



**WELCOME TO THE
UNIVERSITY
OF INSPIRED
LEARNING**

www.manipal.edu

Course Business Statistics

Segment-1 Descriptive Statistics

Faculty Dr. Tarushree Bari



Overview of the Course Curriculum

Segment 1: Descriptive Statistics

Types of Data , Visual Representation of Data ,Numerical Representation of Data

Segment 2: Probability and Uncertainty

Assigning Probabilities to Events, Conditional Probability
Rules of Probability and Bayes' Rule

Segment 3: Probability Distributions

Random Variables, Expected Values and Variance and Distribution



Overview of the Course Curriculum

Segment 4: Sampling and Estimation

Random Sampling Methods

Segment 5: Hypothesis Testing

Concepts in Hypothesis Testing

Segment 6: Regression Analysis

Simple Linear Regression, Multiple Regression



Overview of Topics

Different Types of Data

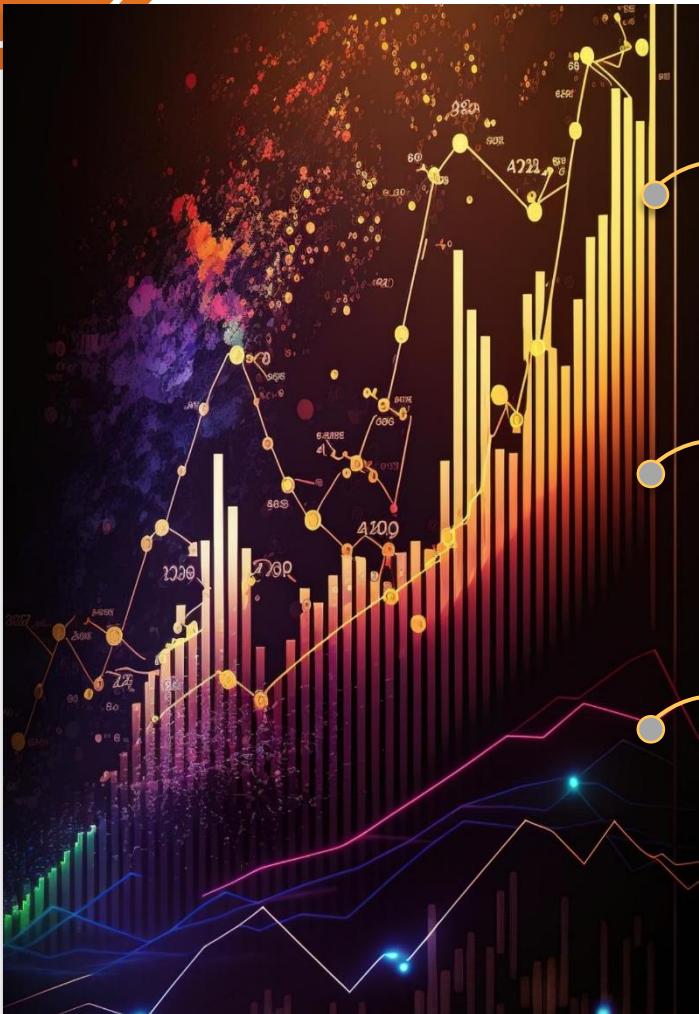
Measures of Central Location

Visual Representation of Data

Measures of Association



Descriptive Statistics



Statistics

Statistics is a discipline that involves the collection, analysis, interpretation, presentation, and organization of data

It provides methods and techniques for making decisions and drawing conclusions in the face of uncertainty.

It can often categorised into descriptive and inferential statistics.

describing
→ central value
→ variance
→ plots

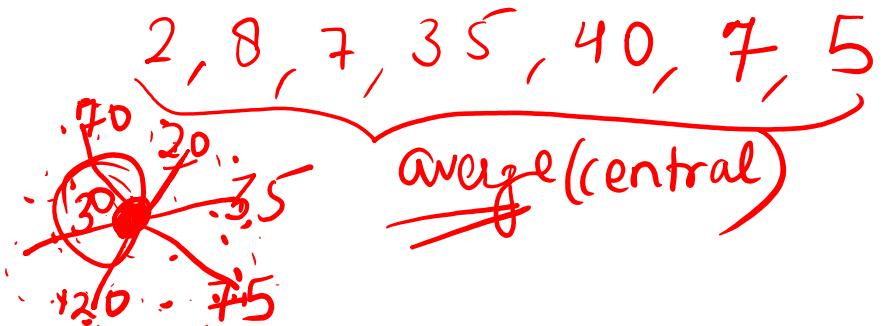
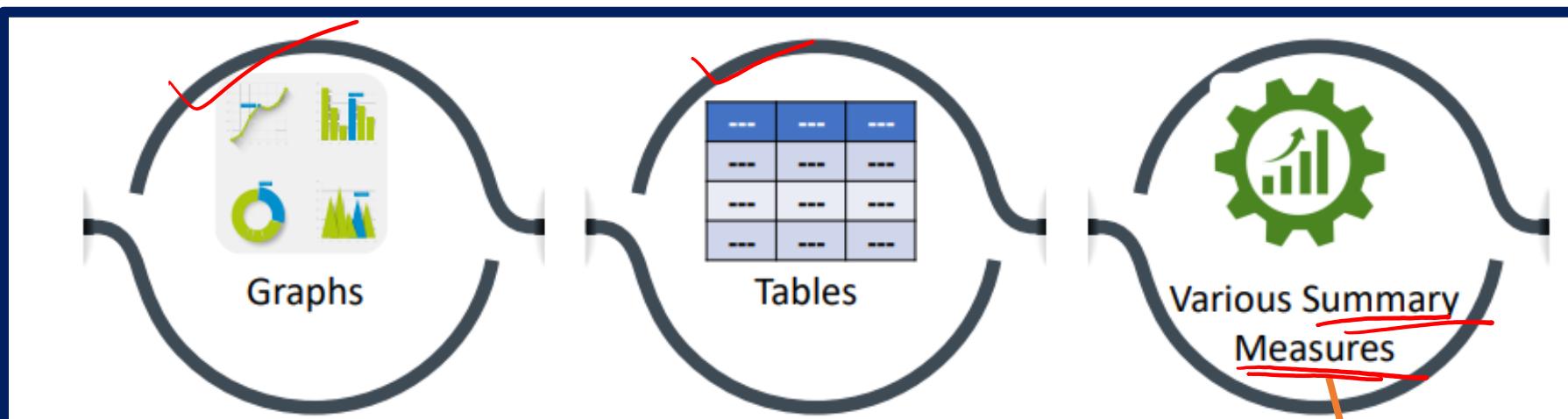
→ conclusion
→ apply test



Descriptive Statistics



Methods of describing characteristics of the data.



- ✓ Measures of central tendency
- ✓ Measures of variation

m → Central



Descriptive Statistics



Measures of central tendency

MEAN

$$\bar{x} = \frac{\sum x}{N}$$

- most stable measure
- affected by extreme values
- may not exist as a data point in the set

MEDIAN

- value that divides ranked data points into halves: 50% larger than it, 50% smaller
- may not exist as a data point in the set
- influenced by position of items, but not their values

MODE

- most frequent data point
- mode exists as a data point
- unaffected by extreme values
- useful for qualitative data
- may have more than 1 value

2 + ~~900~~ + 3 + 5 + 7 - 8 - 15 =
79

2, 3, 8, 4, 6

2, 3, ~~4~~, 6, 8

Median

2, 3, 4, 5, 7, 8

$$\frac{4+5}{2} = \text{Median}$$

20, 40, 30, 20, 20,
50, 20



Descriptive Statistics



Measures of central tendency

$$\begin{array}{r} \frac{20}{2^{\text{nd}}} \\ \frac{25}{2^{\text{nd}}} \\ \frac{480}{3} \\ \frac{30}{4} \\ \frac{35}{5} \\ - \\ - \end{array}$$

2022 1 April

Σ

Disadvantages

- Affected by extreme values in small data sets which may end up in erroneous representation of the data

Mean

Advantages

- Possesses some useful mathematical properties $\sum(x)$ sum
- Suitable for both small data sets (that do not have extreme values) and large data sets

Median

- Its not based on all the observations
- Affected by sample size fluctuations

Mode

- Very suitable for small data sets if there are extreme values
- No calculation beyond sorting and counting needed

- Not unique
- Limited applications

- Suitable for design applications
- No calculation needed



Measures of central tendency: Example:

Sales Performance Analysis

Patient satisfaction regarding waiting times



Descriptive Statistics



Popular ways to describe the position of a data point within a data set

1
2
3
4

5
6
7
8

9
10
11
12

13
14
15
16

17
18
19
20

Middle value

Q1

Q2

Q3

two equal half

Percentiles
~~100~~ Decided

Or

Quartiles
Equal

The n^{th} percentile of a data set is that value below which $n\%$ of the data lie.

