

A red ribbon graphic with a central rectangular box containing the text "Data collection:". The ribbon has a 3D effect with shadows and highlights.

Data collection:

A green oval graphic with a 3D effect, containing the text "Lecture 2".

Lecture 2

It is difficult for us to study all population of interest to reach a conclusion regarding certain parameter (variable) and the effect of different factors on such parameter (variable), as it needs time, money, efforts, and manpower, through **census** of the population which is done in our country every 10 years (1957, 1977, 1987, 1997, 2007????) but also census will give us only a **demographic characteristics** of the population (no medical information can be gathered from census, as it is done by professionals but not doctors or medical staff), so a sample is taken from the population by **sampling** which is as representative as possible for the population and it is done properly we can **generalize** our finding on the population.

- **Sampling:**

There are two main types of sampling;

- 1) Probability sampling (random sampling): which is the best method that allows us to infer from the sample drawn to the population.
- 2) Non-probability sampling (Non-random sampling).

Non-Random sampling:

- 1) Quota method (Accidental sampling):** The selection is done by taking a predetermined sample size at a predetermined time and place.
- 2) Systematic method:** By which the selection is done through a predetermined regular interval (e.g. each 3rd, each 10th, each house in the corner, etc...).

- **Random sampling:**

- In this type of sampling, each person in the population has an equal chance (equal probability) to be included in the sample as the others. So there is no **Bias** (the selection is biased if it systematically favors certain outcome) or no other thing that prefer any person to be included in the sample;

- **1-Simple random sampling:**

- In simple random sampling;
- Listing
- Coding
- Selecting achieved by **card, random digit table, and by computer** (already arranged numbers or system wise numbers).

2-Systematic random sampling:

Sometimes, the sampling frame does not exist, in case if we are dealing with **infinite** population (population composed of endless number, such as patients attending the outpatient clinic), while in **finite** population, there is a sampling frame. So it is convenient to carry out sampling in systematic way (through regular interval), the interval is determined according to the total number of population assumed and the number of the sample required;

Assumed total number of population

----- = nth

Number of sample required

The starting point from the first 5 digits is chosen at random by simple random sampling, suppose it was taken as 4, so the sample will comprise individuals with numbers 4, 9, etc....

- **3-Stratified random sampling:**

This type of sampling is used when we have a population composed of quite different strata or distinct subgroups, the selection of a sample that does not respect these distinct subgroups will give us a sample that may be totally composed of males or of females or of different percentages of males and females as that of the population, from each subgroup by **epsem** method (equal probability selection method) or **pps** method (probability proportional to size).

- **4-Multi-stage random sampling:**

When we have a large population extended over a large geographical area, it is better to carry a multi-stage random sample we select by simple random sample an epsem or pps scheme.

- **5-Cluster random sampling:**

This type of sampling when we have a very little extra cost, effort, manpower, so we can investigate or choose the whole last stage rather than only taking a sample of them. The last stage is referred to as **cluster**.

- **Reasons for sampling;**

- 1) Sample can be studied quickly.
- 2) Less expensive than studying the entire population
- 3) Sample results are usually more accurate than results obtained from population.
- 4) Probability methods can be used to estimate the error in the resulting statistics.
- 5) To reduce the heterogeneity.

- **Data organization (Ordered array):**

It is the enlistment or the arrangement of the data according to their magnitude from the smallest to the largest or from the largest to the smallest. The benefits of ordered array are;

- 1) Determine the smallest value (XS) and the largest value (XL).
- 2) Determine the range.
- 3) Easy to present the data by table.
- 4) To find the value of median.

END