MAM Sampling

Unit 4

Terminology

Population: The set of all observations under study is called Population.

Population Size: The number of observations in the Population is called size of Population and is denoted by ‘N’. It can be finite or infinite.

Sample: A part / subset of the Population is called Sample.

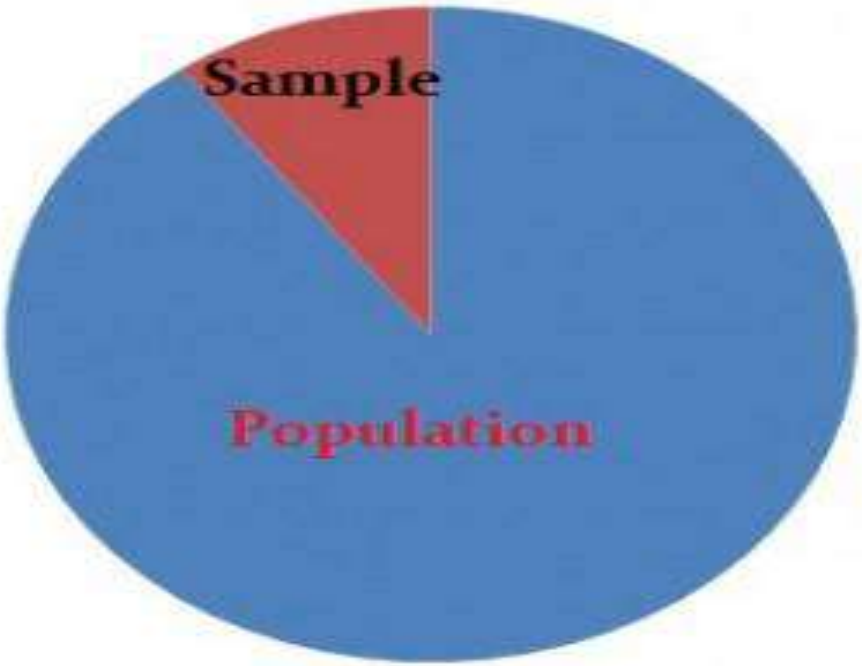
Sample Size: The number of observations in the Sample is called Sample Size denote by ‘n’ and is always finite.

Parameter: All the statistical constants like mean, median, mode, standard deviation computed for the Population are called parameters. Most of the time these parameters are unknown because we don't know all the observations in the Population.

Statistic: All the statistical constants like mean, median, mode, standard deviation computed for the sample are called Statistics. For a sample, statistics can be computed as we know all observations in the sample.

Sampling

- **What Is 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
Population

Advantages of Sampling

Some of the key advantages of sampling are:

i) it costs less

ii) takes less time

iii) data are acquired quickly

iv) fewer mistakes are likely

v) a more detailed study can be done.

- In statistics, you'll be working with samples. A sample is just a part of a population. For example, if you want to find out how much the average American earns, you aren't going to want to survey everyone in the population (over 300 million people), so you would choose a small number of people in the population. For example, you might select 10,000 people.

- Samples are parts of a population. For example, you might have a list of information on 100 people (your "sample") out of 10,000 people (the "population"). You can use that list to make some assumptions about the entire population's behavior.
- However, it's not that simple. When you do sampling, your sample size has to be ideal—not too large or too small. Then once you've decided on a sample size, you must use a sound technique to collect the sample from the population.

- When you conduct research about a group of people, it's rarely possible to collect data from every person in that group. Instead, you select a sample. The sample is the group of individuals who will actually participate in the research.

- To draw valid conclusions from your results, you have to carefully decide how you will select a sample that is representative of the group as a whole. There are two types of sampling methods:

- **Probability sampling** involves random selection, allowing you to make statistical inferences about the whole group.

- **Non-probability sampling** involves non-random selection based on convenience or other criteria, allowing you to easily collect initial data.

- **Probability sampling methods (Random sampling Methods)**
- Probability sampling means that every member of the population has a chance of being selected. If you want to produce results that are representative of the whole population, you need to use a probability sampling technique.

Random Sampling Methods

1. 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.
- **Example**
- You want to select a simple random sample of 100 employees of Company X. You assign a number to every employee in the company database from 1 to 1000, and use a random number generator to select 100 numbers.