Q1 – 25th

Q2 – 50th

Q3 – 75th

Median



Descriptive Statistics

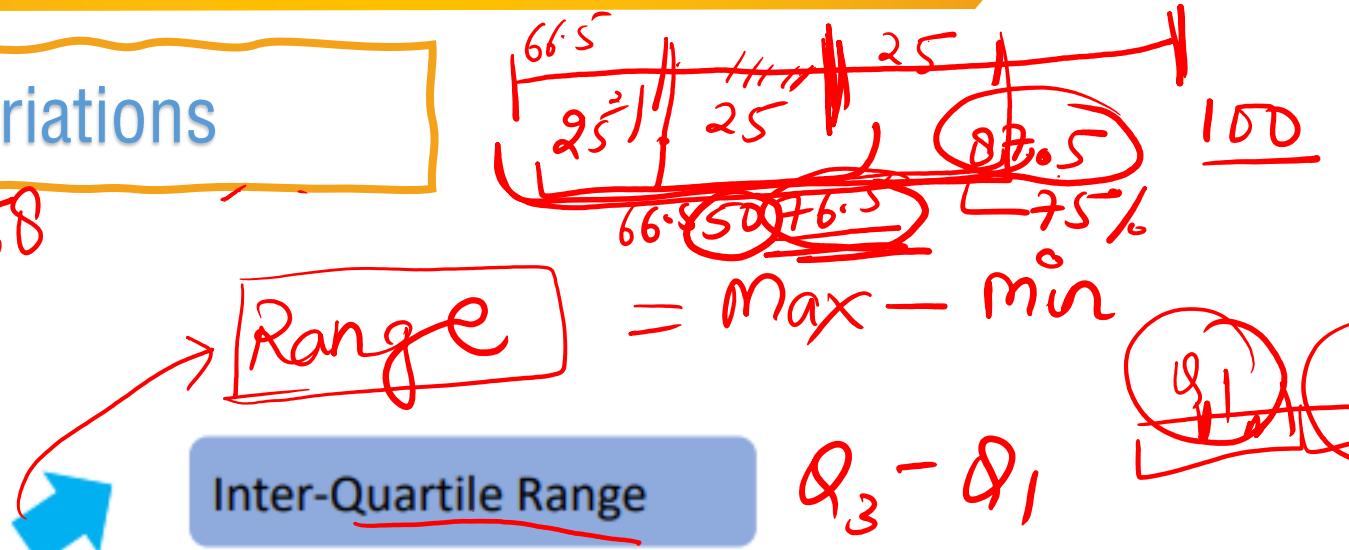
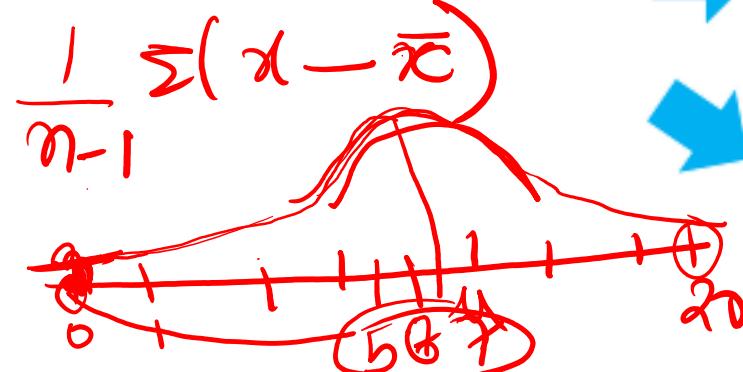


Measures of Variations



mean $\rightarrow (\bar{u})$

Measures of Variation



Inter-Quartile Range

Mean Absolute Deviation *rarely*

Variance σ^2

Standard Deviation

$$SD = \sqrt{\text{variance}}$$



$$Q_3 - Q_1$$

$$\frac{\sum |x - \bar{u}|}{n}$$

$$\frac{\sum (x - \bar{u})^2}{n} = \text{variance}$$

$$= \sqrt{\sigma^2} = \sigma = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

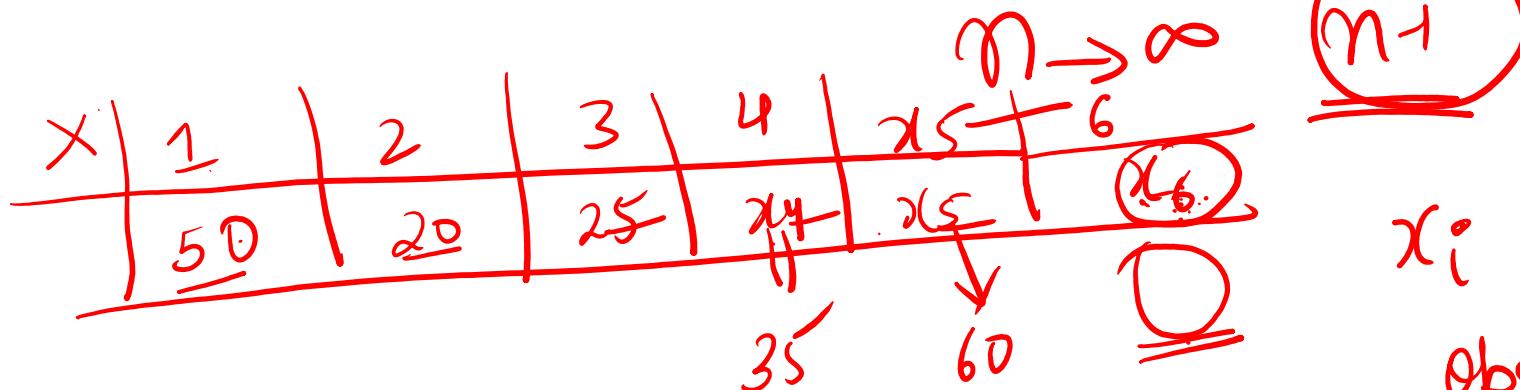
$$\text{Variance} = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{1}{n-1} \sum (x - \bar{x})^2$$

~~x_1~~ ~~x_2~~ ~~x_3~~ ~~x_4~~
 ~~x_5~~ ~~x_6~~ ~~x_7~~ ~~x_8~~
 ~~$\bar{x}(u)$~~ ~~n~~ ~~$n-1$~~

