17

BIOSTATICS

KSU COLLEGE OF MEDICINE
2019 - 2020

ACKNOWLEDGMENTS

TEAM MEMBERS

YAZEED ALKHAYAL

SHAHAD ALDUMKH

REVIEWER

ASEEL BADUKHON

Special thanks to SARAH ALENEZY & 436 TEAM

TABLE OF **CONTENTS**

THIS			LECTURE
	BIOSTATISTICS		
IS			
IMPORTANT!!!!			EXTREMELY

LECTURE **OBJECTIVES**

By the end of this lecture, I am able to:

- Definition of statistics and biostatistics
- To understand different Levels of measurements
- To understand different Types of data
- To use these concepts appropriately

Basic concepts & terminology in Biostatistics

Statistics

(Like any other sciences e.g. mathematics, physics or medicine)

- The science of conducting studies to collect, organize, summarize, analyze, present, interpret, and draw conclusions from ***data***.

Any values (observations or measurements) that have been collected

What are Statistics?

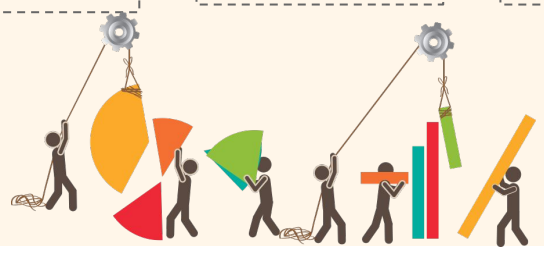
Collecting Data	Characterizing Data	Presenting Data	Interpreting Data
Examples: Sample, Survey, Observation, Simulation	Organize/ Classify, Count, Summarize	Tables, Charts, Statements	Infer, Conclude, Specify Confidence

Why do we need data analysis? To make a decision!

If you made any mistake at any of these steps then, your decision making process and your conclusion will be wrong hence, the importance of methodology.
Remember! the software, the computers or even a high profile statistician can not help you if you have a poor quality data.

Areas where statistics are used

- Mathematics, Physical sciences, Astronomy, Education.
- Government Census, Planning, National Defense
- Business, Economics, Engineering, Marketing, Computer Science
- **Public Health & Medicine Epidemiology, Pharmacology, Genetics.**
- Environment, Agriculture, Ecology, Forestry, Animal Populations



Biostatistics

- Biostatistics is the science that helps in managing medical uncertainties and variability of data
- Statistics arising out of biological sciences, particularly from the fields of medicine and public health.
- The methods used in dealing with statistics in the fields of medicine, biology, and public health for planning, conducting, and analyzing data which arise in investigations of these branches.
- Any health related subject where we're using statistics is called biostatistics or medical statistics or clinical statistics.
- In case of complicated surgeries or accident victims, doctors/surgeons can't guarantee if patients are gonna survive this is an example of **medical uncertainties.**

Basic concepts:

Data: Set of values of one or more variables recorded on one or more observational units (singular: Datum)

Sources of data:

1. Routinely kept records
2. Surveys
3. Experiments
4. External source

What is difference between uncertainties and variability of data?
Variability: It is a quantitative description of the range or spread of a set of values and can expressed through statistical metrics such as variance, standard deviation. Also variability can't be reduced, but it can be better characterized. Uncertainties: incomplete understanding of the context of the risk assessment decision. It can be qualitative or quantitative. It can be reduced or eliminated with more or better data.

Categories of data:

1. Primary data: observation, questionnaire, record form, interviews, survey. (you're collecting data by your own)
2. Secondary data: census, medical record, registry. (already collected by someone else)

Datasets and Data Tables

- Dataset: Data for a set of variables collection a group of persons
- Data Table: A dataset organized into a table, with one column for each variable and one row for each person.

Typical Data Table:

Definitions for variables:

Age: Age in years

BMI: Body mass index (weight/height^2) in kg/m^2

FFNUM: The average number of times eating "fast food" in a week.

TEMP: High temperature for the day

GENDER: 1-Female 2-Male

EXERCISE LEVEL: 1-Low 2-Medium 3-High

QUESTION: What is your satisfaction rating for this Biostatistics session?
1-Very Satisfied
2-Somewhat Satisfied
3- Neutral
4- Somewhat dissatisfied
5- Dissatisfied

OBS	AGE	BMI	FFNUM	TEMP (oF)	GENDER	EXERCISE LEVEL	QUESTION
1	26	23.2	0	61.0	0	1	1
2	30	30.2	9	65.5	1	3	2
3	32	28.9	17	59.6	1	3	4
4	37	22.4	1	68.4	1	2	3
5	33	25.5	7	64.5	0	3	5
6	29	22.3	1	70.2	0	2	2
7	32	23.0	0	67.3	0	1	1
8	33	26.3	1	72.8	0	3	1
9	32	22.2	3	71.5	0	1	4
10	33	29.1	5	63.2	1	1	4
11	26	20.8	2	69.1	0	1	3
12	34	20.9	4	73.6	0	2	3
13	31	36.3	1	66.3	0	2	5
14	31	36.4	0	66.9	1	1	5
15	27	28.6	2	70.2	1	2	2
16	36	27.5	2	68.5	1	3	3
17	35	25.6	143	67.8	1	3	4
18	31	21.2	11	70.7	1	1	2
19	36	22.7	8	69.8	0	2	1
20	33	28.1	3	67.8	0	2	1