2) Systematic Random Sample

Designing a Systematic Random Sample is sometimes quite difficult and time consuming and therefore, Systematic Random Sample, like Simple Random Sample, also uses a list of all members of the population in its sampling frame. However, instead of using random numbers to select the sample elements, the researcher applies a skip interval to the list to produce a sample of the required size.

$$\text{Skip interval} = \frac{\text{number of elements in the population}}{\text{the required sample size}}$$

$$K = \frac{N}{n}$$

K = skip interval

N = Universe size

n = Sample size

For example if we have to select a sample of 100 persons from a universe of 1000 population, then the skip is 10. In this case one number between 1 and 10 has to be selected. Suppose 5 is selected, then the first sample would be 5^{th} and the next one 15^{th}, 25^{th}, 35^{th}, 45^{th}, and so on. One of the advantages of this method is that it is more convenient than other methods and simple to design. Again, it is used with very large populations.

- 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. For this we need a **sample frame**.

- **Example**
- All employees of the company are listed in alphabetical order. From the first 10 numbers, you randomly select a starting point: number 6. From number 6 onwards, every 10th person on the list is selected (6, 16, 26, 36, and so on), and you end up with a sample of 100 people.

- **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).

3) Stratified Random Sample

In Stratified Random Sampling, the target population of N units is first divided into k subpopulations of N_1, N_2, ..., N_k units. These populations are non-overlapping and together they comprise the whole population. So that $N_1 + N_2 + ... + N_k = N$

The sub-populations are called strata. The number in each stratum should be known. A sample is drawn from each stratum independently. The sample sizes within 'k' strata are denoted by n_1, n_2, ..., n_k respectively. If the total sample size 'n' is to be drawn from the target population than $n_1 + n_2 + ... + n_k = n$

If a simple random sample is drawn in each stratum, the whole procedure is described as stratified random sampling.

- This method helps in conducting and managing a large scale survey to be conducted in a country like India. The agency conducting the survey may have field offices in different locations; each one can supervise the survey for a part of the population.

- The basic idea is that it sub-divides the heterogeneous population into homogeneous sub-populations. If each stratum is homogenous in itself, a precise estimate of any stratum mean can be obtained from a small sample, thus, saving a lot of time and cost.

Stratified sampling

- This sampling method is appropriate when the population has mixed characteristics, and you want to ensure that every characteristic is proportionally represented in the sample.
- You divide the population into subgroups (called strata) based on the relevant characteristic (e.g. gender, age range, income bracket, job role).
- From 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.

Property of Strata's:

strata's are **homogeneous within** (that is any two units from the same stratum more or less will have similar properties) and **heterogeneous between** (that is there is lot of difference between one stratum to another).

- **Example**
- The company has 800 female employees and 200 male employees. You want to ensure that the sample reflects the gender balance of the company, so you sort the population into two strata based on gender. Then you use random sampling on each group, selecting 80 women and 20 men, which gives you a representative sample of 100 people.

4. Cluster sampling

- Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole population (cluster is also called mini population, only the size of cluster is small in comparison to population). 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.