$$\frac{(3-2)^2 + (8-2)^2 + (4-2)^2 + (1-2)^2}{4}$$

$$\frac{\sum x_i}{n} = \bar{x}$$

$$= (1+6+2-1) \rightarrow \text{no use}$$



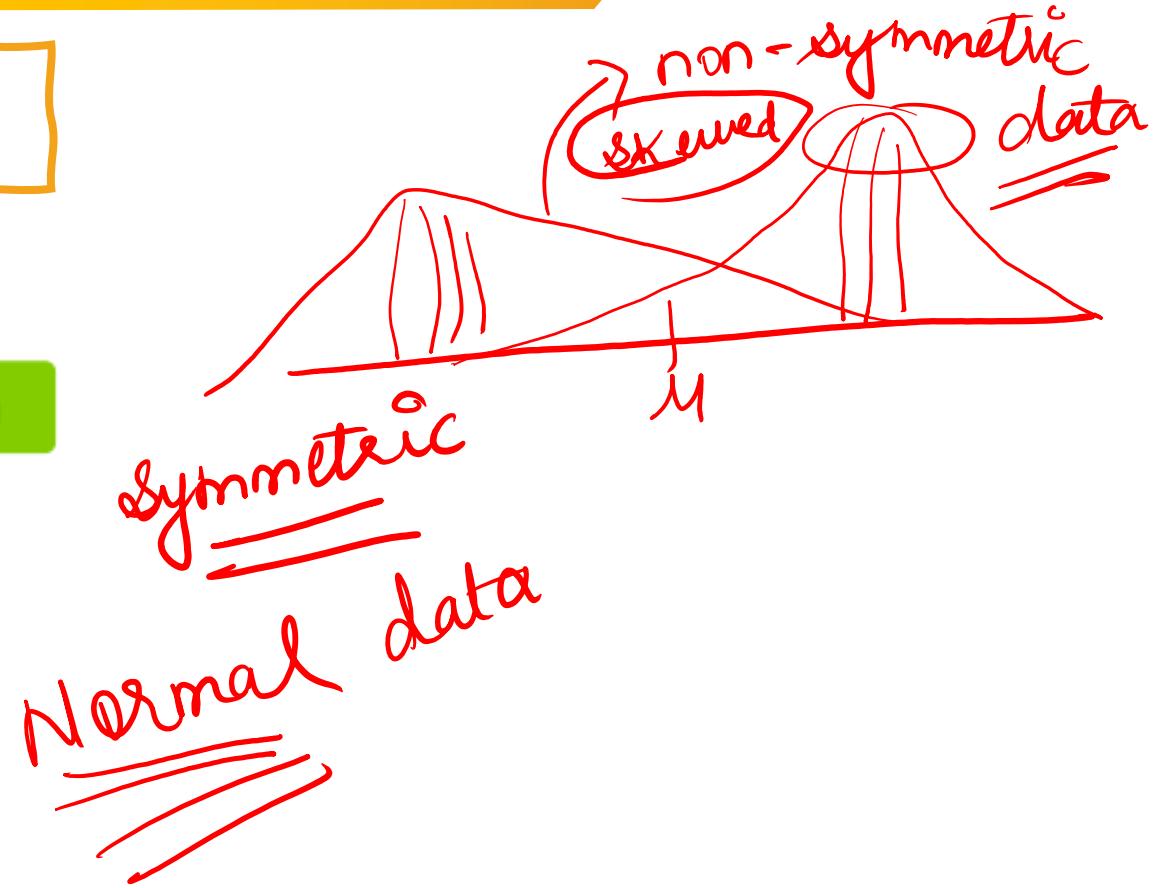
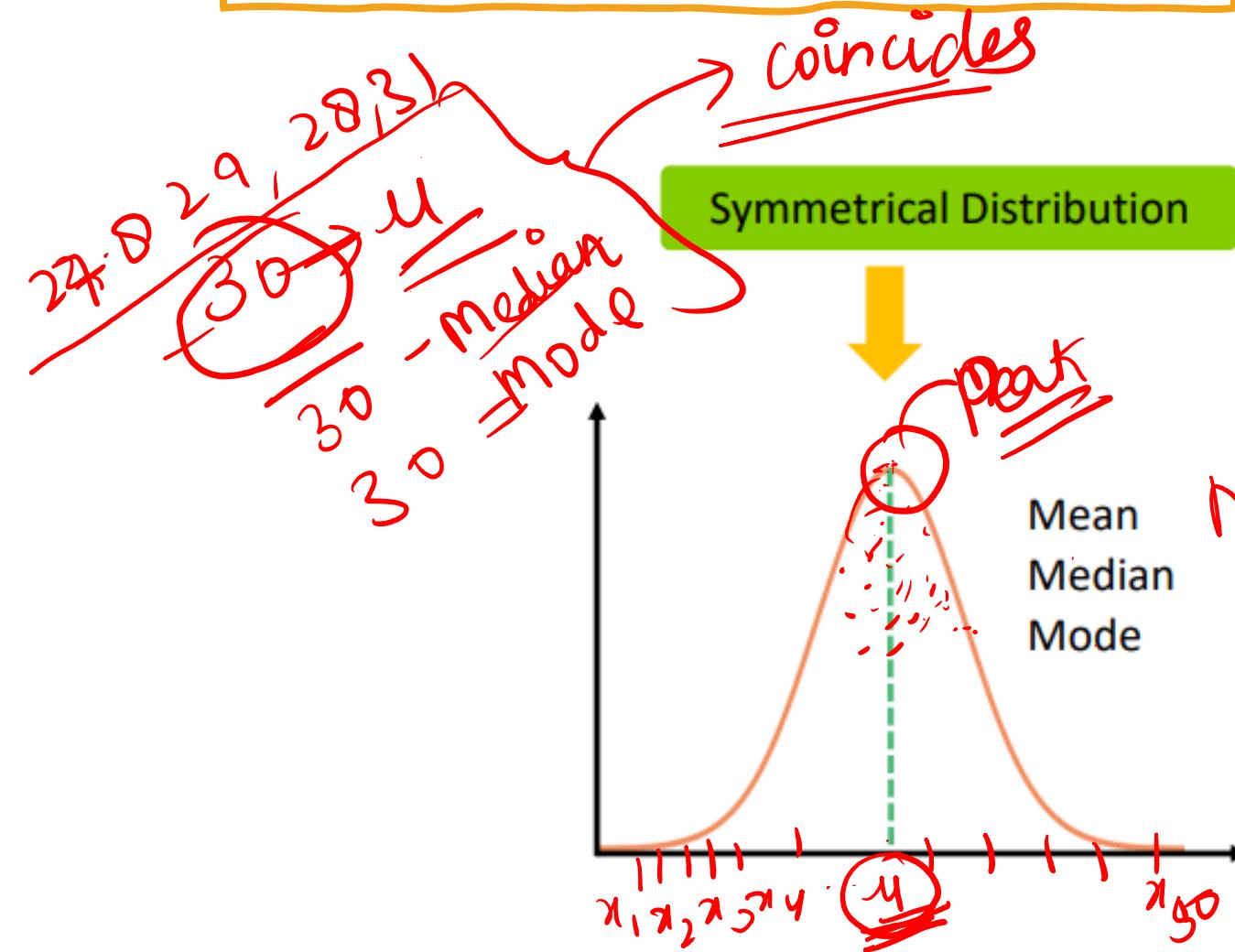
x_i x
 \downarrow
Observation



