of your project.

If the population is very large, demographically mixed, and geographically dispersed, it might be difficult to gain access to a representative sample.

Sampling frame

The sampling frame is the actual list of individuals that the sample will be drawn from. Ideally, it should include the entire target population (and nobody who is not part of that population).

Sample size

The number of individuals in your sample depends on the size of the population, and on how precisely you want the results to represent the population as a whole.

You can use a sample size calculator to determine how big your sample should be. In general, the larger the sample size, the more accurately and confidently you can make inferences about the whole population.

Types of Sampling

Probability sampling

Probability sampling means that every member of the population has a chance of being selected. It is mainly used in quantitative research. If you want to produce results that are representative of the whole population, you need to use a probability sampling technique.

Simple random sampling

In a simple random sample, every member of the population has an equal chance of being selected. Your sampling frame should include the whole population.

To conduct this type of sampling, you can use tools like random number generators or other techniques that are based entirely on chance.

Systematic sampling

Systematic sampling is similar to simple random sampling, but it is usually slightly easier to conduct. Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.

If you use this technique, it is important to make sure that there is no hidden pattern in the list that might skew the sample. For example, if the HR database groups employees by team, and

team members are listed in order of seniority, there is a risk that your interval might skip over people in junior roles, resulting in a sample that is skewed towards senior employees.

Stratified sampling

Stratified sampling involves dividing the population into subpopulations that may differ in important ways. It allows you draw more precise conclusions by ensuring that every subgroup is properly represented in the sample.

To use this sampling method, you divide the population into subgroups (called strata) based on the relevant characteristic (e.g. gender, age range, income bracket, job role).

Based on the overall proportions of the population, you calculate how many people should be sampled from each subgroup. Then you use random or systematic sampling to select a sample from each subgroup.

Cluster sampling

Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, you randomly select entire subgroups.

If it is practically possible, you might include every individual from each sampled cluster. If the clusters themselves are large, you can also sample individuals from within each cluster using one of the techniques above.

This method is good for dealing with large and dispersed populations, but there is more risk of error in the sample, as there could be substantial differences between clusters. It's difficult to guarantee that the sampled clusters are really representative of the whole population.

Sampling Error

A sampling error is a statistical error that occurs when an analyst does not select a sample that represents the entire population of data and the results found in the sample do not represent the results that would be obtained from the entire population

Determination of Sample Size

Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample. In practice, the sample size used in a study is usually determined based on the cost, time, or convenience of collecting the data, and the need for it to offer sufficient statistical power. In complicated studies there may be several different sample sizes: for example, in a stratified survey there would be different sizes for each stratum. In a census, data is sought for an entire population, hence the intended sample size is equal to the population. In experimental design, where a study may be divided into different treatment groups, there may be different sample sizes for each group.