**SAMPLING METHODS**

**Population**

Population refers to the entire group of individuals, items, data points, or elements that are the subject of a study or analysis. In other words, population is the full, comprehensive group under study. For example, all people living in India indicates the population of India.

**Sample**

Sample is a subset of a larger population that is selected for the purpose of conducting research, data analysis, or making inferences about the entire population. It includes one or more observations that are drawn from the population. For example, some people living in India is the sample of the population.

**Sampling**

Sampling refers to the process of selecting a subset of elements from a larger population to make inferences or draw conclusions about the entire population. A representative sample is chosen from the population, and the findings from the sample are used to make generalizations about the population.

**Sampling methods**

Sampling methods are techniques used in statistics and research to select a subset of individuals or items from a larger population. There are two primary types of sampling methods they are:

* Probability sampling
* Non-probability sampling

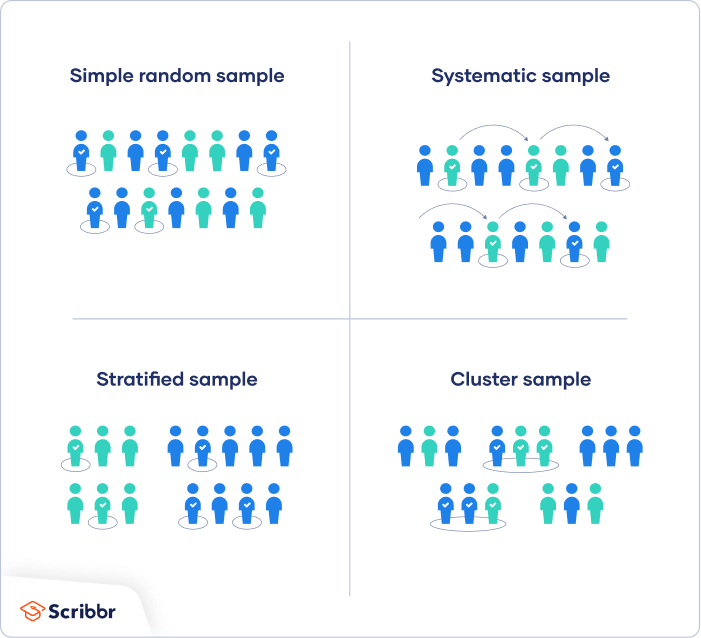
**Probability sampling**

Probability sampling is a sampling technique that involves randomly selecting a small group of people (a sample) from a larger population. Probability sampling means that every member of the population has a chance of being selected.

**Types of probability sampling**

The different types of probability sampling are: -

* Simple random sampling
* Systematic sampling
* Stratified sampling
* Cluster sampling
* Multi stage sampling

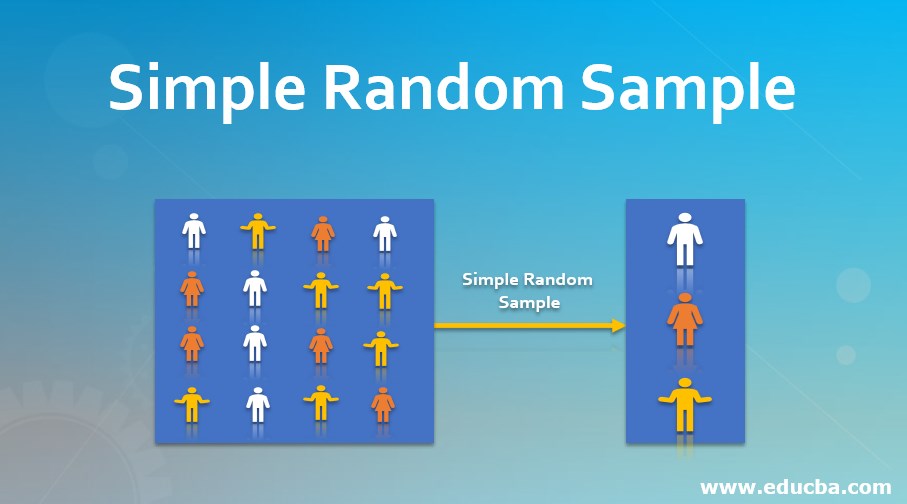


Video link: [Types of sampling](https://youtu.be/be9e-Q-jC-0?si=MyA3WL3A08SpIAol)

**Simple random sampling**

Simple random sampling is a type of probability sampling in which the researcher randomly selects a subset of participants from a population. Each member of the population has an equal chance of being selected.