Descriptive Statistics



Measures of shapes

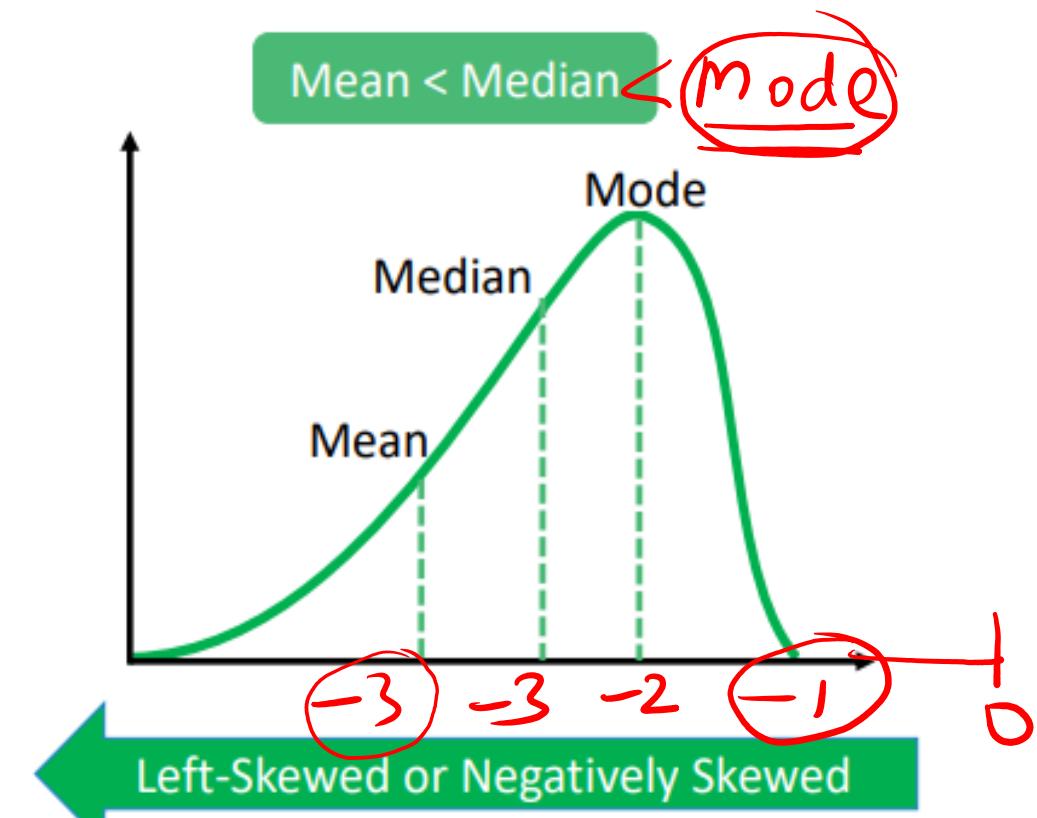
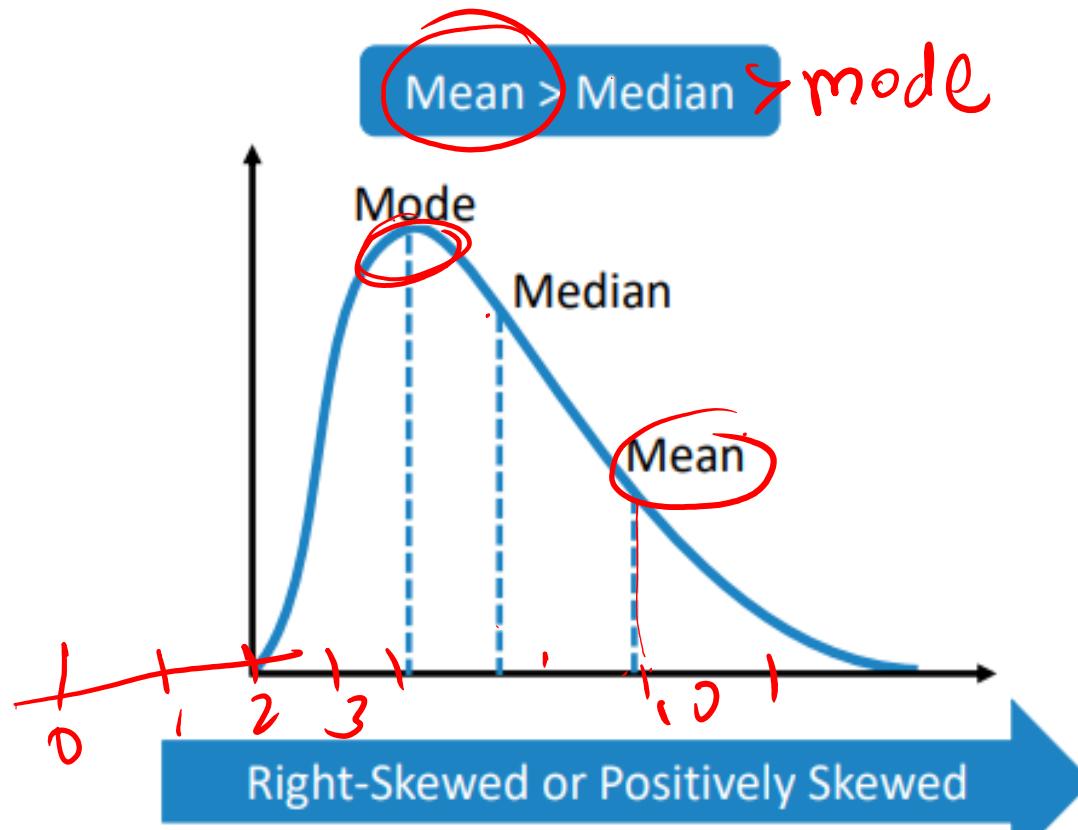




Descriptive Statistics



Skewness



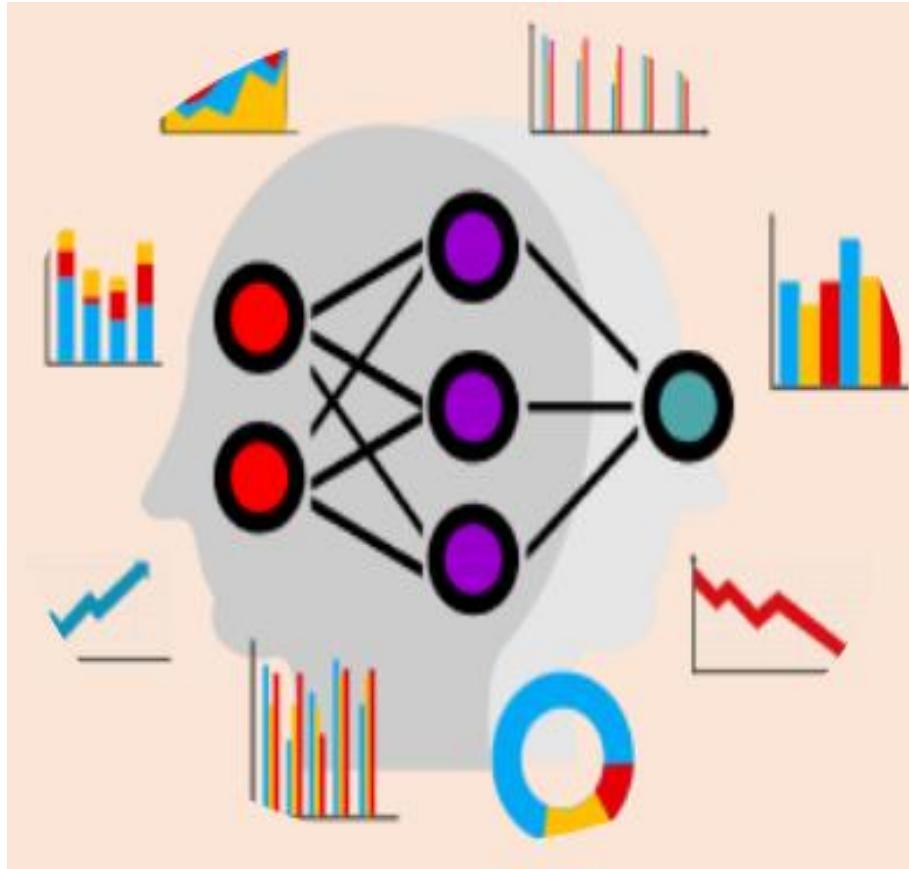


Descriptive Statistics



Inference Statistics

test → Conclusion



Inferential statistics speak about drawing some meaningful inferences from the data after performing suitable statistical tests.



