

# Population Sample

## Population

Population is defined by exact definitions of its objects. Objects of a population are given by enumeration or by explicit rule (for example given common property) that enables us to decide for any object whether belongs to given population or not (e.g.: everybody in a town, all male workers in a factory, everybody with hypertension).

A population can be finite or infinite. Finite populations are, for example, demographic populations. Population can also be considered as an abstract idea of a collection of **objects**. These objects are described by **variables**:

- Qualitative variables: Properties that cannot be measured but only described in words (family status, level of pain).
- Quantitative variables: Can be measured and expressed as a number (weight, BP, BMI).

## Sampling

Sampling is a process which enables us to make a statement of the whole by just examining a part of the population, to a certain degree of validity. That is why it's important to make your group representative of the whole population. Before one can start sampling, important questions to ask are:

- Who is included/excluded?
- Geographical considerations?
- Time period?
- Sample size?

## Sampling Frame

A further step is to examine if any lists of the **population units** (sampling frames) exist, such as a household or a person, depending on the survey. A **sampling frame is a list of all elements in the population**. It is very time consuming to construct a frame de novo and thus we usually use the same approach that previous studies used for a specific population. A common sampling frame are the Electoral Register Lists.

## Types of Sampling

There are 2 main types of sampling:

1. Probability (random) sampling
2. Non-probability sampling

Examples of non-probability sampling can include:

- Convenient sampling, e.g.: selecting near you.
- Purposive sampling, that is you subjectively choose people who fit your study.
- Selecting the extremes from the normal (deviant case).

Probability (random) sampling:

- Every object of the population has a known non-zero chance of being selected, in contrast to non-probability sampling choice of selection of sampling units depends entirely on the decision of the sampler. Therefore, inductive statistics is based on probability (random) sampling methods only.
- How to select a random sample:
  1. Define the population.
  2. Construct a sample frame if one does not exist already.
  3. Give each element a unique ID starting from 1 to the number of elements in the population N. Use only the minimum number of digits.
  4. Select people at random using tables of random number or computer generated random numbers.

## Types of Probability (random) Sampling

- **Simple random sampling (SRS)**: Each object has an equal chance of being selected. This is the type of sampling assumed by most of the statistical packages and it is the standard to which all methods are compared.
- **Multi-stage sampling**: As the name suggests the sample is selected in stages. This type of sampling is used in some way in the vast majority of all social surveys, which involve interviews.
  - Process:
    1. The population is divided up into *hierarchical units*. Clusters of the elements that occur naturally in the population usually form these units, e.g. they live near each other.

2. We then select a cluster and then from the selected cluster we select the people (randomly).
- Advantages:
    1. Reduction in cost relative to simple random sampling.
    2. When there are no suitable sampling frames for the entire population, we can construct a list of all the areas in the population but we don't need to construct a sampling frame - (list of all the elements for the selected areas), therefore substantially reducing the cost of drawing up a sampling frame. An area is often called a Primary Sampling Unit (PSU).
  - Disadvantages:
    1. For the same sample size, the value (measured in terms of precision) for multi-stage is usually less than simple random sampling.
  - **Stratified random sampling:** The population is divided up into (non-hierarchical) groups prior to selecting the sample, e.g. by state or phone area codes and separate samples are selected from each group, randomly.
    - Advantages:
      1. Due to the breakdown of data it can lead to administrative convenience, such as organizing and allocating resources nearby to the strata (e.g.: cities) or accommodating to specific needs of each strata.
      2. Different sampling techniques may be used in each of the stratum, according to their distinct natural characteristics and problems. This will eventually result in smaller standard errors and sampling errors, in comparison to SRC.
      3. All the groups of a population are adequately represented, regardless if they are minorities.
      4. We get a better cross-section of the population.
  - **Systematic random sampling:** This is a widely used sampling technique. It consists of taking every  $n^{\text{th}}$  sampling unit after a *random start*, from a list of uniquely ID-allocated population members.
    - Advantages:
      1. An easy, almost foolproof and flexible method to sample.
      2. If the list is stratified beforehand, the sample will reflect this ordering and as such can easily give an a stratified, systematic random sample.
    - Disadvantage: If there is any previous ordering (sorting) in the list and it is unknown to the researcher, this may bias the resulting estimates.

## Links

### Related articles

- Descriptive Studies
- Analytical Studies

### Bibliography

- BENCKO CHARLES UNIVERSITY, PRAGUE 2004, 270 P, V, et al. *Hygiene and epidemiology. Selected Chapters*. 2nd edition. Prague. 2008. ISBN 9788024607931.