Editing file

Feedback

# Basic concepts and terminology in biostatistics

## Objectives:

1. Definition of statistics and biostatistics and its application.
2. To understand different levels of measurements.
3. To understand different types of data.
4. To use these concepts appropriately.

21st lecture

## Color Index:

- Boys' Slides
- Girls' Slides
- Doctors Notes
- Golden Notes
- Important
- Extra

# Statistics

Statistics is the science of conducting studies to collect, organize, summarize, analyze, present, interpret and draw conclusions from data.

**Data:** Any values (observations or measurements) that have been collected.
Singular: datum

## What is statistics?

| | | |
|---|---|---|
| 1 | Collecting Data | e.g., Sample, Survey, Observe, Simulate<br>Depends on the study design |
| 2 | Characterizing Data | e.g., Organize Classify, Count, Summarize |
| 3 | Presenting Data | e.g., Tables, Charts, Statements |
| 4 | Interpreting Results | e.g., Infer, Conclude, Specify Confidence |

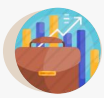
Data Analysis

Why?

Decision-Making

## Areas where statistics are used

Business, Economics, Marketing

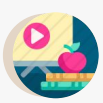


Engineering, Mathematics

Computer Science

Planning, National Defense

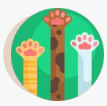
Astronomy



Physical Sciences, Education

Government Census

Public Health & Medicine
Epidemiology, Pharmacology, Genetics



Animal Populations

Environment, Agriculture, Ecology, Forestry

# Biostatistics

| Biostatistics | is the science that **helps** in managing medical **uncertainties** and **variability** of data |
|---|---|

## Biostatistics

- 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.

**Biostatistics**

### Medical Statistics aka: clinical statistics -same as biostatistics-

Deals with application of statistical methods to the study of diseases (risk factors, prognostic factors, etc..), efficacy of new treatments or vaccine, etc..

### Health Statistics - methodology is the same as medical statistics

Deals with application of statistical methods to varied information of public health importance.

### Vital Statistics

Is the ongoing collection of government agencies of data relating to vital event such as births and deaths which are deemed reportables by local health authorities.

## Basic Concepts

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

### Sources of data

- Routinely kept records
- Surveys (census) population based
- Experiments
- External source

### Categories of data

**Primary data:**
What principle investigator collect (new data)
observation, questionnaire, record form, interviews, survey

**Secondary data:**
Collected from other source which data already is collected
census, medical record, registry

# Variables

## Datasets and Data Tables

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

| 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 |

**Table 1**: 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
0- 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

# Variables

## Types of variables and 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.
- 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 measurement:

1 Nominal scale variables

2 Ordinal scale variables

3 Interval scale variables

4 Ratio scale variables

### Nominal scale variables

- A type of categorical data in which objects fall into **unordered** categories.
- Studies measuring nominal data must ensure that each category is **mutually exclusive** (no overlap like Male / Female) and the system of measurement needs to be exhaustive.
- Variables that have only two responses i.e. Yes or no, are known as **dichotomies**. (Binary) but also can have more than 2 options like (blood group, name of cities or name of cars...etc)

### Ordinal scale variables

- 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.
(Low stress - Moderate stress - Severe Stress)

