BIOSTATISTICS

Prof Dr Najlaa Fawzi

STATISTICS

Statistics is the field of study concerned with

the collection, organization , presentation, and summarization of data, and the drawing of inferences about whole body of data when only a small part of the data is observed or examined or considered.

Types of Statistics

Descriptive Statistics and Inferential Statistics

Descriptive Statistics - methods of organizing, summarizing, and presenting data in an informative way.

Inferential Statistics: A decision, estimate, prediction, or generalization about a population, based on a sample.

Why Study Statistics?

1-Carry out your own research.

2-Evaluating published papers.

3-Ethical consideration (e.g. through statistics you can compare between old & new drugs choosing the most appropriate).

4-Professional & personal satisfaction (to value a result of a particular work if it is a good or not).

Data : are the building blocks of statistics , and refer to individual values presented , measured

of values.

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 that are to be collected and or analyzed

Value : it is the numerical representative of the measurement of the variable.

Data (singular): The value of the variable associated with one element of a population or sample. This value may be a number, a word, or a symbol.

Data (plural): The set of values collected for the variable from each of the elements belonging to the sample.

Sources of data

1. Routinely kept records

2- Experiment: The investigator controls or modifies the environment and observes the effect on the variable under study.

3-Survey: Data are obtained by sampling some of the population of interest. The investigator does not modify the environment.

4-Census: A 100% survey. Every element of the population is listed. Seldom used: difficult and time-consuming to compile, and expensive.

5. External source

Categories of data

1. Primary data: observation, questionnaire, record form, interviews, survey,

2. Secondary data: census, medical record, registry etc.

Routinely kept records

1-Registration of vital events

2-Notification of diseases.

3- Hospital records

4- Disease registers

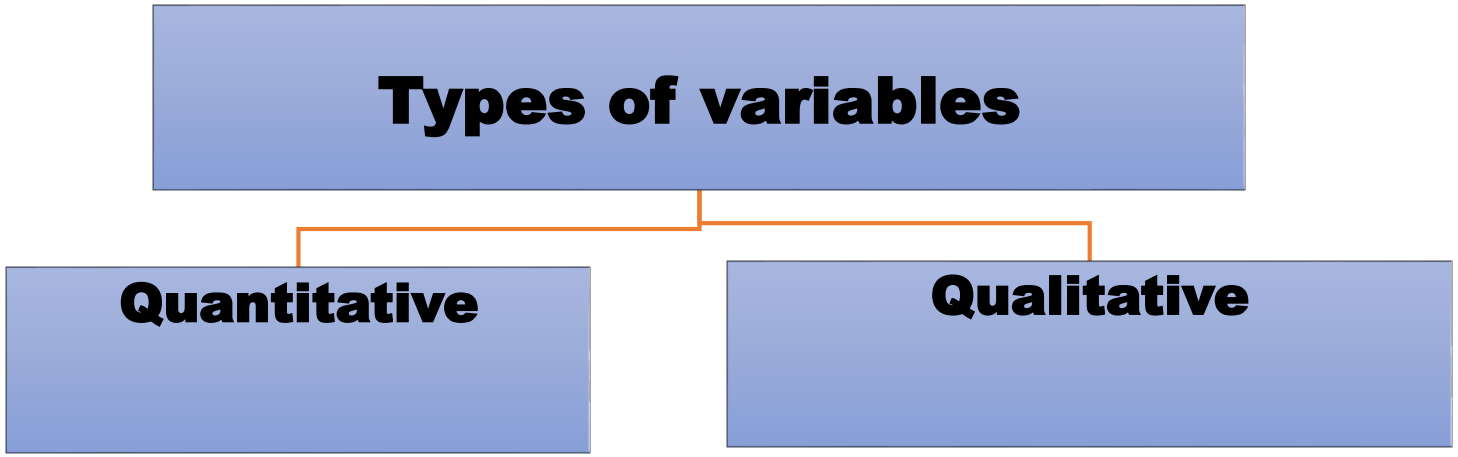
5-Epidemiological surveillance

VARIABLE:

it is a characteristic that take different values in different persons, places, times, or occasions,

anything that varies within a set of data

e.g.name , age, height, blood urea, weight, type of car ,etc.


Types of variables
Quantitative
Qualitative

1- Quantitative: that can be measured in the usual Sense. Is a variable that can be measured directly

For example:

- **survival time**
- **systolic blood pressure**
- **number of children in a family**
- **height**
- **age**
- **body mass index**

a-Discrete variable

Discrete variables have a set of possible values that is either finite or countably infinite

number of pregnancies
shoe size
number of missing teeth

For a discrete variable there are gaps between its possible values
Discrete values often take integer whole numbers values

Continuous variable : Is a variable that has large no. of possible values & several in-between values

- **Weight**
- **Height**
- **Mid-arm circumference**
- **Blood sugar level**
- **ºC/ ºF temperature scale**
- **Body mass index (BMI)**
- **Hemoglobin level**
- **Serum cholesterol level**

Continuous not have gaps or interruption, e.g. S. cholesterol

weight).