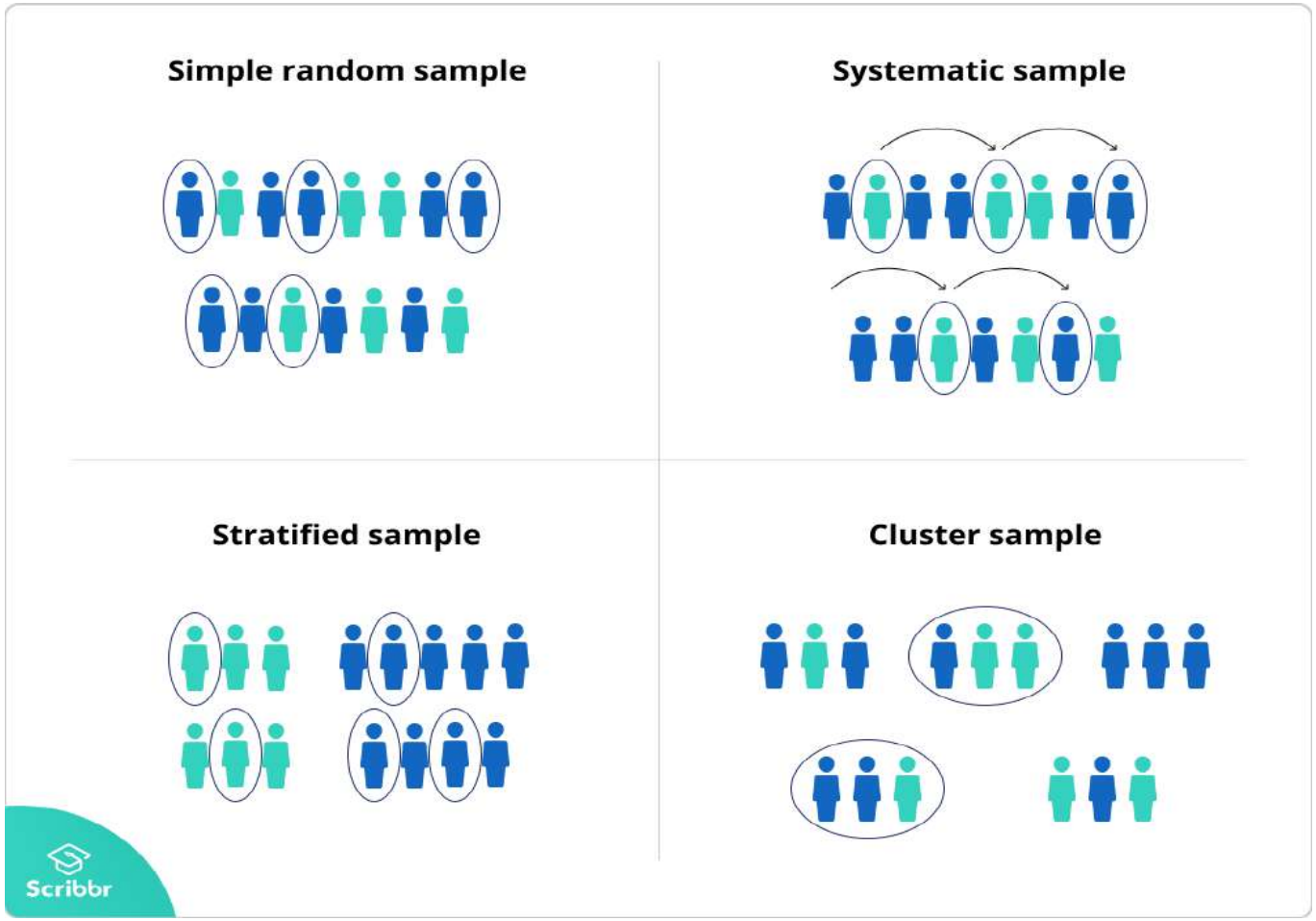
- This method is good for dealing with large and dispersed populations, but there is more risk of error in the sample, as there could be substantial differences between clusters. It’s difficult to guarantee that the sampled clusters are really representative of the whole population.

Property of Cluster's:

Cluster's are **heterogeneous within** (that is any two units from the same cluster can have different properties) and **homogeneous between** (that is there is no difference between one cluster to another).

- **Example**
- The company has offices in 10 cities across the country (all with roughly the same number of employees in similar roles). You don’t have the capacity to travel to every office to collect your data, so you use random sampling to select 3 offices – these are your clusters.

Random sampling Methods



- **Non-probability sampling methods**
- 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 you can't use it to make valid statistical inferences about the whole population.

- Non-probability sampling techniques are often appropriate for exploratory and qualitative research. In these types of research, the aim is not to test a hypothesis about a broad population, but to develop an initial understanding of a small or under-researched population.

1. Convenience sampling

- A convenience sample simply includes the individuals who happen to be most accessible to the researcher.
- This is an easy and inexpensive way to gather initial data, but there is no way to tell if the sample is representative of the population and the results cannot be generalized

Example

- You are researching opinions about student support services in your university, so after each of your classes, you ask your fellow students to complete a survey on the topic. This is a convenient way to gather data, but as you only surveyed students taking the same classes as you at the same level, the sample is not representative of all the students at your university.
- Example—A researcher polls people as they walk by on the street.

2. Voluntary response sampling

- Similar to a convenience sample, a voluntary response sample is mainly based on ease of access. Instead of the researcher choosing participants and directly contacting them, people volunteer themselves (e.g. by responding to a public online survey).
- Voluntary response samples are always at least somewhat biased, as some people will inherently be more likely to volunteer than others.

- **Example**
- You send out the survey to all students at your university and a lot of students decide to complete it. This can certainly give you some insight into the topic, but the people who responded are more likely to be those who have strong opinions about the student support services, so you can’t be sure that their opinions are representative of all students.
- Example—A TV show host asks his viewers to visit his website and respond to an online poll.

- **3. Purposive sampling (Judgment sample)**
- This type of sampling involves the researcher using their judgment to select a sample that is most useful to the purposes of the research.
- It is often used in qualitative research, where the researcher wants to gain detailed knowledge about a specific phenomenon rather than make statistical inferences. An effective purposive sample must have clear criteria and rationale for inclusion.