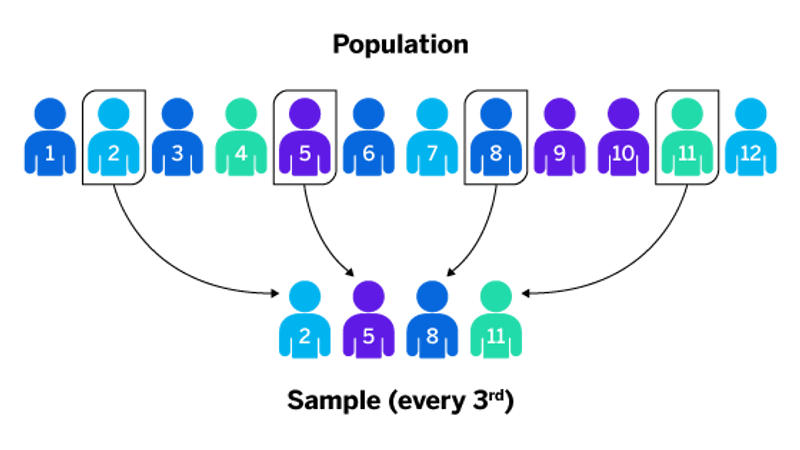
For eg, in our context if we want to analyze the price of instant food in amazon e- commerce site, the population size will be so big that the process will be time consuming and economical. Therefore, we fix a sample size and select a sample randomly from the population. This method is simple random sampling.



**Systematic sampling**

Systematic sampling is a probability sampling method where the researcher chooses elements from a target population by selecting a random starting point and selecting sample members after a fixed ‘sampling interval.’

For eg, if we have 1000 data of food products and we need a sample of 20 then we will start at a random point, assume 1 observation and then we define an interval 50 and we take every 50th observation into the sample. So that we will get a sample of 20.