Hb, glucose level. Variable has unit.

2-Qualitative: that which can't be measured in usual sense ,but we can describe them in categories &there is no numerical value.

e.g. eye colors ,type of anemia, blood group ,socio-economic status

Note: Arithmetic operations, such as addition and averaging, are not meaningful for data resulting from a qualitative variable.

Summary of Types of Variables

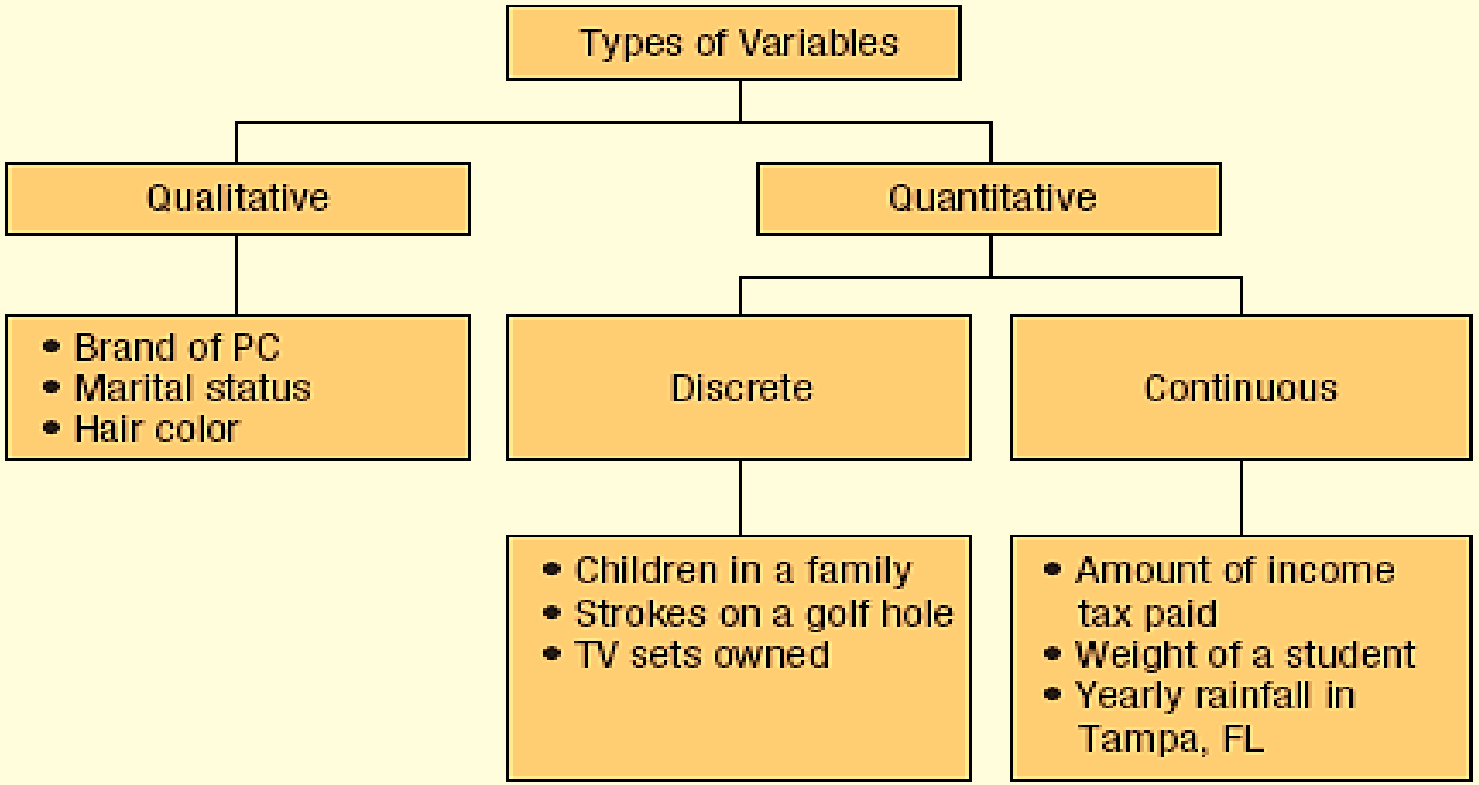
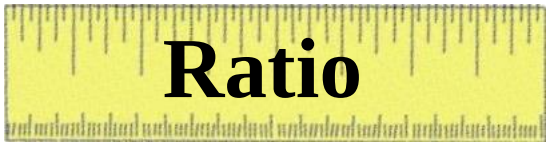


CHART 1–2 Summary of the Types of Variables

Measurements and Measurement Scales

Measurement : this may be defined as the assignment of numbers

Four Levels of Measurement



1-The Nominal Scale: the lowest measurement scale is the nominal scale,as the name indicates it consists of "naming" observations or classifying them into mutually exclusive and collectively categories .