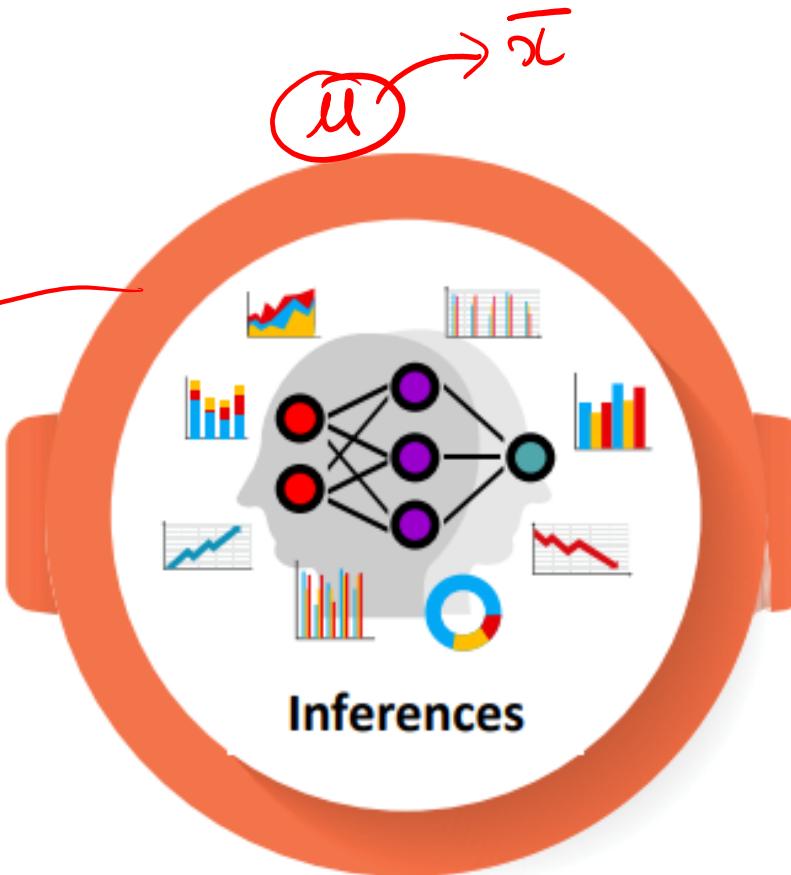
Descriptive Statistics



Types of Inferences



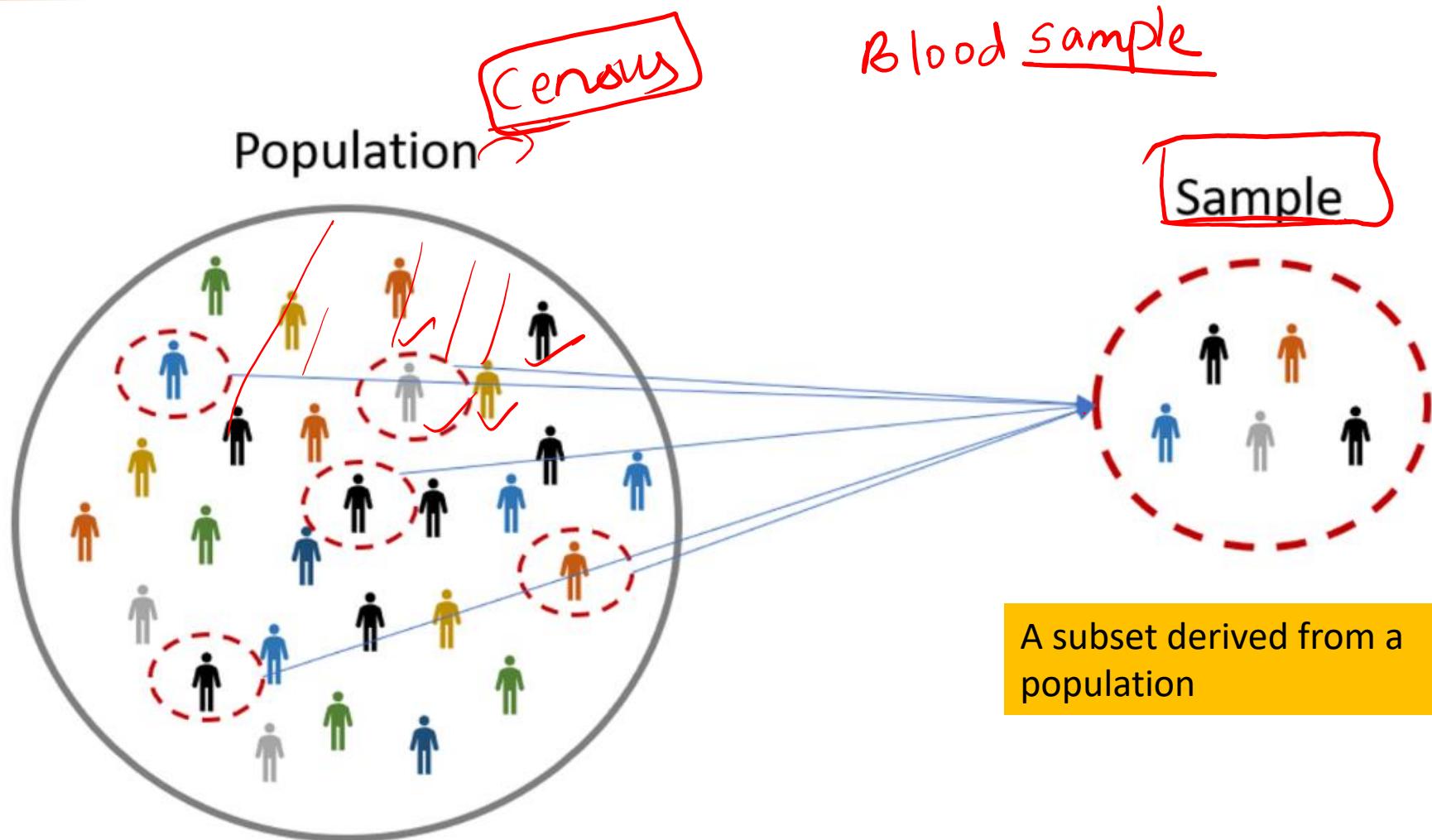
Estimation



Hypothesis Testing



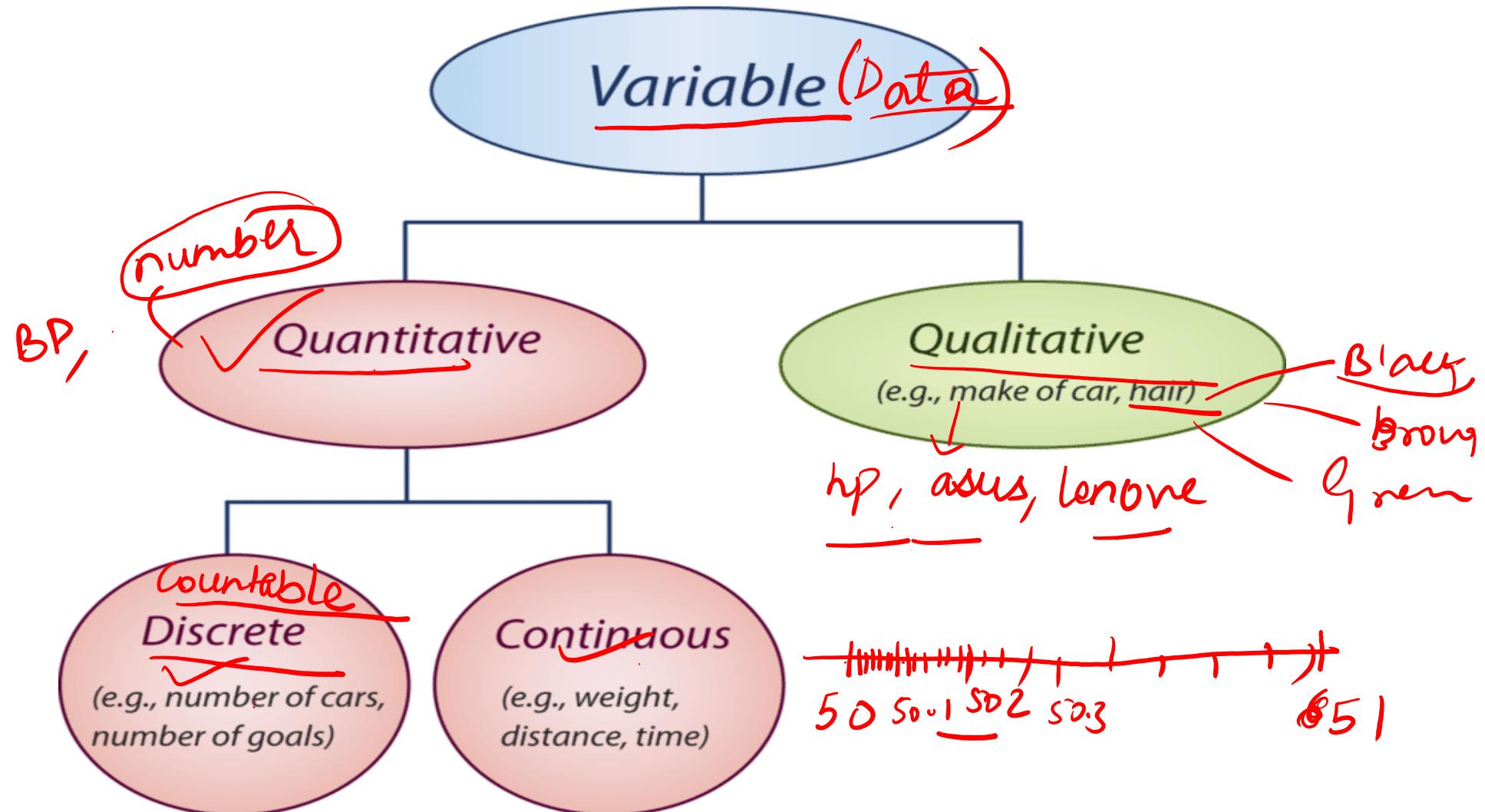
Descriptive Statistics



A set of all the elements under a study
of interest or research



Descriptive Statistics





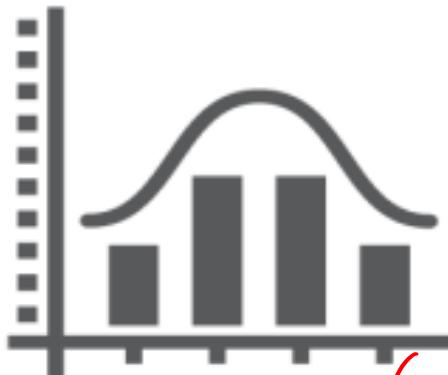
Descriptive Statistics

Quantitative Data

It is also called as **numerical data**.