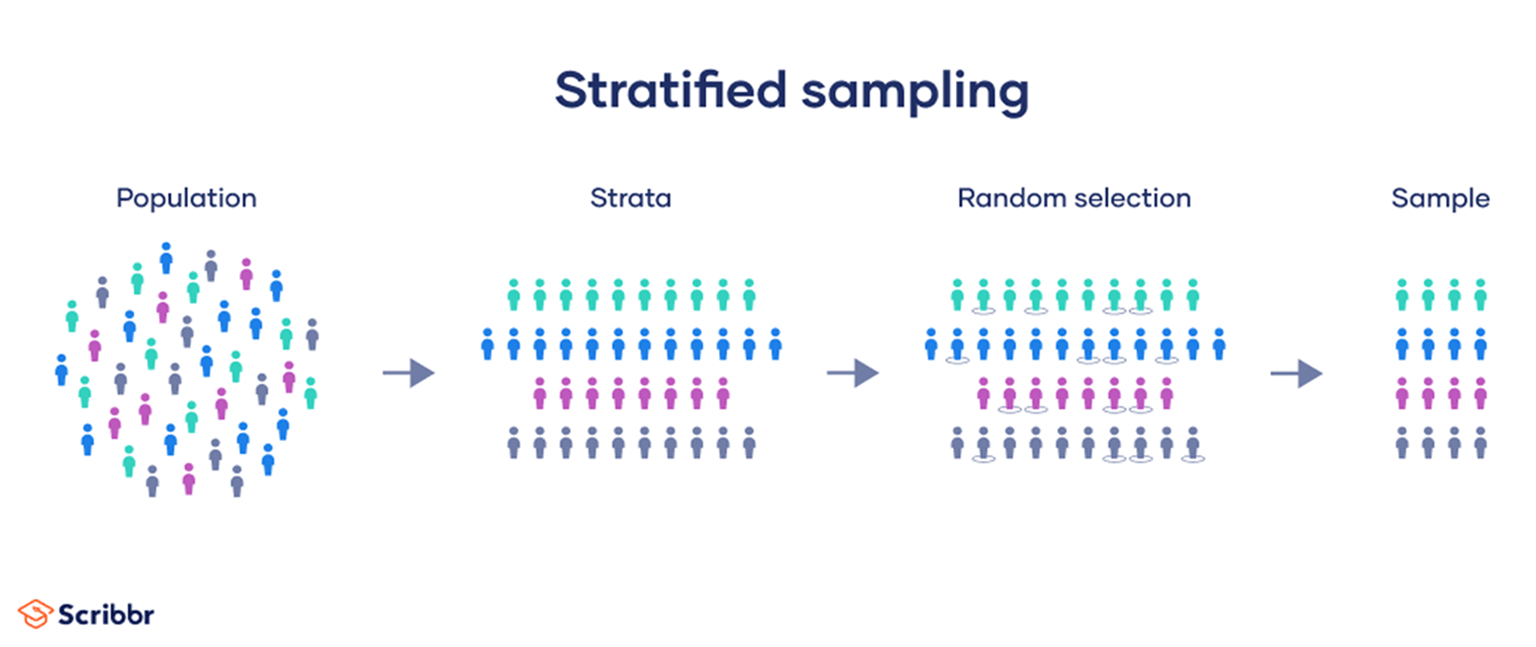


**Stratified sampling**

In a stratified sample, researchers divide a population into homogeneous subpopulations called *strata* based on specific characteristics (e.g., race, gender identity, location, etc.). Every member of the population studied should be in exactly one stratum.

Each stratum is then sampled using another probability sampling method, such as cluster sampling or simple random sampling, allowing researchers to estimate statistical measures for each sub-population.

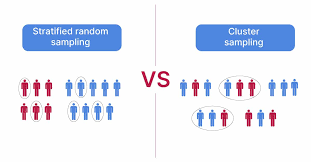
For eg, suppose we have the price details of 1000 food products. In stratified sampling we divide the 1000 food products into different strata (groups) based on their characteristics i.e., meat products, fish products, dairy products etc. Then from each strata a sample is selected based on the sample size.



**Cluster sampling**

In cluster sampling, researchers divide a population into smaller groups known as clusters. They then randomly select among these clusters to form a sample.

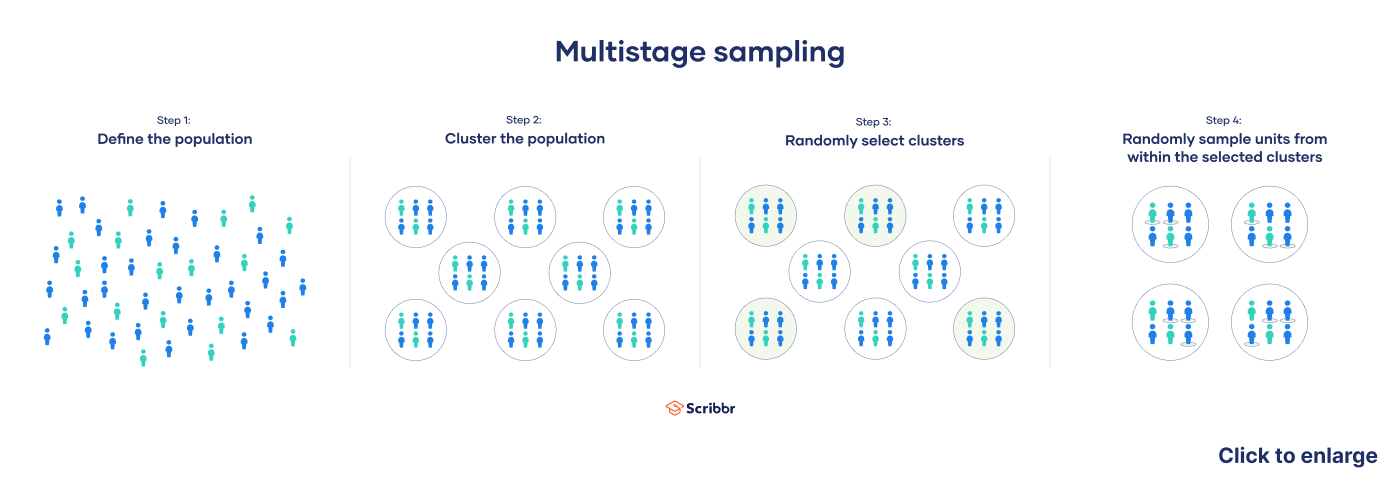
For eg, suppose we have the price details of 1000 food products. In stratified sampling we divide the 1000 food products into different clusters. The observations in each cluster need not be of same characteristics like in stratified sampling. Then clusters are selected into the sample.



