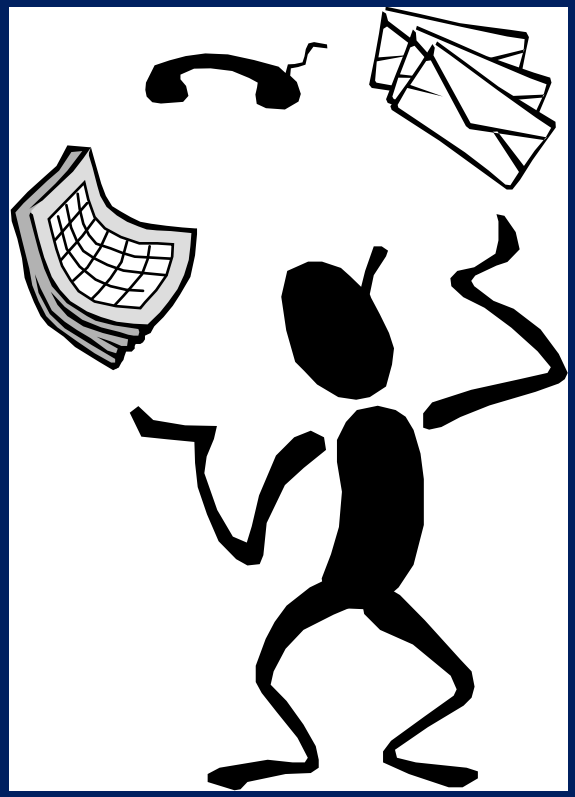
Basic concepts and terminology in Biostatistics

By
Dr. Shaik Shaffi Ahamed
Professor
Dept. of Family & Community Medicine
KKUH

Objectives of this session

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



Statistics is 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 is Statistics?

1. **Collecting Data**
   e.g., Sample, Survey, Observe, Simulate
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


Business, Economics, Engineering, Marketing, Computer Science
Mathematics,Physical Sciences, Astronomy,Education
Public Health & Medicine Epidemiology, Pharmacology, Genetics,,
Areas where STATISTICS are used
Environment, Agriculture, Ecology, Forestry, Animal Populations
Government Census, Planning, National Defense

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.

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 (census)
3. Experiments
4. External source

Categories of data

1. Primary data: observation, questionnaire, record form, interviews, survey,
2. Secondary data: census, medical record, registry

Datasets and Data Tables

Dataset: Data for a set of variables collection in 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

OBS	AGE	BMI	FFNUM	TEMP(^{0}F)	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

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

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:

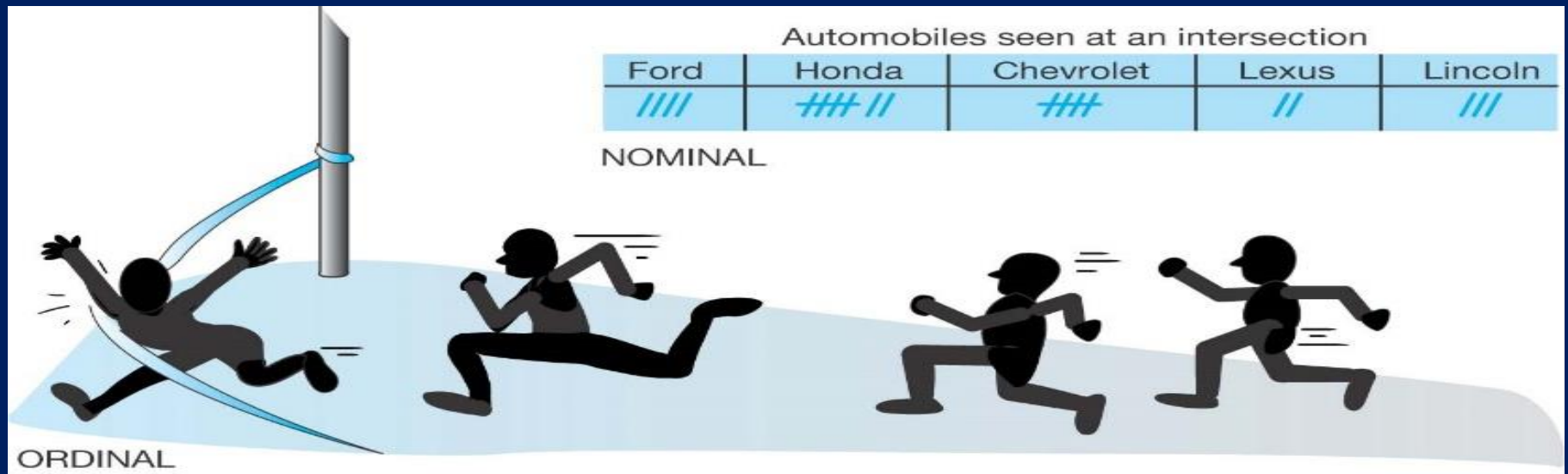
1. Nominal	**2. Ordinal**
3. Interval	**4. Ratio**