The practice of using numbers to distinguish among the various medical diagnosis , male – female , married – not married.

Categorical measurements which contain only two classes (sex) are termed [dichotomous].

Outcomes may also be described with more than two categories, such as the classification of anemia.

Although we talk about nominal data as being on the measurement scale, we do not actually measure nominal data; instead, we count the number of observations with or without the attribute of interest.

Nominal scale data Examples: **Country of Birth, Eye Color, Race, Blood group**

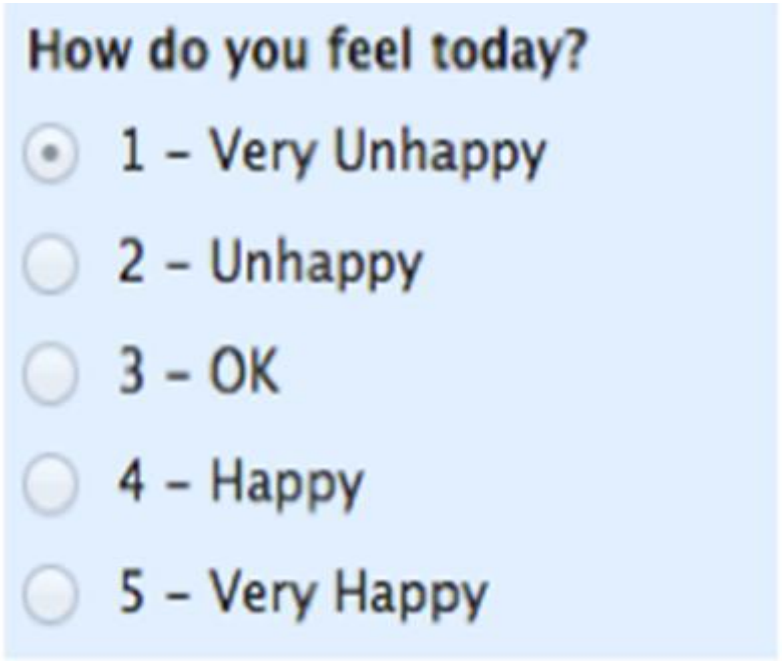
2-Ordinal Scale allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, but still they do not allow us to say "how much more." A typical example of an ordinal variable is the socioeconomic status of families.

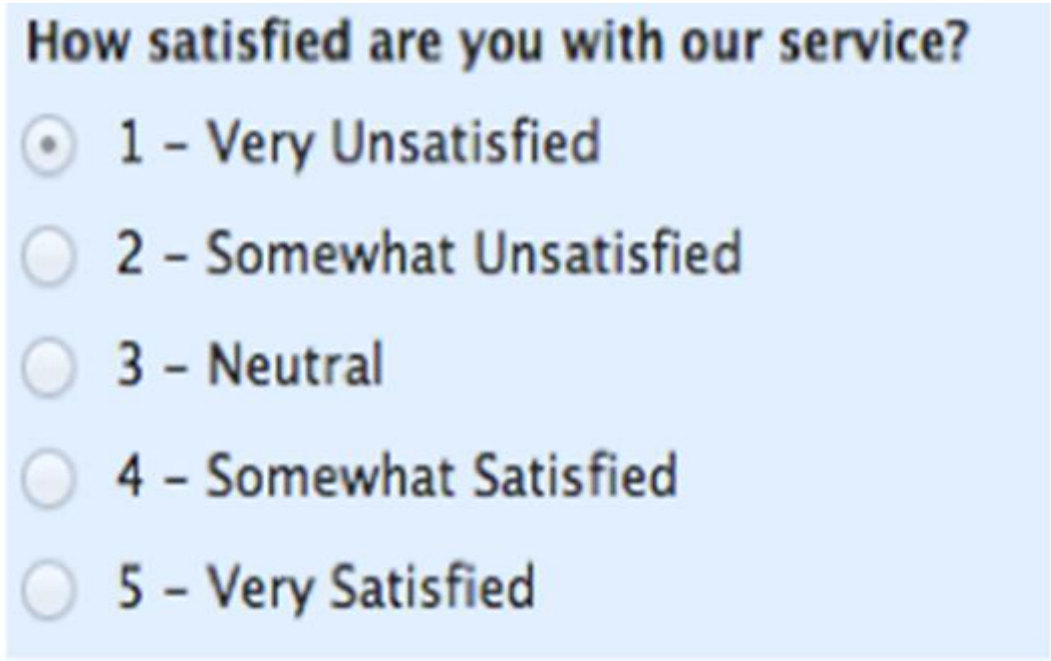
Involves data arranged in some order, but the differences between data values cannot be determined OR are meaningless!