It can be described by a number.

Examples:



Salaries for employees

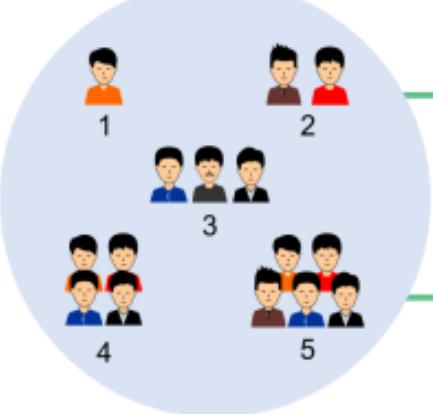


Average salary of employees





Discrete Variables



Number of stocks in an investment portfolio

Assumes only **countable** or **finite number of values**



Number of customers waiting in a queue





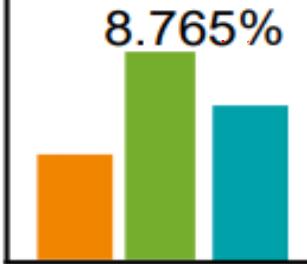
Descriptive Statistics

Continuous Variables



Return on a portfolio of stocks can vary continuously

Example:



Annual return on a
portfolio of stocks

8.764321 - - - .

Variable of interest could assume an infinite number of possible values



Question and Answer Session



Mid-Webinar Break



DIRECTORATE OF
ONLINE EDUCATION
(A unit of Manipal Academy of Higher Education, Manipal)

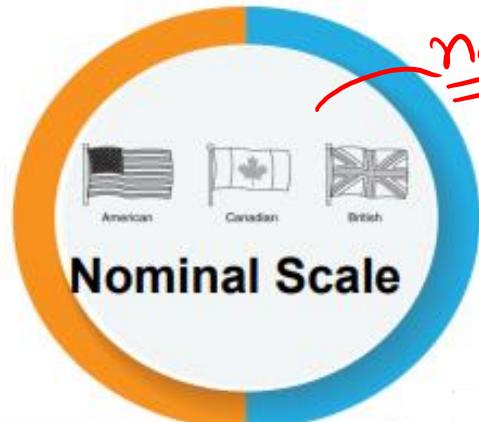
00:00



Descriptive Statistics

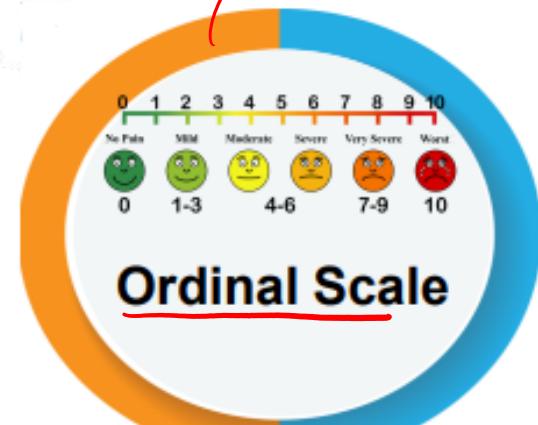


Measurement of scales



name

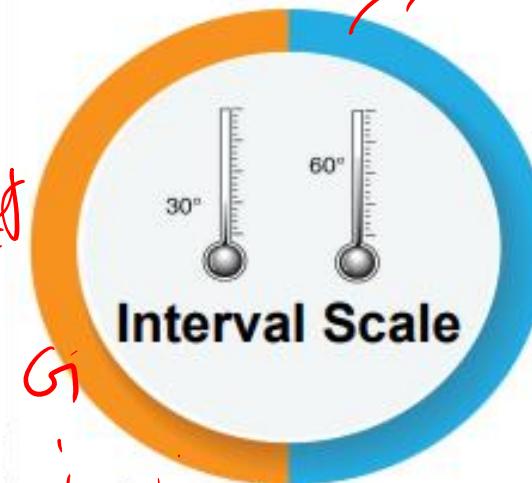
ISD



order / pref

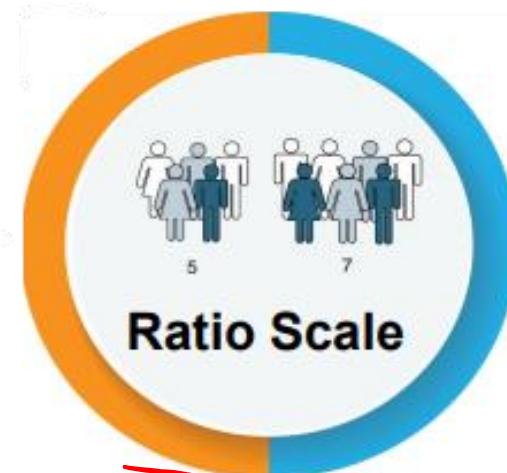
*G
I
C
B
A
P
A+*

X_S, S, M, L, X_L



arbitrary 0

—



Ratio Scale



Descriptive Statistics

Nominal Scale

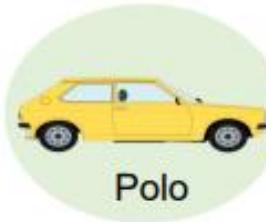
Basic level of measurement to classify or identify an object



John



Famous Five



Polo



Registration
Number



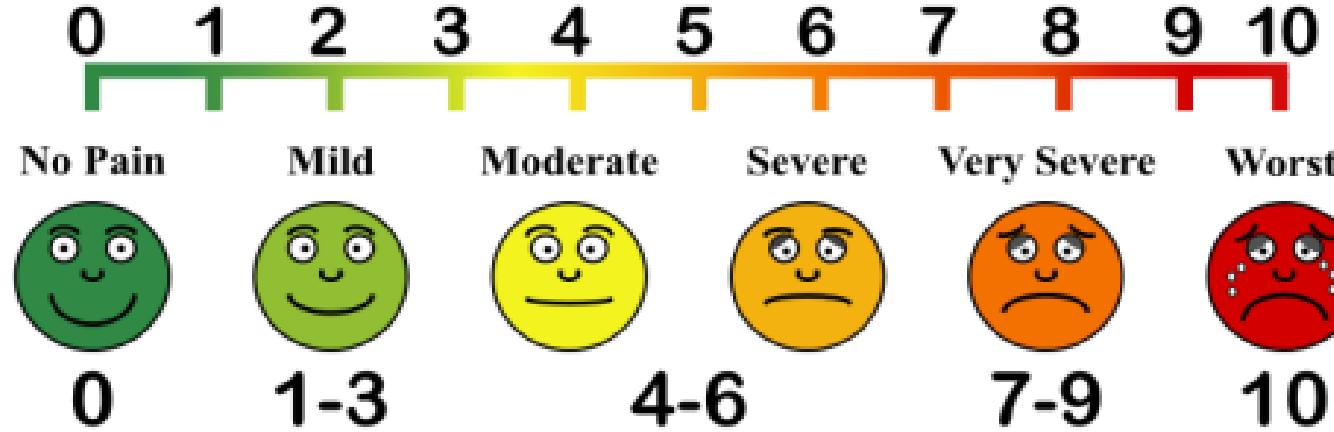
Bank Account
Number

- male–female,
- well–sick,
- under 65 years of age–65 and over,
- child–adult,
- married–not married.



Descriptive Statistics

Ordinal Scale



- The ranking of favorite sports,
- the order of people's place in a line,
- the order of runners finishing a race
- the choice on a rating scale from 1 to 5.
- socioeconomic status as low, medium, or high.
- The intelligence of children may be above average, average, or below average.

Used to order and rank the elements in data

Categorical or qualitative data



Interval Scale



It is **quantitative** in nature.



It has an **arbitrary zero value**.



The **interval** between the two variables is **exact** and not relative.