Scales of Measurement

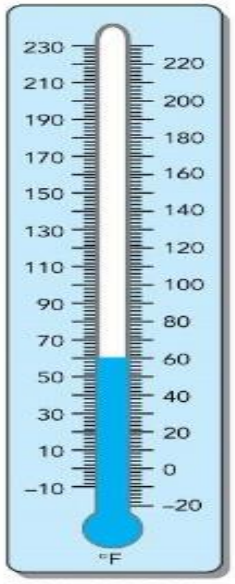
Automobiles seen at an intersection

Ford	Honda	Chevrolet	Lexus	Lincoln
////	###//	###	//	///

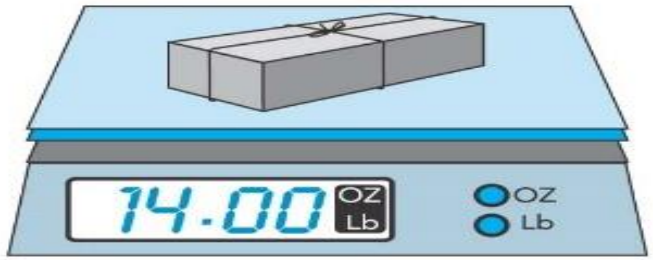
NOMINAL



ORDINAL



INTERVAL



RATIO

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 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

• 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

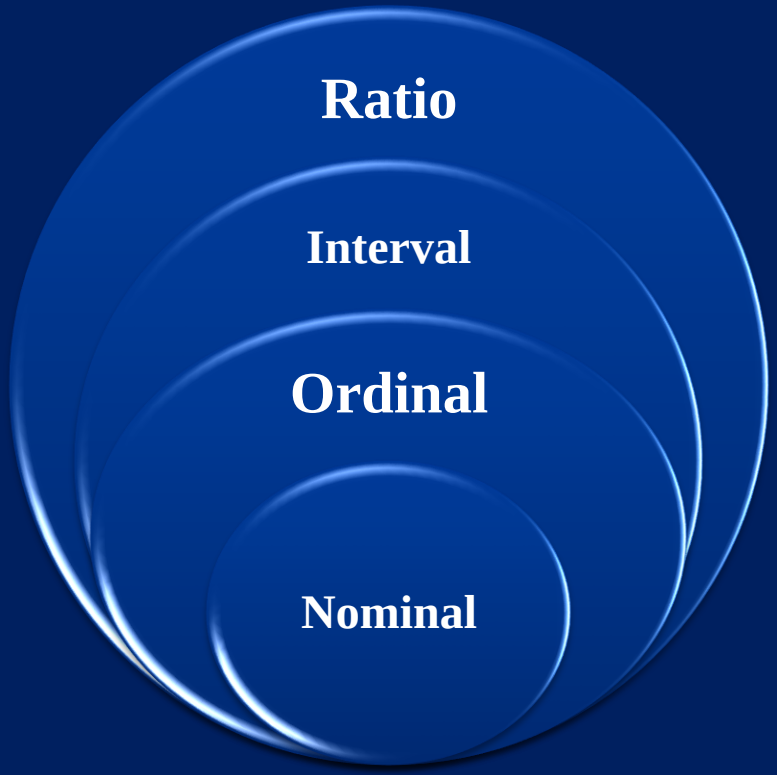
- Fahrenheit temperature scale- **zero is arbitrary**- 40 degrees is not twice as hot as 20 degrees.
- IQ tests. No such thing as zero IQ. 120 IQ not twice as intelligent as 60.
- Question- Can we assume that attitudinal data represents real, quantifiable measured categories? (i.e.. That '*very happy'*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

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

Hierarchical data order

These levels of measurement can be placed in hierarchical order.

