Lec **6** **Biostatistics**

• **Sampling**: -

It is a Procedure to draw a representative sample from the population in such a way that the sample elements selected represent the population.

Sampling Bias: -

Sampling bias is said to occur if the sample unit does not possess the population characteristics. Due to which the result may be inaccurate

Characteristics of good sampling

1. The sample must represent the population correctly.
2. It must be free from bias
3. It must possess the least sampling error
4. It must be optimum in size

Advantages of sampling

1. Reduce time and cost to collect data
2. More comprehensive data is obtained than in a census.
3. Administrative control
4. Better motivation
5. Reliability of data
6. Less non response error

Principles of sampling

1. Selection of sample must be systematic and objective manner
2. Sample unit must be clearly define and easily identifiable
3. Sample units must be independent of each other
4. Same units of sample must be used through out the study
5. The selection process must be on sound criteria.
6. It should avoid error, bias

Process of sampling

1. Defining the population concern

2. Specifying a sampling frame. (Listing the elements of the population)
3. Specifying a sampling method
4. Determining a sample size
5. Implementing the sampling plan
6. Sampling and data collecting

ESSENTIAL OF SAMPLING

1. Representative :- ensure by random selection
2. Adequate: - sample size
3. Independence: - same chance of selection
4. Homogeneity: - No basic difference in nature

Types of sampling

A. Probability sampling

Probability sampling is one in which every unit of the population has an equal probability / chance of being selected for the sample.

• The list of every unit of population must be available.
• It offers high degree of representativeness.
• This method is expensive , time consuming and relatively complicated since its requires a large no. Of size
• This method enable to generalize its finding

B. Non- probability sampling

Non-Probability sampling is one in which every unit of the population doesn't have an equal probability / chance of being selected for the sample.

• It doesn't claim of representativeness.
• The list of entire population is not available to draw a random sampling.
• The researcher who decides which sample unit has to be selected

Sampling methods

a variety of sampling methods can be employed, individually or in combination

A. Probability sampling

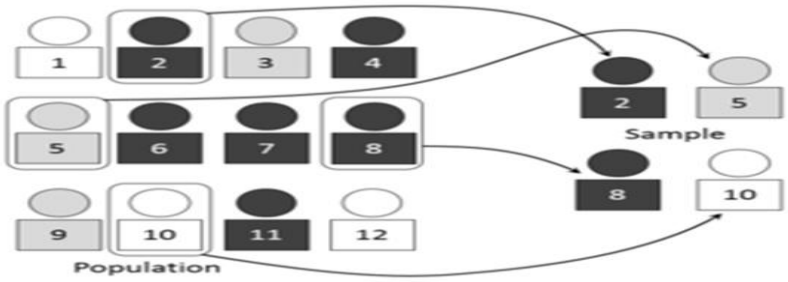
1. Simple random sampling

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.



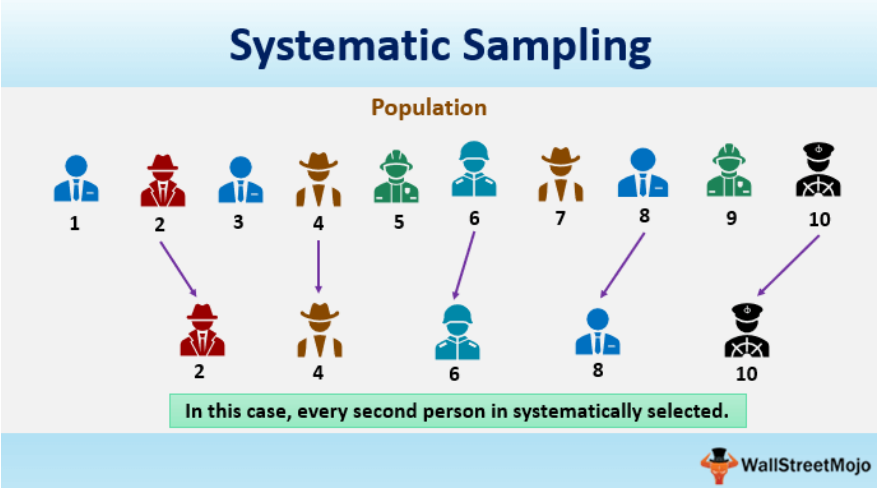
Simple random sampling

2. 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.

Example

All employees of the company are listed in alphabetical order. From the first 10 numbers, you randomly select a starting point: number 2. From number 2 onwards, every 10th person on the list is selected (2, 12, 22, 36, and so on), and you end up with a sample of 100 people.



3. 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.

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

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.

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.

B. 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 it has a higher risk of sampling bias, and 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, so it can't produce generalizable results

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.

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.

3. Purposive sampling

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.