Ratio Scale



Quantitative in nature

$$\frac{50}{60}$$

$$\underline{4}$$

$$\frac{5}{6}$$

$$\underline{0} \rightarrow$$



Contains an
absolute zero

Variables have all the
arithmetic operations



Examples:

Height, weight, sales of any
product

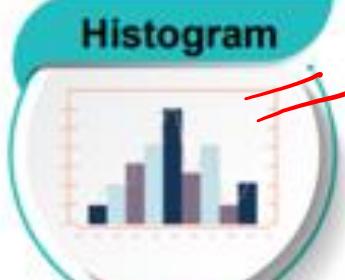




Descriptive Statistics



Types of charts



Histogram



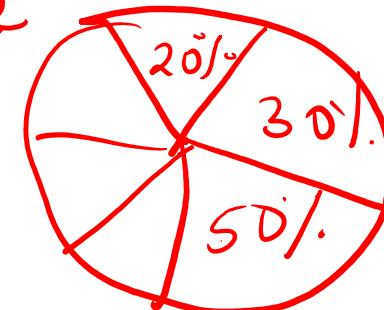
Pie Chart



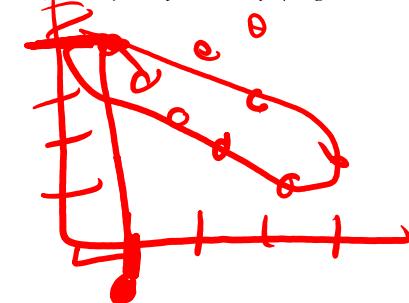
Bar Chart

Represents

Quantitative Data

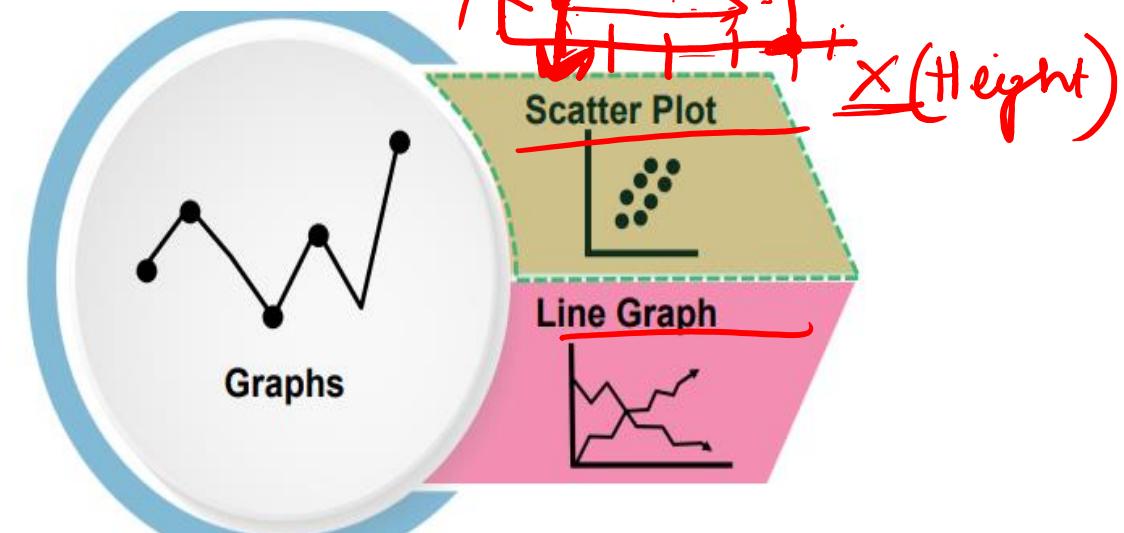
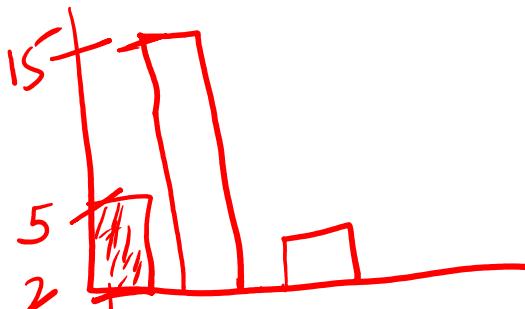
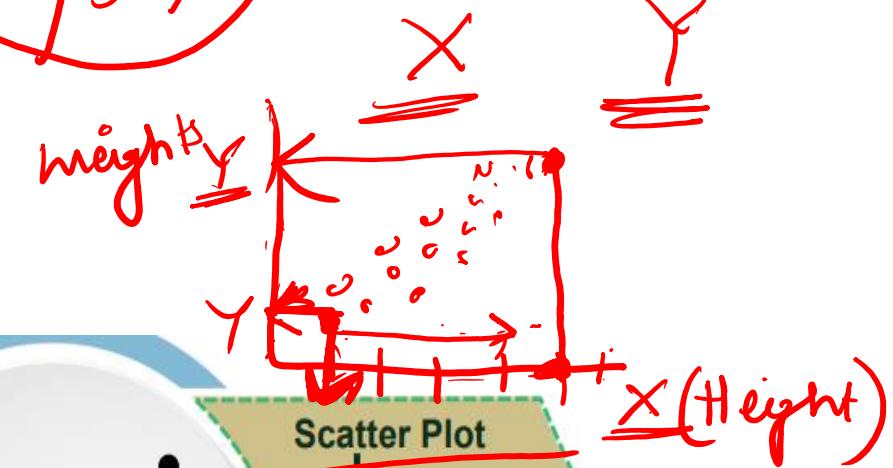


single



Represent

Qualitative or Categorical Data



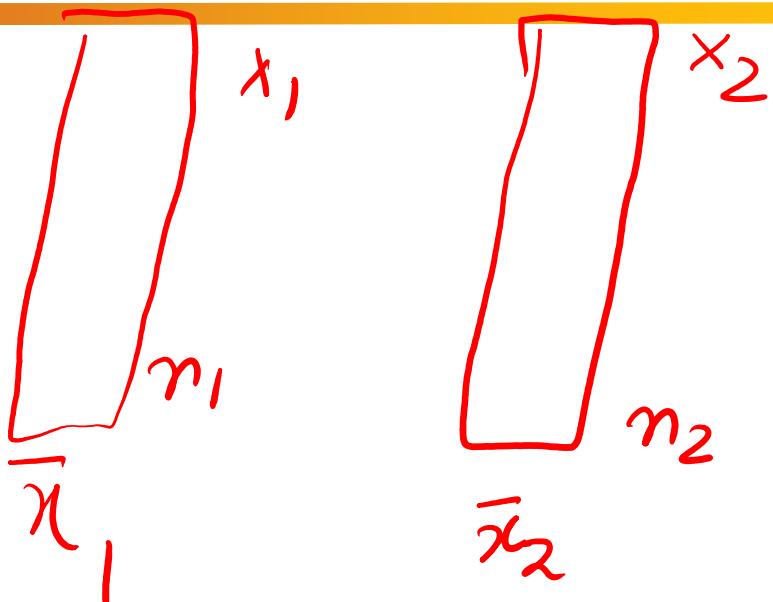
Graphs

Scatter Plot

Line Graph



Descriptive Statistics



$$\text{Cm} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

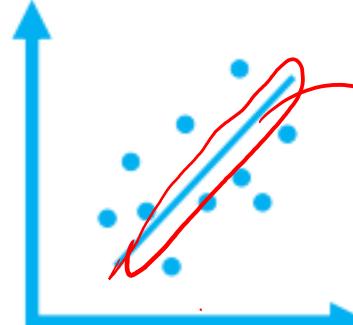


Descriptive Statistics



Measures of Association

X ← or Yes? → Y



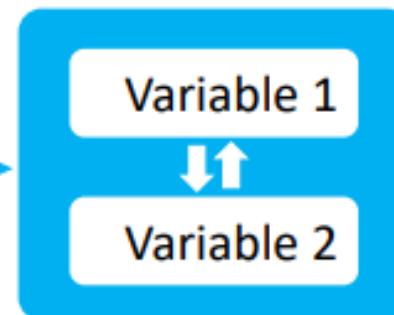
linear relationship



Help to investigate



Useful to summarise



Relationship

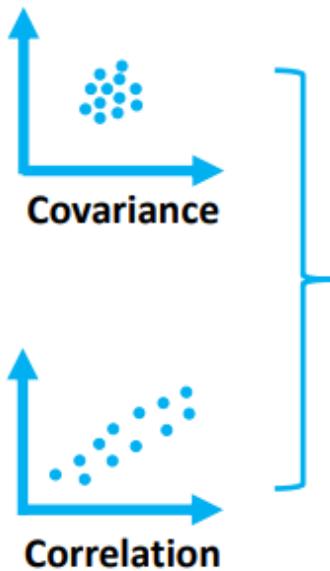
Linear relationship



Measures of Association

covariance

& correlation



Help to describe the strength and direction of the linear relationship between two quantitative variables



Descriptive Statistics



Covariance

$$\frac{(x - \bar{x})(y - \bar{y})}{n}$$

$$M = \frac{0.1}{-0.8} \quad 0.5 \quad 0.9 \\ b = \text{cov}(x, y)$$

①

$$\text{cov}(x, y) = \frac{E(x \cdot y) - E(x)E(y)}{n}$$

Correlation =
• Standardized
• Unitless

$$\text{cor}(x, y) = \frac{\text{cov}(x, y)}{\text{SD}(x) \cdot \text{SD}(y)}$$

$-1 < r < +1$



Correlation

The coefficient of correlation is obtained by dividing the covariance by the product of the standard deviation of both variables.