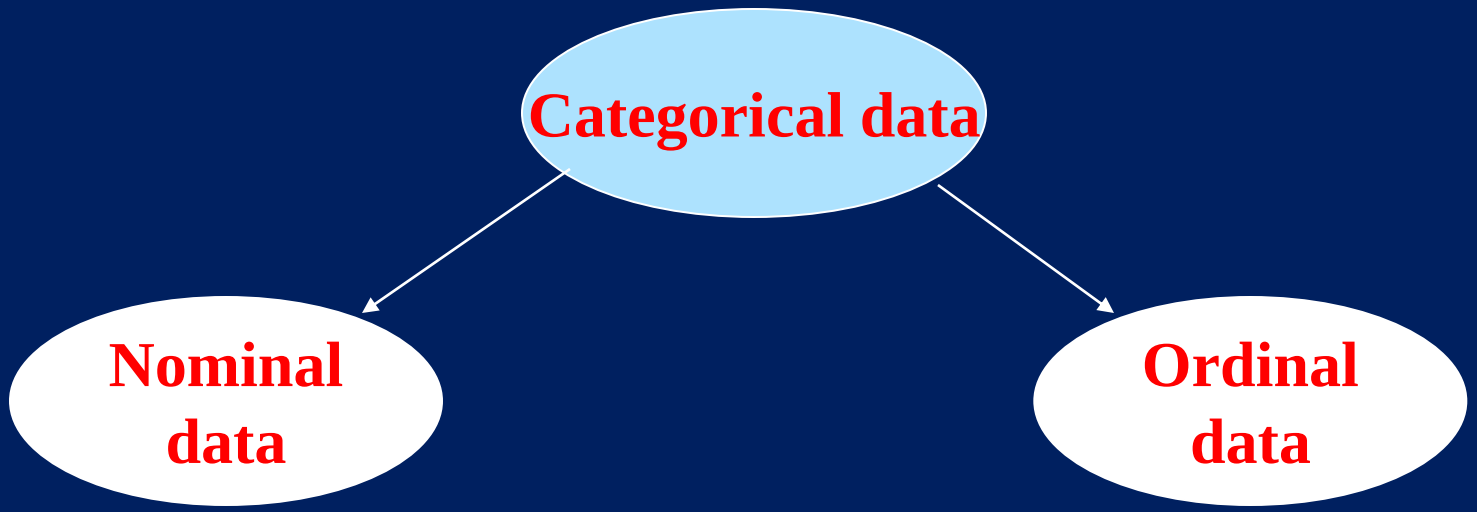


Hierarchical data order

- 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.
- Ratio data adds to interval with including an absolute zero.

Categorical 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.


Categorical data
Nominal data
Ordinal data

Examples of Nominal Data

- Type of car

  BMW, Mercedes, Lexus, Toyota, etc.,

- Ethnicity

  White British, afro-caribbean, Asian, Arab, Chinese, other, etc.

- Smoking status

  Smoker, non-smoker

Binary Data

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

Examples:

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

Examples of Ordinal Data

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

Examples of categorical (nominal & ordinal) data

- Eye color
  Blue, brown, black, green, etc.
- Smoking status
  Smoker, non-smoker
- Attitudes towards the death penalty
  Strongly disagree, disagree, neutral, agree, strongly agree.