### Interval scale variables

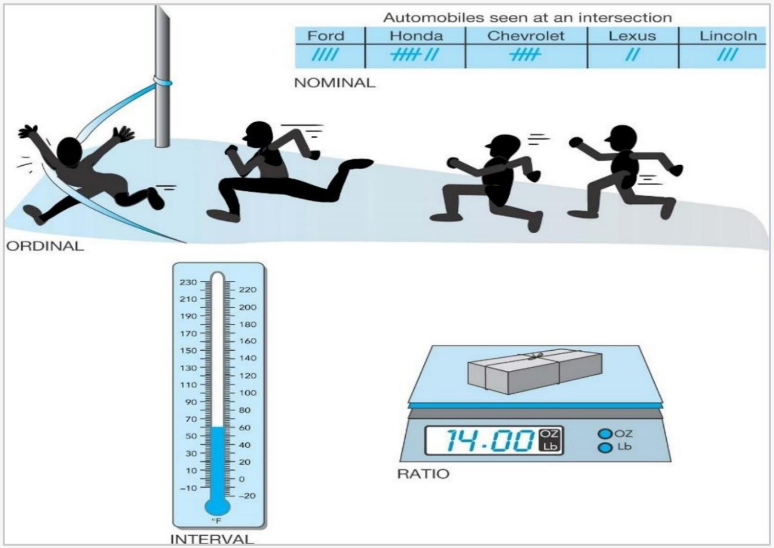
- **Fahrenheit temperature scale:** (zero is arbitrary) -40 degrees is not twice as hot as 20 degrees. You can't compare since no zero reference (there is negative).
- **IQ tests:** No such thing as zero IQ.
120 IQ not twice as intelligent as 60.
- Question- Can we assume that attitudinal data represents real, quantifiable measured categories?
(i.e.. That 'very happy' is twice as happy as plain 'happy' or that 'very unhappy' means no happiness at all). "Statisticians not in agreement on this". **NO!. You can't quantify it, negative values are included.**

### Ratio scale variables

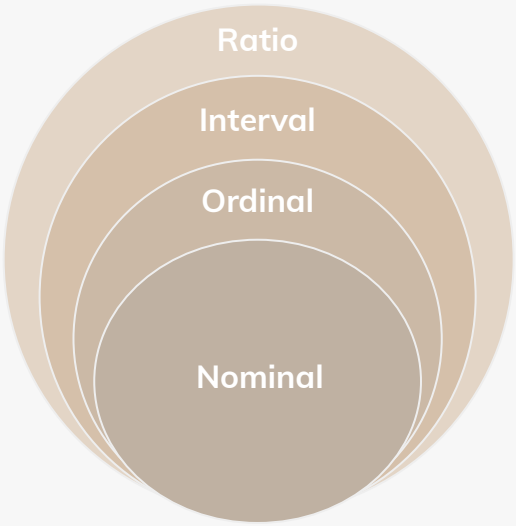
- The distance between any two adjacent units of measurement (intervals) is the same and there is a meaningful zero point.
- Income: someone earning SAR20,000 earns twice as much as someone who earns SAR10,000.
- Height
- Weight
- Age
- **Negative values are not included**
- **It can be compared (best scale for comparison)**

