



☒ Sheet

☐ Slides

Number

2

Done by:

Omayma Hassanin

corrected by:

Mariam AL-Bayai

Doctor

Mahmoud Al Hussami

Note that I have included anything from the slides within "".

### **Definition of Biostatistics:**

Biostatistics is a mathematical approach through which special techniques are used in order to organise, analyse, and generalise the results, which are presented to decision makers in the society.

Biostatistics is a growing science, especially in medicine, nursing, healthcare, and pharmaceutical and biological applications. It is a branch of public health, as it is a main tool in epidemiological studies.

Biostatistics is also a study of the determinants and distribution of any health event. Therefore, data is a crucial requirement in bio-statistical studies through which determinants and distribution of health events are studied. In other words, in biostatistics, data, which are gathered by the methods and designs of epidemiology, are analysed.

“Concerns of biostatistics :

-Gathering, organising, summarizing and analysing data from a set or group or subgroup of people. Consequently, results would be inferred and generalised to the population.

-We seek to draw inferences about a body of data when only a part of the data is observed.”

Because a study on the whole population is very difficult to be performed, a sample of the population which is representative of this population would be used in a bio-statistical study. For instance, if a study to know the mean of weights of Jordanian individuals is performed, a representative sample of the Jordanian population would be taken, and the statistical mean of the weights would be calculated from the sample.

Note that it is very important for the sample of the population to be representative of the population as a whole, as the data collected from this sample is used to deduce generalisations about the population, and a study based on a representative sample would have minimal errors.

## **Purpose of Biostatistics:**

“To describe and summarize information thereby reducing it to smaller, more meaningful sets of data.

To make predictions or to generalize about occurrences based on observations.

To identify associations, relationships or differences between the sets of observations.”

The purpose of biostatistics depends on the type of bio-statistical approach used.

Biostatistics has two main types:

- 1- Descriptive: through descriptive statistics, data is organised and summarised in tables or graphs.
- 2- Inferential: through inferential statistics, generalizations can be made, as in, a generalization made from a representative sample to a general population.

In other words, two main types of purposes of biostatistics: inferential and descriptive.

## **Data:**

Data used in biostatistics” are numbers which can be measurements or can be obtained by counting.” The data primarily gathered are called raw data. Raw data cannot be immediately used as it should be organised first, and the mathematical median cannot be calculated from raw data before being organised and reviewed as errors within the raw data might be present.

So, “Biostatistics is concerned with the interpretation of the data and the communication of information about the data.” → Interpreting and communicating data to decision makers such as the ministry of health to assist them in taking important decisions, eg. Building a hospital in a certain district requires taking many factors into consideration, such as the number of people in the district at which the hospital will be built, the diseases more prevalent in that district, the needs of the individuals, etc.

## **Sources of Data:**

Data can be obtained from:

### **1- Surveys→**

Surveys are administered questionnaires regarding an issue to a group of participants, we collect their opinions regarding this issue at a specific period of time. Such surveys can also be called: Cross-sectional surveys.

So, if we have a series of surveys regarding a particular issue to be studied, the survey is considered longitudinal (surveillance). Surveillance is the most accurate type of survey. Through surveillance, we study and observe a group of people over a particular period of time (Observational). However, a longitudinal study is not only presented through a surveillance, it can be presented through clinical trials (Experimental).

→ There are three factors which are present in clinical trials but are absent in observational studies:

- a- Randomisation among the groups studied.
- b- Control group.
- c- Manipulation of the independent variable.

\*00:00→ 11:33 min

### **2- Analysis of records:**

In records (medical records for example), data have been already collected and are relatively perfect. However, such data in records could need further organising, be expired, have some errors, or collected in different seasons.

### **3- Counting:**

For example, a census of the whole population through which all individuals in the population are counted. Data collected from the whole population through the census is 100 percent accurate.

### **4- Experiments:**

In hospitals, experiments are called clinical trials, while in labs they are called experiments. Human beings are very complex and it is impossible to study them in labs. Studying humans in clinical trials could subject them to unethical tests, such as using violence against the people on which experiments are done. For this reason, IRB (Institutional Review Board) has been created in hospitals, universities, and every

research institution to see the researchers stick with the law and ethics and to prevent violation of ethics when conducting clinical trials. (Example of clinical trials which violate ethics: 1) the experiments done on prisoners by Germans to test the durability of soldiers to cold weather in case they invaded Russia (they studied how long they can live and survive); 2) In 1984, experiments done on 500 hundred black men suffering from syphilis to see its prognosis, even though penicillin was discovered in 1935 and is very effective against syphilis symptoms they didn't give them penicillin. 3) And an experiment where they injected retarded patients with cancer cells in order to see if they will develop cancer).