Nominal data (Binary) & Ordinal data

Examples

What is your gender? *(please tick)*

Male ☐

Female ☐

Did you enjoy the teaching 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 ☐

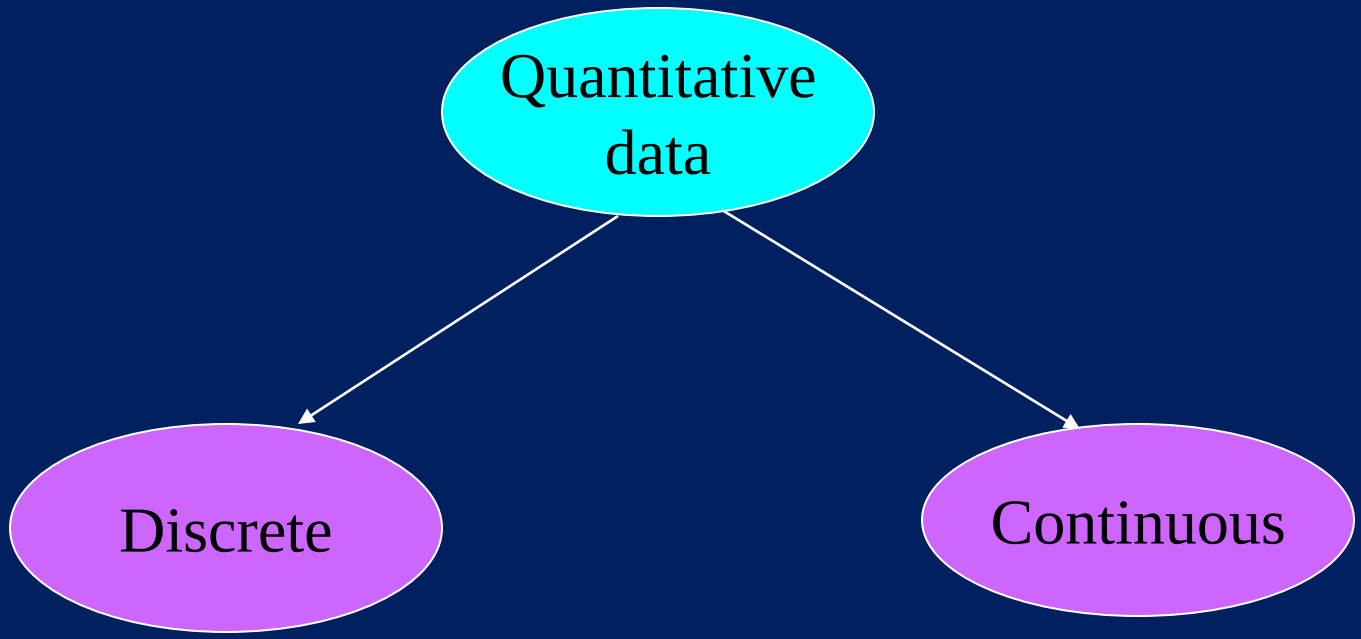
Very dissatisfied ☐

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.

Examples

- Pulse Rate
- Height
- Age
- Exam marks
- Time to complete a Bio-statistics exam
- Number of cigarettes smoked


Quantitative data
Discrete
Continuous

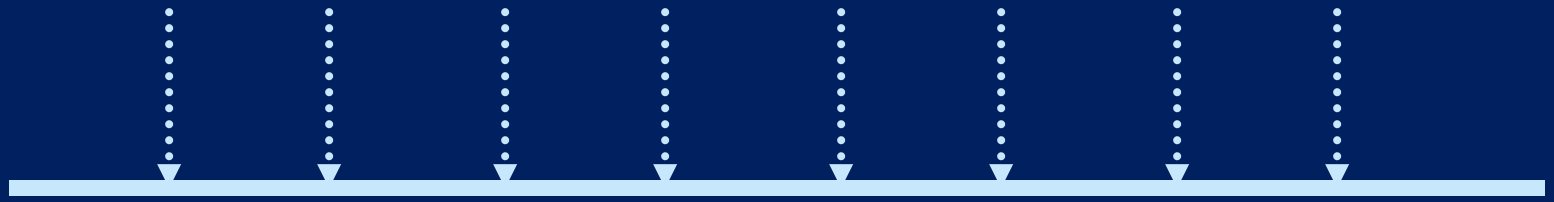
Discrete Data

Only certain values are possible (there are gaps between the possible values). Implies counting.

Continuous Data

Theoretically, with a fine enough measuring device. Implies measuring.

Discrete data -- Gaps between possible values



Number of Children

Continuous data -- *Theoretically*, no gaps between possible values

Hb

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.