**Multi stage sampling**

In multistage sampling, or multistage cluster sampling, you draw a sample from a population using smaller and smaller groups (units) at each stage. It’s often used to collect data from a large, geographically spread group of people in national surveys.

For eg, in multistage sampling the population is divided into clusters and then clusters are selected and from the selected clusters the samples are selected using any of the sampling methods. If the population is 1000, then it is divided into clusters and some of the clusters are selected. Then from each cluster a sample of fixed sample size is selected using any sampling techniques.



**Non -probability sampling**

In a non-probability sample, individuals are selected based on non-random criteria, and not every individual has a chance of being included.

This type of sample is easier and cheaper to access, but it has a higher risk of sampling bias. That means the inferences you can make about the population are weaker than with probability samples, and your conclusions may be more limited.

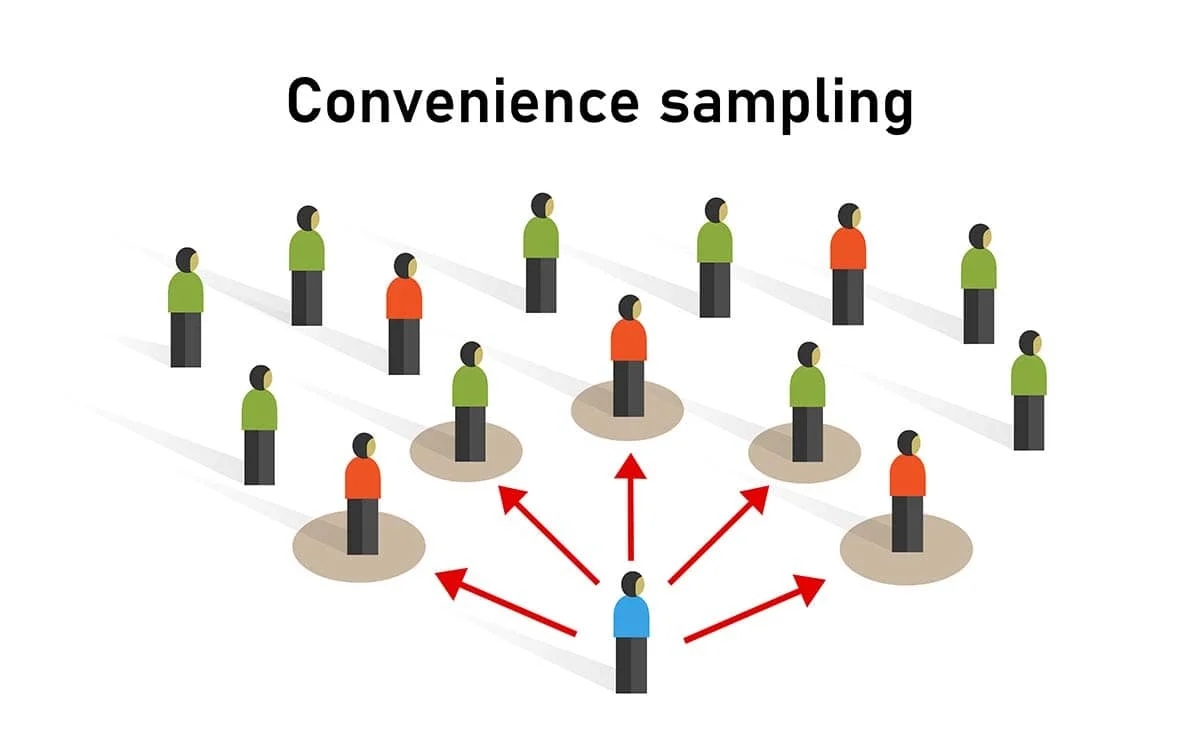
**Types of non-probability sampling**

The different types of non-probability sampling are: -

* Convenience sampling
* Purposive sampling
* Snowball sampling
* Quota sampling

**Convenience sampling**

Convenience sampling is a non-probability sampling method where units are selected for inclusion in the sample because they are the easiest for the researcher to access. This can be due to geographical proximity, availability at a given time, or willingness to participate in the research. Sometimes called accidental sampling, convenience sampling is a type of non-random sampling.