# Variables

## Scales of Measurement




## Hierarchical data order



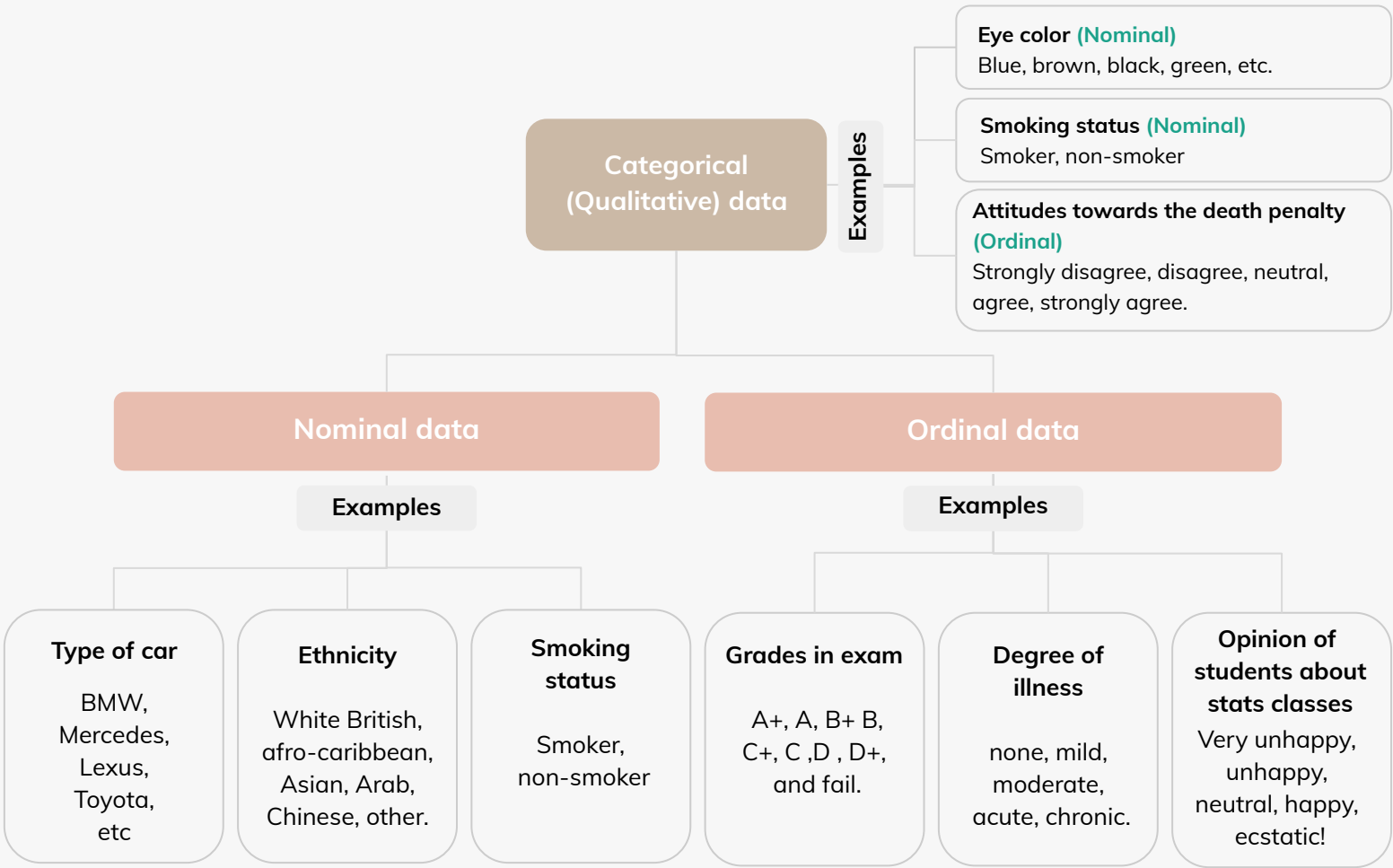

**These levels of measurement can be placed in hierarchical order.**

- 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. No absolute zero value.
- Ratio data adds to interval with including an absolute zero.

# Categorical Data

## 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.




## Binary Data (dichotomies)

A type of categorical data in which there are only **two categories.**

Examples:

- Smoking status- smoker, non-smoker, past smoker
- Attendance: present, absent
- Result of a exam: pass, fail
- Status of student: undergraduate, postgraduate

Dr: it's important in the exam to know that it can be more than two categories "blood groups as example"

# Quantitative Data

## Nominal data (Binary) & Ordinal data (Examples)

**What is your Gender?** (please tick)

- ☐ Male
- ☐ Female

**Did you enjoy the session?** (please tick)

- ☐ Yes
- ☐ No

**What is the level of satisfaction with the new curriculum at a medical school received?** (please tick)

- ☐ Very satisfied
- ☐ Somewhat satisfied
- ☐ Neutral
- ☐ Somewhat dissatisfied
- ☐ Very dissatisfied

## Quantitative Data

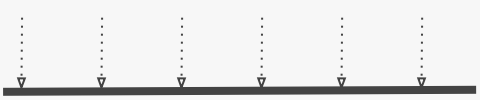
- The objects being studied are **'measured'** based on some **quantitative** trait.
- The resulting data are set of numbers.
- **Interval and Ratio** scales will be used to measure quantitative data.