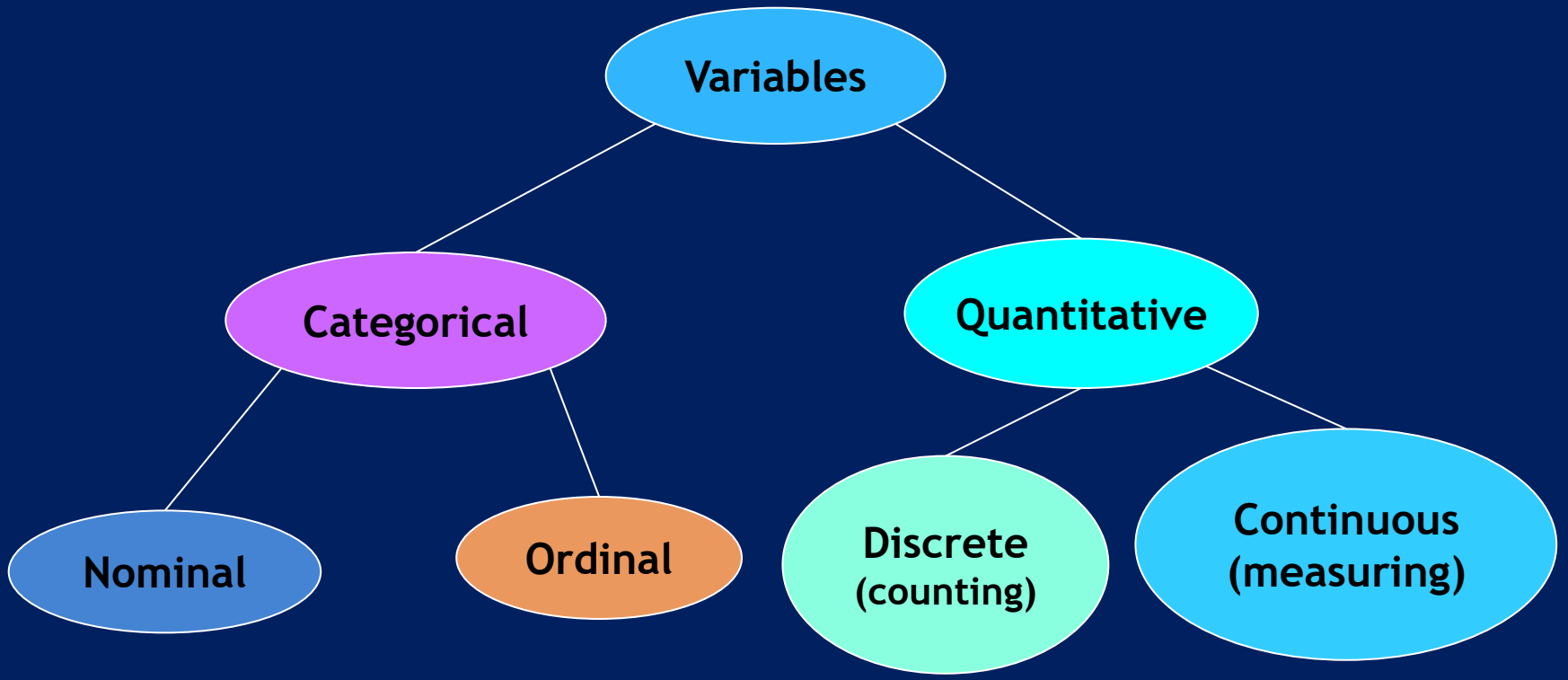
We would not expect to find 2.2 children in a family or 88.5 students passing an exam or 127.2 crimes being reported to the police or half a bicycle being sold in one day.

Example of Continuous Data

- Age (in years)
- Height(in cms.)
- Weight (in Kgs.)
- Sys.BP, Hb., Etc.,

Generally, continuous data come from measurements.

Relationships between Variables



CONTINUOUS DATA

QUALITATIVE DATA

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

Ht. (In 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

A science called clinimetrics in which qualities are converted to meaningful quantities by using the scoring system.

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 pipe, cigar etc.,

(3) APACHE (Acute Physiology and Chronic Health Evaluation) 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:
Categorical(Qualitative)
&
Quantitative
Whereas the scales to measure these two variables are:
Nominal, Ordinal, Interval and Ratio scales