Types of variables & data

- When collecting or gathering data we collect data from individuals cases on particular variables.
- A variable is a unit of data collection whose value can vary. (from someone to another)
- Variables can be defined into types according to the level of mathematical scaling that can be carried out on the data.
- **There are four types of data or levels of measurements:**
  - 1. Nominal 2. Ordinal 3. Interval 4. Ratio

Scales of Measurement

Remember the highlighted words!

Nominal scale variables

(names) Its mean I can't mix between the data. For example: females and males data, smoker and non smoker data.

A type of categorical data in which objects fall into **unordered categories**
Studies measuring nominal data must ensure that each category is mutually exclusive and the system of measurement needs to be exhaustive
Variables that have only two responses i.e. Yes or No, are known as *Dichotomies*.

Ordinal scale variables

Examples of ordinal: grades, cancer stages.

Ordinal data is data that comprises of categories that can be rank ordered.

Similarly with nominal data, the distance between each category cannot be calculated but the categories can be ranked above or below each other.

Interval scale variables

Mathematically not strong, zero value isn't fixed (we can't compare)

Fahrenheit temperature scale -zero is arbitrary- 40 degrees is not twice as hot as 20 degrees.
IQ tests. No such thing as zero IQ. 120 IQ is not twice as intelligent as 60 IQ.
Can we assume that attitudinal data represents real, quantifiable measured categories? (i,e, Very happy is twice as happy as plain "Happy", or "Very Unhappy" means no happiness at all). "Statisticians are not in agreement with this.

Ratio scale variables

Mathematically strong, zero value is fixed (we can compare)

The distance between any two adjacent units of measurement (intervals) is the same and there is a meaningful zero point.
Income; someone earning SAR20,000 earns twice as much as someone who earns SAR10,000.
Height Age Weight.

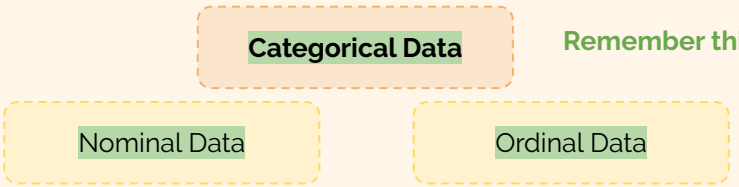
Hierarchical Data Order

- These levels of measurement can be placed in a hierarchical order:

  Ratio > Interval > Ordinal > Nominal.
- Nominal data is the least complex and give a simple measure of whether objects are the same or different.
- Ordinal data maintains the principles of nominal data but adds a measure of order to what is being observed.
- Interval data builds on ordinal by adding more information on the range between each observation by allowing us to measure the distance between objects.
- Ratio data adds to interval with including an absolute zero.

Categorical Data (Qualitative data)

- The objects being studied are grouped into categories based on some qualitative trait.
- The resulting data are merely labels or categories.
- Nominal and Ordinal scales will be used for categorical data or qualitative data.



Examples of Nominal Data:

- Type of car:
  - Mercedes, BMW, Lexus, Toyota, etc.
- Ethnicity:
  - White British, Afro-Caribbean, Asian, Arab, Chinese, other, etc.
- Smoking status:
  - Smoker, non-smoker.

Examples of Ordinal Data:

- Grades in an exam: A+, A, B+, B, C+, C, D+, D, and fail.
- Degree of illness; none, mild, moderate, acute, chronic.
- Opinion of students about stat classes; Very unhappy, unhappy, neutral, happy, ecstatic!

Examples of Binary Data:

A type of categorical data in which there are only *two* categories. (yes or no, A or B and nothing between)

E.g.

Smoking status; smoker, non-smoker.
Attendance; present, absent.
Result of exam; pass, fail.
Status of student; undergraduate, postgraduate.

Examples of categorical (nominal & ordinal) data:

Eye color: (Nominal)

Blue, brown, black, green, etc.

Smoking status: (Nominal)

Smoker, non-smoker

Attitudes towards the death penalty: (Ordinal)

Strongly disagree, disagree, neutral, agree, strongly agree.

Nominal data (Binary) & Ordinal data:

What is your gender?

☐ Male ☐ Female

Did you enjoy the teaching session?

☐ Yes ☐ No

What is the level of satisfaction with the new curriculum at a medical school received?