**Quantitative data**

**Discrete**

Only certain values are possible (there are gaps between the possible values).
Implies counting.
A whole number.
For example:
- Number of family members,
- Number of smoked cigarette

Gaps between possible values

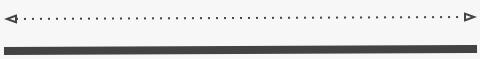


Number of Children

**Continuous**

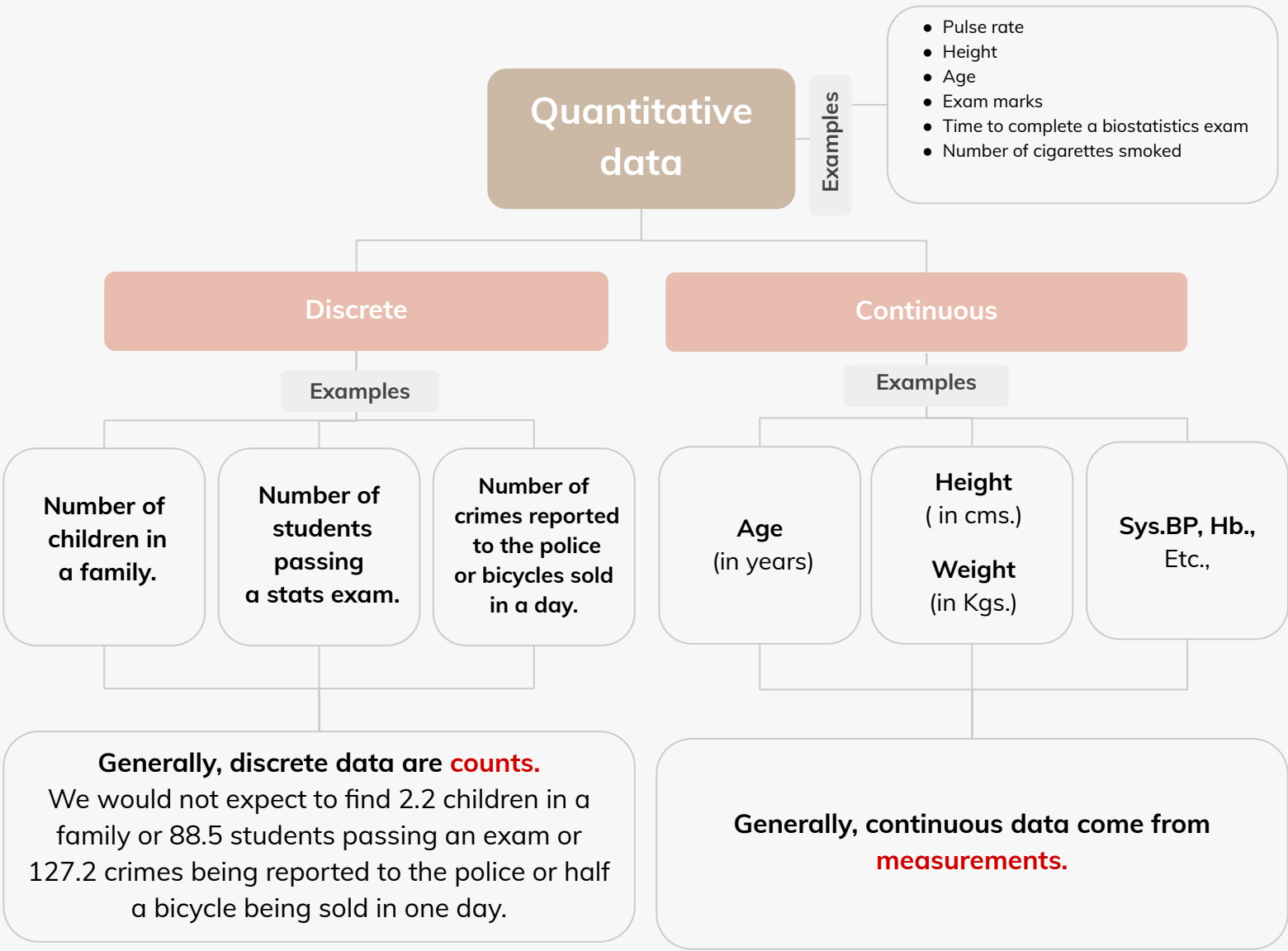
Theoretically, with a fine enough measuring device.
Implies measuring.
There is a decimale.