So we have ethics that control research and because of that IRB was established, before the researcher gather data he must have a permission especially if it's a clinical trial.

### **5- Reports:**

Secondary data could be collected from journals and reports.

Note: Primary data is data the researcher him or herself gathers, while secondary data is data which already exist and have been already gathered through approaches unknown to the researcher ( variables might have different names than the ones in the primary data , might have errors , might be expired ).

### **Variables:**

Variable is the opposite of a constant. It is something which varies, eg. Different incomes for different families, and thus we can study the relationship between the income and quality of life.

"1- A variable is an object, characteristic, or property that can have different values.

2- A quantitative variable can be measured in some way.

3- A qualitative variable is characterized by its inability to be measured but it can be sorted into categories."

### **Random Variables:**

"A variable that cannot be predicted in advance because it arises by chance ." The outcome cannot be predicted. "Observations or measurements are used to obtain the value of a random variable."

There are two types of random variables (different according to paradigms\*):

- a- Discrete random variables: “A discrete random variable has gaps or interruptions in the values that it can have. The values may be whole numbers or have spaces between them.” (Integers) eg. Number of children in a family.
- b- Continuous random variables: “A continuous random variable does not have gaps in the values it can assume. Its properties are like the real numbers.” Eg. How much change someone has (the value can be presented in decimals).

\*Paradigms:

Two major paradigms to obtain the knowledge:

- a- Positivism: Number oriented (Quantitative) → can be statistically measured; correlations can be concluded, for them truth is rigid.
- b- Naturalistic: Themes and narratives (Qualitative).

In biostatistics we worry about quantitative. Qualitative is for categories.

\*11:33 → 31:20 min

### **Population and Samples:**

The whole population studied → parameter; no statistics.

Samples → Statistics used. Eg. mean; standard deviation; mode; etc.

The use of statistics applies when you inference or generalise information collected from a subgroup to the population. In other words, to use statistical approaches, the whole population cannot be studied (costly; takes too much time; means of accessing the whole population can sometimes be absent), because it is more practical to study a representative subgroup of the population and then generalise the data collected to the whole population.

“-A population is the collection or set of all of the values that a variable may have. The entire category under consideration.”

“-A sample is a part of a population. The portion of the population that is available, or to be made available, for analysis.”

## **Population and Sampling:**

“Sampling: the process of selecting portion of the population.” → selecting the participants in a study.

The sample chosen during the sampling process should be representative in order to generalise the data to the whole population→ “Representativeness: the key characteristic of the sample is close to the population.”

Sampling bias is when the researcher excludes a participant or group of participants from the population without any reason. In other words, “excluding any subject without any scientific rational. Or not based on the major inclusion and exclusion criteria.”

Inclusion criteria: scientific criteria on which the researcher decides whom to include in the study.

Exclusion criteria: scientific criteria on which the researcher decides whom to exclude from the study.

An appropriate reason is needed to exclude or include a subject in a study→ otherwise, sampling bias→ negative effect on generalizations.

“Example:

-Studying the self-esteem and academic achievement among college students.

-Population: all student who are enrolled in any college level.

-Sample: students’ college at the University of Jordan.”

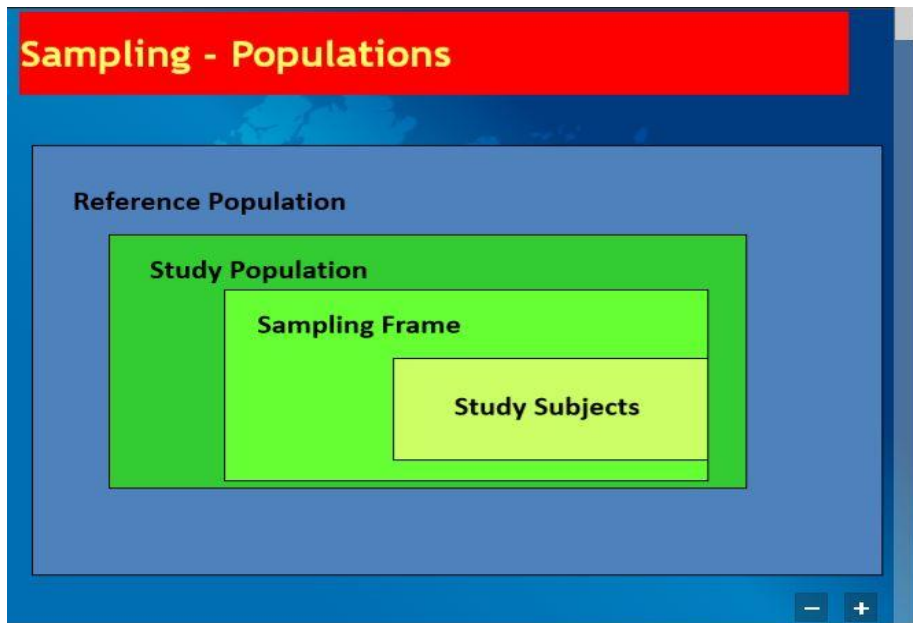
So, sampling is the selection of a number of study units /subjects / participants from the accessible/ defined population.