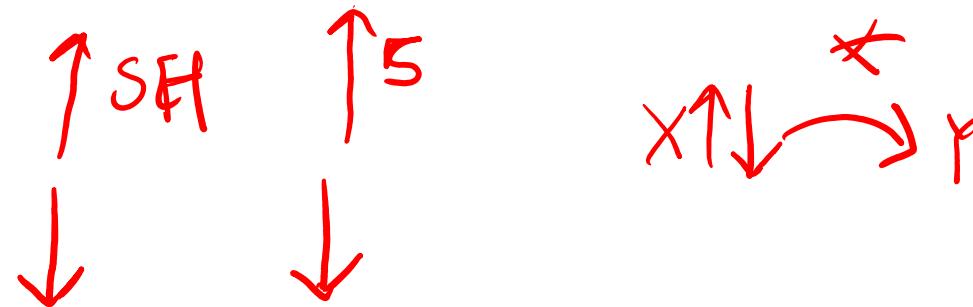
The resulting measure is a unitless quantity that is unaffected by the units of measurement of each of the two variables.



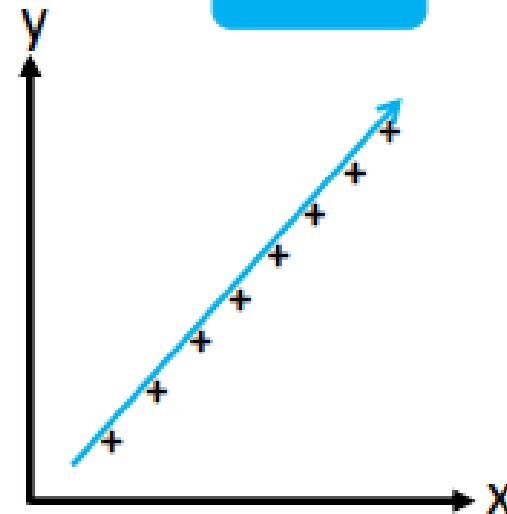
Descriptive Statistics



Correlation

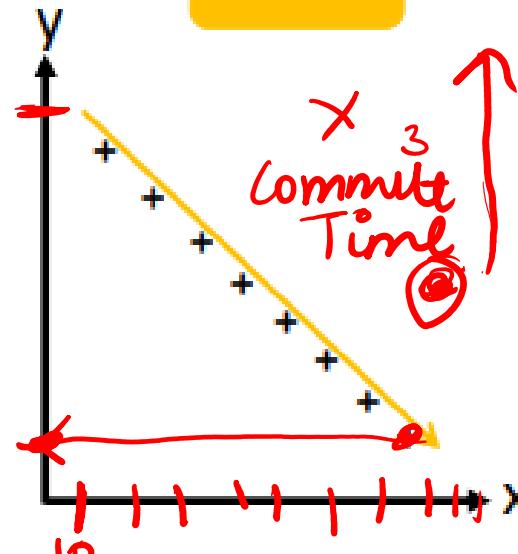


+1



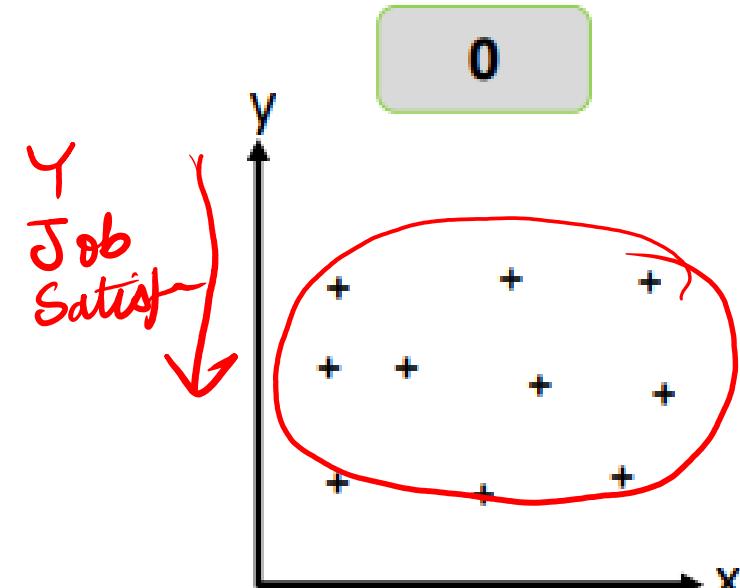
Perfect Positive
Correlation

-1



Perfect Negative
Correlation

0



No Correlation
Uncorrelated



Descriptive Statistics

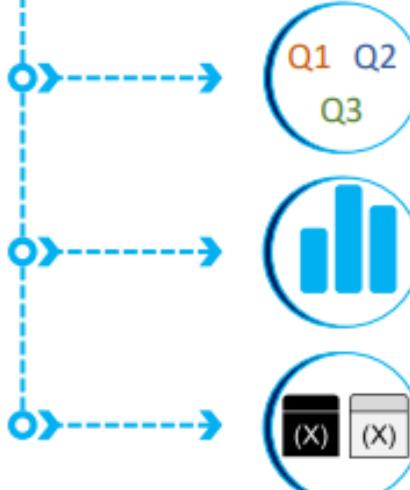


Boxplots



Boxplots

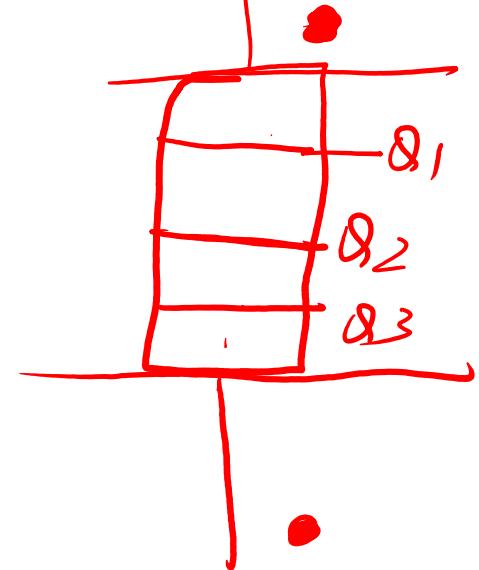
An alternative graphical method
of presenting a distribution



Make use of Quartiles

Graphical representation of
distribution of single variable

Comparison of two or more
variables





Important Questions

No excel

Statistics

Question 1: Discuss various types of scales in data.

Question 2: Explain measure of central tendency, variation

Q - Discuss the descriptive statistics

[nominal

Discrete

-1< Correl <+1



Book References:

- Anderson D.R, Sweeney D.J, Williams T.A, Camm J.D, Cochran J.C (2015). *Statistics for Business & Economics: Revised*, (12th ed.). New Delhi, India: Cengage learning.
- Daniel, W. W., & Cross, C. L. (2013). *Biostatistics: a foundation for analysis in the health sciences*. Tenth edition. Hoboken, NJ, Wiley.
- Albright, S. C., Winston, W. L., & Zappe, C. J. (1999). *Data analysis and decision making with Microsoft Excel*. Pacific Grove, CA: Duxbury Press.

E-References:

<https://www.statsmadeeasy.com/home>