

Introduction to Biostatistics

By

Dr. Asaad Al-Yassen
Assist. Professor of dermato-epidemiology
& occupational health
Department of Community Medicine
University of Basrah

Statistic plays an important role in all areas of science. Statistical models and methods allow us to take sample of data from our area of focus and answer questions about the population of objects or events that we are interested in studying.

Objectives

At the end of this course, you should be able to:

1. Classify variables as qualitative, quantitative, continuous or discrete
2. Summarize the range of a series of quantitative measurements into intervals.
3. Construct a frequency distribution and relative frequency distribution for given a set of raw statistical data
4. Present a summary of raw data in the form of tables and graphs
5. Calculate the arithmetic mean, median and mode of a given set of observations
6. Understand various measures of variability (range, variance, standard deviation, standard error and coefficient of variation).
7. List the descriptive characteristics of the normal distribution curve
8. Use, and interpret the results, of certain statistical tests (test of significant).

Objectives of this lecture:

A student is expected to understand

1. What is meant by statistics and biostatistics?

2. Importance of statistics in Medicine
3. Types of variables (qualitative and quantitative variables).
4. What is meant by descriptive statistics?

Statistics: is the science of collecting, summarizing, presenting and analyzing data, and using them to test hypothesis.

Roughly speaking, the field of statistics can be divided into:

* **Mathematical Statistics:** the study and development of statistical theory and methods.

* **Applied Statistics:** the application of statistical methods to solve real problems involving randomly generated data and the development of new statistical methodology motivated by real problems.

Biostatistics: is the branch of applied statistics directed toward applications in the health sciences and biology.

Importance of statistics in medicine

Statistics is important for:

1. Planning, conducting , and interpretation of medical research .
2. Understanding and Evaluating medical literature
3. Definition of normal,(what is the normal value? & what is the abnormal value?)
4. Studying the reliability of laboratory tests.
5. Studying the effectiveness of treatment.
6. calculated frequency of cases or disease.

Data: are observations of random variables made on the elements of a population or sample.

* Data are the quantities (numbers) or qualities (attributes), measured or observed, collected and /or analyzed.

* The word “data” is plural, “datum” is singular.

* A collection of data is often called a data set (singular).

Data: are measurements or observations.

A **data set** is a collection of measurements or observations.

A **datum** (singular) is a single measurement or observation .

Population: A population is the set of all possible individuals, objects, or measurements of interest in a particular study (the largest collection of anything).

Sample: A sample is a portion, or part, of the population of interest.

Variable: is a characteristic that takes different values in different persons, places or times.

Examples: Gender, intelligence, age, blood urea, height, weight..... etc.

Variables may be classified into two types as follow:

Types of Variables

➤ **Qualitative Variables**

Variables which can't be measured in usual sense but can be described called **categorical** variables

Examples: Gender , ethnicity, religious affiliation, blood group,.....etc.

➤ **Quantitative Variables**

Variables that can be measured , variables that have numeric value

Examples: age, height, blood urea,etc.

Quantitative variables can be subdivided into two types:

- **Discrete** - characterized by gaps or interruptions in between the values (i.e. can't assume fraction like 2.3 persons).
(counts, how many)

Examples: number of children in a family, number of bedrooms in a house, age last birthday

- **Continuous** –variables that don't have gaps or interruption , i.e. can take on any value (for e.g. we can say the weight is 25.8 kg).
(*measurements, how much*)

Examples: weight, height, serum cholesterol , BP, age.

SCALES used to measure variables include:

Nominal Scale

Each measurement is assigned to limited numbers of unordered categories & fall in only one category (i.e. the information of an individual put the individual in one category only). e.g. gender, religious affiliation & blood groups.

Variables measured on nominal scales are also called categorical .

Ordinal Scale

Each measurement is assigned to one of a limited number of categories that are ranked in a graded order. Differences among categories are not necessary to be equal & often not measurable.

Involves data that may be arranged in some order, but differences between data values cannot be determined or are meaningless.

Examples of ordinal Scale :

Socioeconomic Status

1 = Low

2 = Middle

3 = High

Health Status

1 = Poor

2 = Fair

3 = Good

4 = Excellent

Interval Scale

Each measurement is assigned to one of unlimited number, that are equally spaced with no true zero point,(i.e. it does not begin from zero due to the presence of minus numbers), e.g. temperature.

Ratio Scale

A Ratio scale is the most precise level of measurement. Measurements begin at true zero point & the scale has equal intervals (Differences among ratios are meaningful for this level of measurement).

Examples: money, height , Weight, blood pressure

Scale	Characteristic Question	Examples
Nominal	Is A different than B ?	Marital status Eye color Gender Religious affiliation Race
Ordinal	Is A bigger than B ?	Stage of disease Severity of pain Level of satisfaction
Interval	By how many units do A and B differ?	Temperature
Ratio	How many times A bigger than B ?	Distance Length Time until death Weight

Descriptive Statistics

Presentation of data can be:

- Tabular: using tables.
- Graphical: using graphs.

- Pictorial: using pictures or charts.

Summarization of data by calculation of:

- a. Measures of Central Location
- b. Measures of Variation (dispersion)