The population is two types:

- 1- Reference or target population→ To whom the researcher is going to generalise the results.
- 2- Accessible/ study population→ The group of people from which the researcher draws the sample.

Sometimes the reference sample is not always present, while the accessible population is present. Eg. Reference population → all Jordanian people, including those living abroad; accessible population → only those who live in Jordan.

\*31:20 → 40:20 min



In the diagram:

- Study population= accessible population → the group from which the sample is drawn.
- The reference population is larger than the study/ accessible population.
- Sampling frame: usually, a frame list is created for the accessible/ study population, especially if there are few study subjects and the researcher wants to do a simple random sampling.

“Definition of population: a complete set of persons or objects that possess some common characteristic that is of interest to the researcher.” As mentioned above, there are two types:

- 1- Target population: “The entire group of people or objects to which the researcher wishes to generalize the findings of a study. Target population should meet the criteria of interest to the researcher.



Example: all people who were admitted to the renal unit for dialysis in AlBasheer hospital during the period of 2014 – 2016.”

2- Accessible/ study population: “The group of people or objects that is available to the researcher for a particular study.”

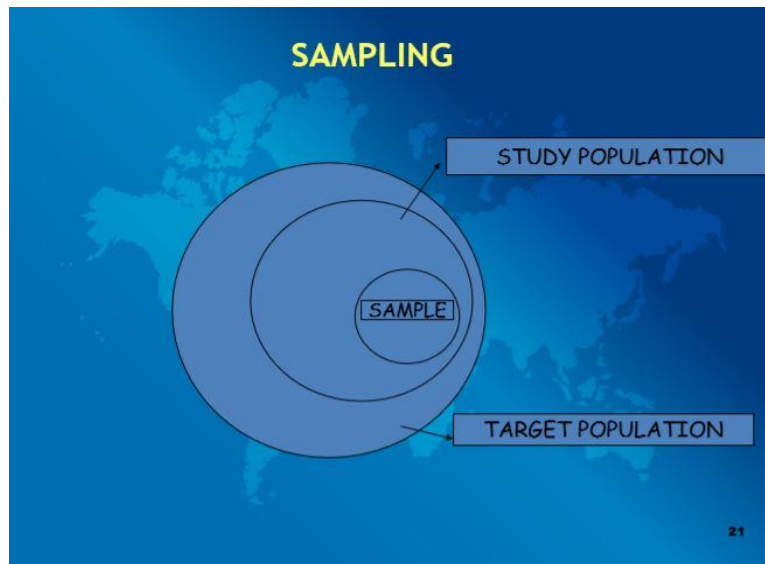


Diagram:

- Target/ reference larger than study/ accessible population.
- Sample drawn from the study population.

### **Sampling:**

A method of choosing or selecting the elements.

“Element is the single member of the population (population element or population member are used interchangeably)”.

Sampling frame is the list of the study population. → “Sampling frame is the listing of all elements/ units that compose the study population, i.e., a list of all medical students at the university of Jordan, 2014-2016.”

## **Sampling Methods:**

Sampling depends on the sampling frame which is the list of all accessible population.

**Data that we collect from the sample are 2 types : Primary and Secondary Data:**

**Primary data:** “It is the data that has been compiled by the researcher using such techniques as surveys, experiments, depth interviews, observation, focus groups.”

“A lot of data is obtained using surveys. Each survey type has advantages and disadvantages.”

Types of surveys:

- 1- Mail: Lowest rate of response (negatively affects generalisations); usually the lowest cost.
- 2- Personally administered surveys (interviews): the best type, “can probe , most costly , interviewer effect ( the interviewer might influence the response )”

This one has 2 types :

- a- Structured/ Semi-structured
- b- Unstructured

3- Telephone: response rate is low; fastest.

4- Web: also, one of the best; fast and inexpensive; response rate is not always good.

**Secondary data :** next lecture

\*40:20→ 48:00 min