BP 801T Biostatistics & Research Methodology

Ms. Saroj. R.Valvi
Asst. Professor
Dept. of Pharmaceutics
JES's College of Pharmacy, Nandurbar

Unit I

- **Introduction to:** Statistics, Biostatistics, Frequency distribution
- **Measures of central tendency:** Mean, Median, Mode
- **Measures of dispersion:** Dispersion, Range, Standard deviation
- **Correlation:** Karl Pearson's coefficient of correlation, Multiple correlation

Statistics:- It deals with collection, organization, analysis, summarization, interpretation and presentation of data.

Types:-

1) Descriptive statistics: Summarizes features of collected information. Process of using and analyzing a set of collected information.

2) Inferential statistics:- Used to make predictions among large groups using collected information. It is used to deduce properties of population by testing hypothesis.

Statistics

- It is the branch of mathematics that deals with the collection, organisation, analysis, interpretation and presentation of data.

- **Estimator:-** Statistic for calculating estimate about the population based on observed data.
- **Pivotal quantity:-** Function of sample & is observable & unobservable parameter.

- **Hypothesis:-** Assumption based on insufficient evidence.
- **Null hypothesis:-** No change occur over time. No statistical significance between two variables.
- **Alternative hypothesis:-** There is a statistical significance between two variables.

Biostatistics

- It is the branch of statistics which is applied to biological or medical science.

Classification of Biostatistics:-

- **Descriptive Biostatistics**
- **Inferential Biostatistics**

Steps in Biostatistics:-

- Generation of hypothesis
- Collection of experimental data
- Classification of the collected data
- Categorization and analysis of collected data
- Interpretation of data

Measures of Central Tendency

- **Mean (arithmetic average):-** It is the value which we get by dividing the total of the values of various given items in a series by the total number of items.

Formula for Finding the Mean of the Ungrouped Data

$$\text{Mean} = \frac{\text{Sum of the VariablesTotal}}{\text{Number of Variates}}$$

$$\text{Mean} = \frac{x_1+x_2+x_3+x_4+....+x_n}{n}$$

Symbolically, $A = \frac{\sum x_i}{n}$; i = 1, 2, 3, 4, ..., n.

- **Median:-** Is the value of the middle item of series when it is arranged in ascending or descending order of magnitude.