☐ Very satisfied

☐ Somewhat satisfied

☐ Neutral

☐ Somewhat dissatisfied

☐ Very dissatisfied

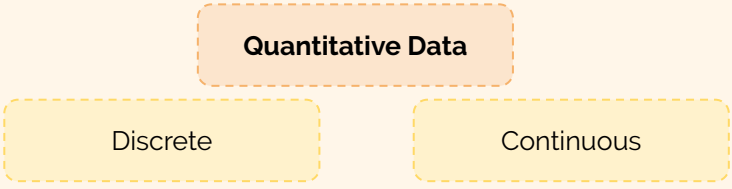
Quantitative Data

- The objects being studied are 'measured' based on some quantitative trait.
- The resulting data are a set of numbers.
- Interval & Ratio scales will be used to measure quantitative data.

Examples:

Pulse rate	Exam marks
Height	Time to complete a biostatistics exam
Age	Number of cigarettes smoked

What is difference between Discrete and Continuous Quantitative data?
Discrete: can take on only integer (target) values (counted data). For example: the number of students in a hall (you can't have half a student).
Continuous: can take on any value (measured data) For example: heights, weight..etc (you can have half data)

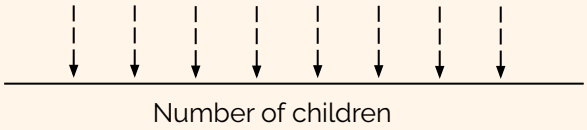


Discrete Data: (Whole numbers)
Only certain values are possible (there are gaps between the possible values). Implies counting.

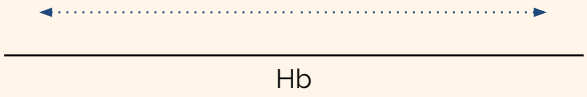
Continuous Data: (Decimal points)
Theoretically, with a fine enough measuring device. Implies measuring.

Quantitative Data (cont.)

- Discrete data - Gaps between possible values:



- Continuous data - Theoretically, no gaps between possible values



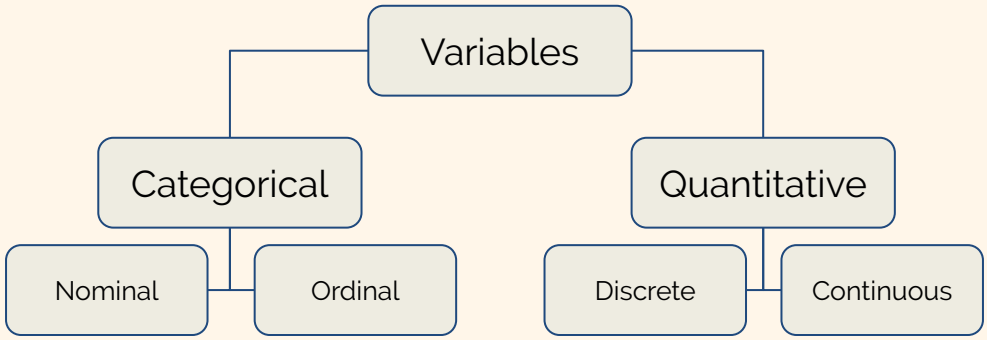
- **Examples of Discrete Data:**
  - Number of children in a family
  - Number of students passing a stats exam
  - Number of crimes reported to the police
  - Number of bicycles sold in a day

Generally, discrete data are counts. We wouldn't expect to find 2.2 children in a family, or 88.5 students passing an exam, or 127.2 crimes being reported to the police, or half a bicycle sold in a day.

- **Examples of Continuous Data:** (All clinical data examples)
  - Age (in years)
  - Height (in cm)
  - Weight (in kg)
  - Sys.BP, Hb, etc.

Generally, continuous data comes from measurements.

Relationships between variables:



CONTINUOUS DATA ⇢ QUALITATIVE DATA

Weight (kg): underweight, normal, overweight

Height (cm): short, medium, tall

Table 1 Distribution of blunt injured patients according to hospital length of stay:

hospital length of stay	Number	Percent
1 – 3 days	5891	43.3
4 – 7 days	3489	25.6
2 weeks	2449	18.0
3 weeks	813	6.0
1 month	417	3.1
More than 1 month	545	4.0
Total	**14604**	**100.0**
Mean = 7.85 SE = 0.10		

Clinimetrics

Clinimetrics

- A science called clinimetrics in which qualities are converted to meaningful quantities by using the scoring system. e.g Glasgow Coma Scale

Examples:

1. Apgar score based on appearance, pulse, grimace, activity, and respiration is used for neonatal prognosis
2. Smoking index: no. of cigarettes, duration, filter or not, whether it's a pipe, cigar, etc.
3. APACHE (Acute Physiology and Chronic Health Evaluation) score: to quantify the severity of a condition of a patient.

Data types - Important?

Why do we need to know what type of data we are dealing with?

The data type or level of measurement influences the type of statistical analysis techniques that can be used when analysing data.

To conclude

Type of variables in any data set are:

Categorical (qualitative)

Quantitative

Whereas the scales used to measure these two variables are:

Nominal, Ordinal, Interval, and Ratio scales.