Theoretically,
no gaps between possible values

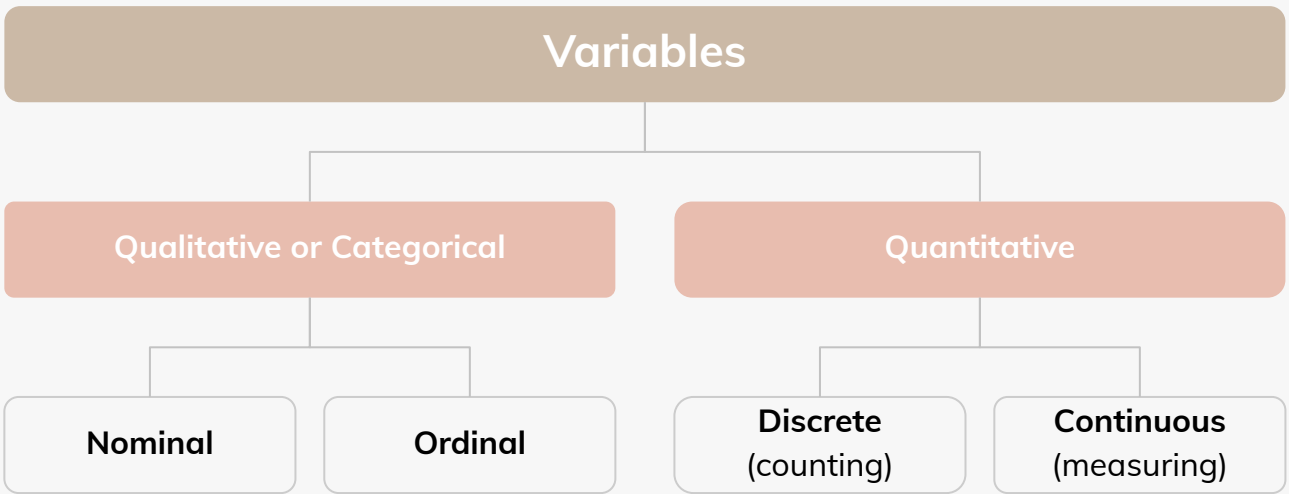


Hb

# Quantitative data




## Relationships between variables



# Relationships between variables

| Quantitative Variables | Qualitative variables |
|---|---|
| Height (cm / feet) (Continuous) | Short / Medium / Tall (ordinal) |
| Weight (kg / pound) (Continuous) | Underweight / Normal weight / Overweight (ordinal) |
| Blood sugar (mg %) (Continuous) | Diabetic / Non-Diabetic (nominal) |
| Blood pressure (mm) (Continuous) | Normal blood pressure / Hypertension (nominal, if more options then it will become ordinal) |
| Hemoglobin (mg %) (Continuous) | Anaemic / Non-Anaemic (nominal) |

**CONTINUOUS DATA** → **QUALITATIVE DATA**

Wt. (In kg.) → Under wt, normal & over wt. (ordinal)

Ht. (In cm.) → Short, medium & tall. (ordinal)

| 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** | | |

**Table 2**: Distribution of blunt injured patients according to hospital length of stay.

Sometimes continues data can be converted into qualitative data.
for example: Stress score into (Low stress - Moderate stress - Severe stress)

# Clinimetrics and Data types

## Clinimetrics

A science called clinimetrics in which qualities are converted to meaningful quantities by using the scoring system. (Categorical data converted into quantitative data)

Examples:

1. **Apgar score** based on **a**ppearance, **p**ulse, **g**rimace, **a**ctivity and **r**espiration is used for neonatal prognosis. Appearance: Skin color (categorical) into the score .
2. **Smoking index**: no. of cigarettes, duration, filter or not, whether pipe, cigar etc.
3. **APACHE** (**A**cute **P**hysiology **a**nd **C**hronic **H**ealth **E**valuation) score: to quantify the severity of 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:** | **The scales to measure these two variables are:** |
| Categorical (Qualitative) & Quantitative | Nominal, Ordinal, Interval and Ratio scales |