Ordinal Scale data Examples

- **Pain level (Mild, Moderate, Severe)**

- **Tumors (Stage 0, ……, IV)**
- **Arthritis (Class 1, ……, 4)**
- **Military Rank (Lt., Capt., Maj., Col., General)**





There is no information about the size of the interval—no conclusion can be drawn about whether the difference between the first and second students is the same as the difference between the second and third.

Interval scale

Data are like ordinal data, in that they can be placed in a meaningful order.
In addition, they have meaningful intervals between items, which are usually measured quantities. For example, on the Celsius scale, the difference between 100° and 90° is the same as the difference
between 50° and 40°.

However, because interval scales do not have an absolute zero, ratios of scores are not meaningful: 100°C is not twice as hot as 50°C because 0°C does not indicate a complete absence of heat.

Interval scale each measurement is assigned to

no true zero point, it does not begin from zero

due to the presence of minus numbers,

e.g. temperature.

With this scale ,it is not only possible to order measurement , but also the distance between any two measurements is known , the difference between measurement of 20 & measurement of 30 is equal to the difference between measurement of 30 &40.

The interval scale unlike the nominal and ordinal scales is a truly quantitative scale.

Example: Women's dress sizes listed on the table.

Size	Bust (in)	Waist (in)	Hips (in)
8	32	24	35
10	34	26	37
12	36	28	39
14	38	30	41
16	40	32	43
18	42	34	45
20	44	36	47
22	46	38	49
24	48	40	51
26	50	42	53
28	52	44	55

Class interval	Frequency
30 – 39	11
40 – 49	46
50 – 59	70
60 – 69	45
70 – 79	16
80 – 89	1
Total	189

Ratio scale

The highest level of measurement is the ratio

Data have the same properties as interval scale data; however, because there is an absolute zero, meaningful ratios do exist.
Most biomedical variables form a ratio scale: weight in grams or pounds, time in seconds or days, blood pressure in millimeters of mercury, and pulse rate in beats per minute are all ratio scale data.

Zero pulse rate indicates an absolute lack of heartbeat.
Therefore, it is correct to say that a pulse rate of 120 beats/min is
twice as fast as a pulse rate of 60 beats/min

Ratio scale (Presence of absolute zero; thus ratios are possible : Examples: Weight, Height, Blood glucose, Hemoglobin level, Serum cholesterol, Mid-arm circumference, Blood pressure, Pulse rate.

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

Often the distinction between interval and ratio scales can be ignored in statistical analyses. Distinction between these two types and ordinal and nominal are more important. WHY ???

Likert Scale

- Is also known as ‘Summative scale
- Is a ‘type of Ordinal scale
- Is generally used to quantify attitudes and behaviour
- ‘Responses are graded on a continuum’ (For example: Strongly agree – Agree – Neutral – Disagree – Strongly disagree)
- No. of responses are usually 3, 5 or 7
- Likert scale is ‘usually a bipolar scaling’ method: It measures positive or negative response to a statement

Independent vs Dependent Variables

The two main variables in an experiment are the independent and dependent variable.

An independent variable is the variable that is changed or controlled in a scientific experiment

A dependent variable is the variable being tested and measured in a scientific experiment.

The independent and dependent variables may be viewed in terms of cause and effect. If the independent variable is changed, then an effect is seen in the dependent variable.

Remember, the values of both variables may change in an experiment and are recorded.

The difference is that the value of the independent variable is controlled by the experimenter, while the value of the dependent variable only changes in response to the independent variable.

Independent Variables	Dependent Variables
Independent variables are variables that determine the value of the variables.	Dependent variables are variables that get determined from independent variables.
Independent variables are considered to be experiment controller these can be manipulated.	Dependent variables are experiment measure and are very difficult to manipulate.
Independent variables takes the form of experiment stimulus having two attributes which is either present ot absent.	Dependent variables have attributes which are direct, indirect or through constructs.
Independent variables are variables that can be termed as casual variables.	Dependent variables are considered as the caused variables.