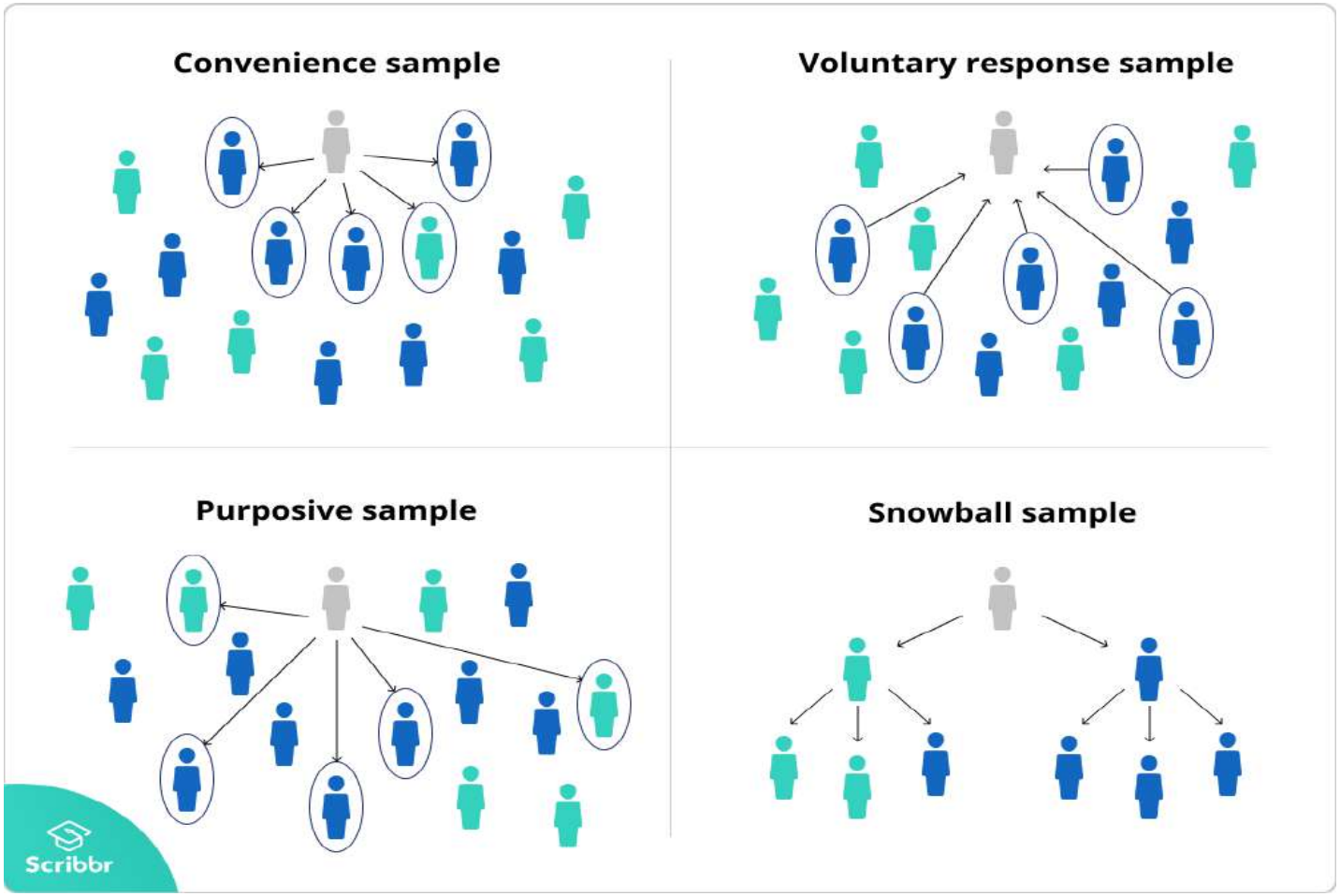
Example

- You want to know more about the opinions and experiences of disabled students at your university, so you purposefully select a number of students with different support needs in order to gather a varied range of data on their experiences with student services.

4. Snowball sampling

- If the population is hard to access, snowball sampling can be used to recruit participants via other participants. The number of people you have access to "snowballs" as you get in contact with more people.
- **Example**
- You are researching experiences of homelessness in your city. Since there is no list of all homeless people in the city, probability sampling isn't possible. You meet one person who agrees to participate in the research, and she puts you in contact with other homeless people that she knows in the area.

Non – Random Sampling



Quota Sample

Quota sampling is like stratified sampling. In quota sampling, the population is categorized into several strata which consist of an expected size, and the samples are considered to be important for the population they represent. The advantages of quota sample are that it involves a short time duration, is less costly, and gives moderate representation to a heterogeneous population.

Types of Sampling