# Summary

## Variables

### Nominal

- Unordered (not ranked)
- Mutually exclusive: ex: female / male
- Exhaustive system of measurement.
- Two responses variables: ex: Yes/ no

### Ordinal

- Can be ranked ordered
- Distance between each category cannot be calculated

### Interval

- Zero is arbitrary
- Fahrenheit temp
- IQ test

### Ratio

Has a meaningful zero point

ex: Height
Weight
Age

## Categorical (Qualitative)

The objects being studied are grouped into categories based on some **qualitative** trait.

Nominal & Ordinal

### Nominal

- Type of car
- Ethnicity
- Smoking status

### Ordinal

- Grades in exam
- Degree of illness

## Quantitative

The objects being studied are 'measured' based on some **quantitative** trait.

Discrete & Continuous

### Discrete

There are gaps between the possible values.

Implies counting

Example:

- Number of children in the family
- number of students passing a test
- Number of crimes reported

### Continuous

No gaps between possible values

Implies measuring

Example:

- Age
- Height
- Weight

## Binary Data

A type of categorical data in which there are only **two categories.**

- Smoking status
- Attendance
- Result of exam
- Status of student

# Questions

**(1)** Which one of the following types of variables include absolute zero?

A) Nominal data

B) Ordinal data

C) Interval data

D) Ratio data

**(2)** Out of those variables, which can be considered as categorical ordinal data?

A) Smoking status

B) Grades in exam

C) Types of car

D) Height

**(3)** Which one of these values can be considered as discrete value?

A) 2.2 children in the family

B) 88 cm height

C) 14 students passed

D) 12.7 crimes reported

**(4)** Which one of the following types of variables has mutually exclusive values?

A) Nominal data

B) Ordinal data

C) Interval data

D) Ratio data

**(5)** The levels of measurement can be placed in hierarchical order which is?

A) Ratio > ordinal > interval > nominal

B) Ratio > interval > ordinal > nominal

C) Interval > ordinal > nominal > ratio

D) Ordinal > interval > nominal > ratio

**(6)** Out of those variables, which can be considered as categorical nominal data?

A) Sex

B) Rate of satisfaction

C) Number of children in the family

D) Height in cm

Answers:
1: D| 2: B | 3: C | 4: A | 5: B | 6: A

# Questions

**(7)** What is the type of scale measurement for a variable “serum creatinine (mg/dl)**?**

A) Ratio scale

B) Interval scale

C) Nominal scale

D) Ordinal scale

**(8)** What is the type of variable measurement scale for cancer stages (I, II, III< IV)**?**

A) Qualitative variable, nominal scale

B) Qualitative variable, ordinal scale

C) Quantitative variable, interval scale

D) Quantitative variable, ratio scale

**(9)** The grades of geometry test were classified as A+, A, B+, B, C+, C, D+, D, F. What is the best scale to measure this type of data**?**

A) Interval

B) Numerical

C) Ordinal

D) Nominal

**(10)** Which is the type of scale measurement for systolic blood pressure**?**

A) Ordinal

B) Nominal

C) Ratio

D) Interval

**(11)** Which of the following deals with application of statistical methods to varied information of public health importance**?**

A) Medical statistics

B) Health statistics

C) Vital statistics

D) Safety statistics

**(12)** What is the type of variable and measurement scale for serum cholesterol**?**

A) Qualitative variable, nominal scale

B) Qualitative variable, ordinal scale

C) Quantitative variable, interval scale

D) Quantitative variable, ratio scale

Answers:
7: A | 8: B | 9: C | 10: C | 11: B | 12: D

Research Team

RESEARCH TEAM
438
Keep it simple, Keep it focused

# Thank you for checking our work!

**Leaders:**

Shuaa Khdary  Sarah AlQuwayz

Abdulrhman Alsuhaibany

**Member:**

Dana Naibulharam

**Note Taker:**

Noura Alshathri

**Contact us:**
Research4390@gmail.com