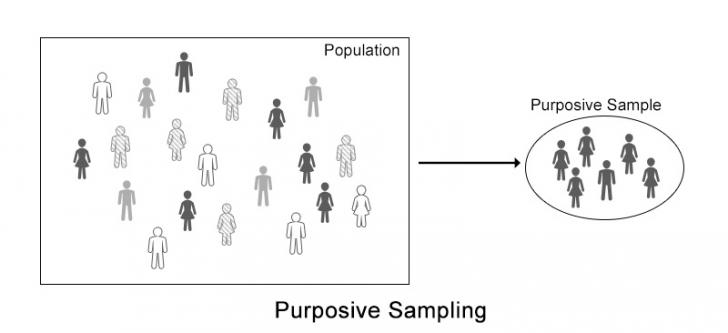


A basic example of a convenience sampling method is when companies distribute their promotional pamphlets and ask questions at a mall or on a crowded street with randomly selected participants.

**Purposive sampling**

Purposive sampling refers to a group of non-probability sampling techniques in which units are selected because they have characteristics that you need in your sample. In other words, units are selected “on purpose” in purposive sampling.

Also called judgmental sampling, this sampling method relies on the researcher’s judgment when identifying and selecting the individuals, cases, or events that can provide the best information to achieve the study’s objectives.

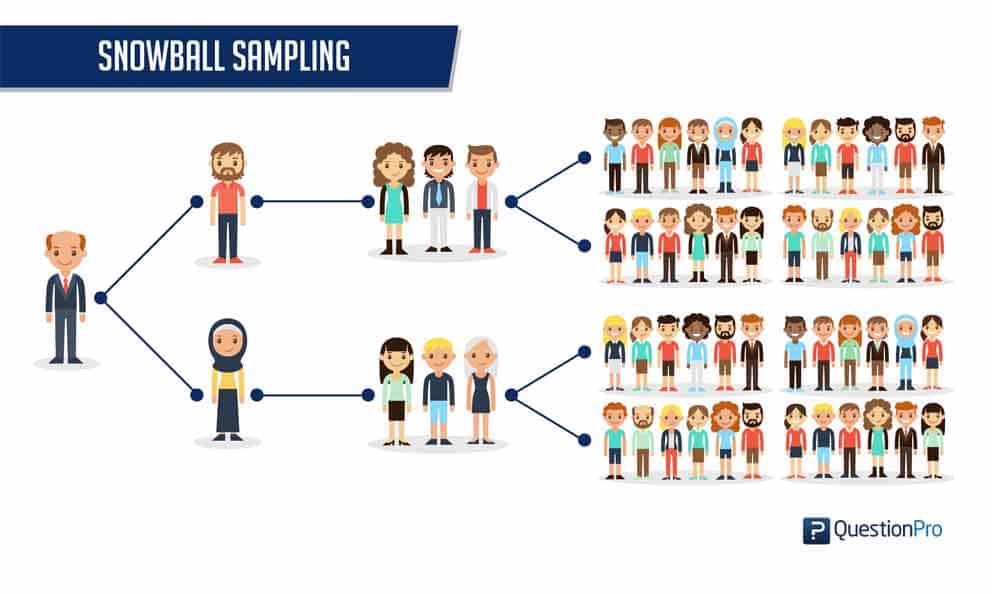


For eg, in an educational survey the individuals selected for the survey will be students, this is a case of purposive sampling.

**Snowball sampling**

Snowball sampling is a non-probability sampling method where new units are recruited by other units to form part of the sample. Snowball sampling can be a useful way to conduct research about people with specific traits who might otherwise be difficult to identify (e.g., people with a rare disease).

Also known as chain sampling or network sampling, snowball sampling begins with one or more study participants. It then continues on the basis of referrals from those participants. This process continues until you reach the desired sample.



Snowball sampling is used when researchers have difficulty finding participants for their studies. This typically occurs in studies on hidden populations, such as criminals, drug dealers, as these individuals are difficult for researchers to access.

**Quota sampling**

Quota sampling is a non-probability sampling method that relies on the non-random selection of a predetermined number or proportion of units. This is called a quota. First the population is divided into mutually exclusive subgroups (called strata) and then recruit sample units until you reach your quota. These units share specific characteristics, determined by you prior to forming your strata.

For eg, A researcher wants to survey individuals about what smartphone brand they prefer to use. He/she considers a sample size of 500 respondents. Also, he/she is only interested in surveying ten states in the country.Then the researcher can divide the population by quotas:

* Gender: 250 males and 250 females
* Age: 100 respondents each between the ages of 16-20, 21-30, 31-40, 41-50, and 51+
* Employment status: 350 employed and 150 unemployed people.