1, 3, 3, **6**, 7, 8, 9

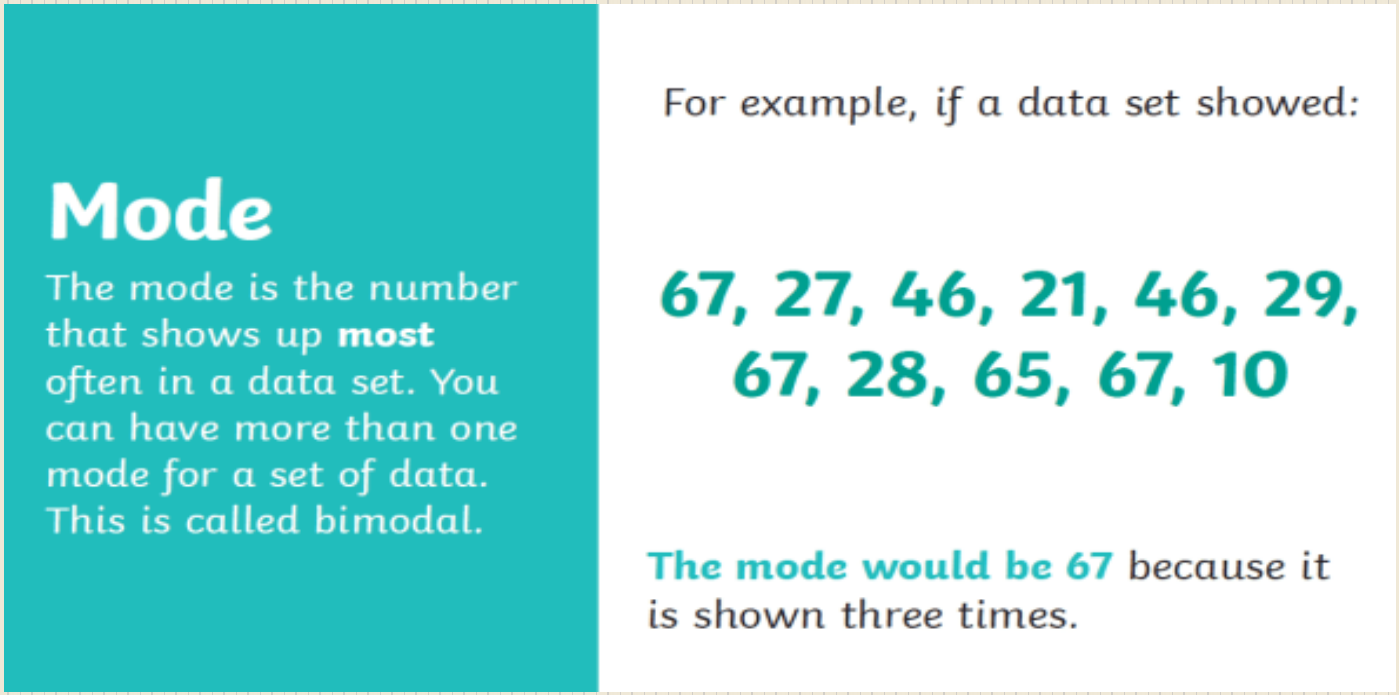
Median = **6**

1, 2, 3, **4**, **5**, 6, 8, 9

Median = (4 + 5) ÷ 2

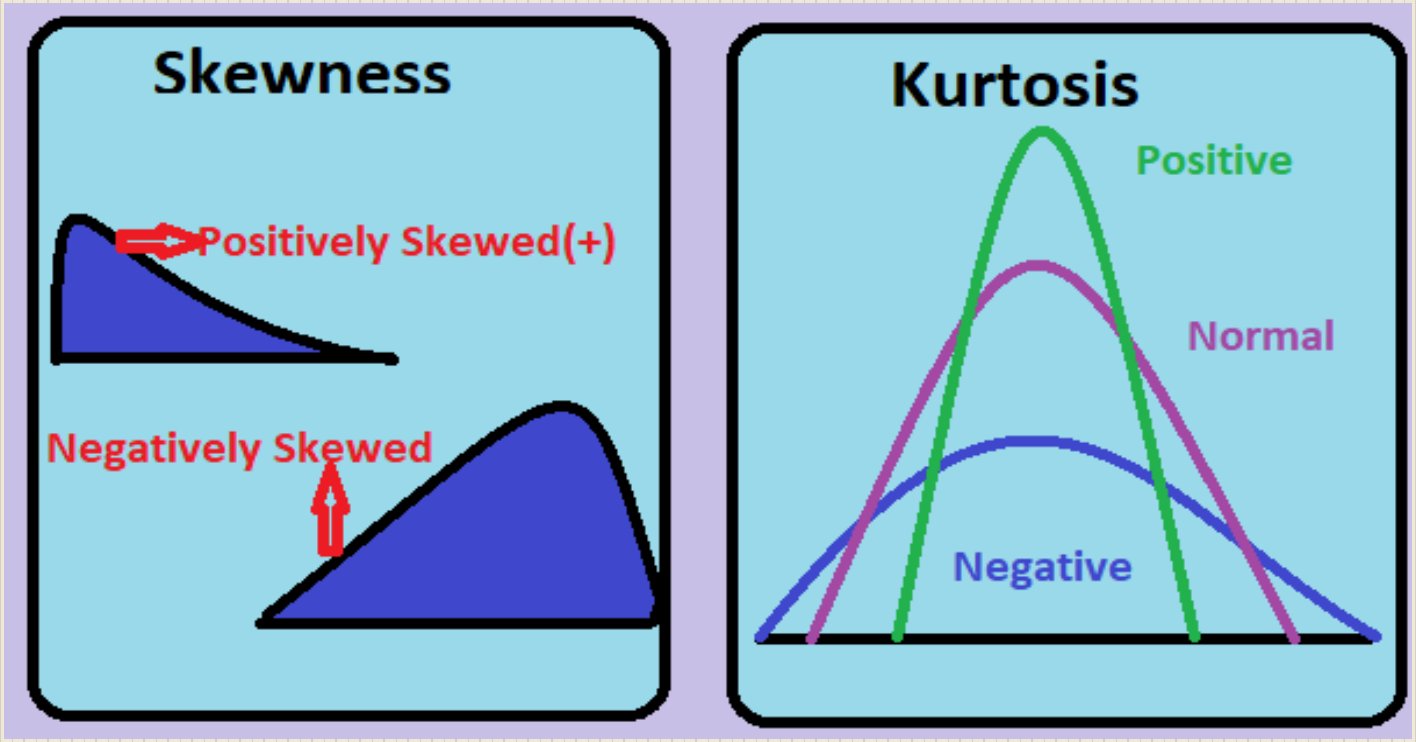
= **4.5**

- **Mode:-** Mode is the most commonly or frequently occurring value in a series. The mode in a distribution is that item around which there is maximum concentration. In general, mode is the size of the item which has the maximum frequency.



Skewness:- Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.

Kurtosis:- Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution.



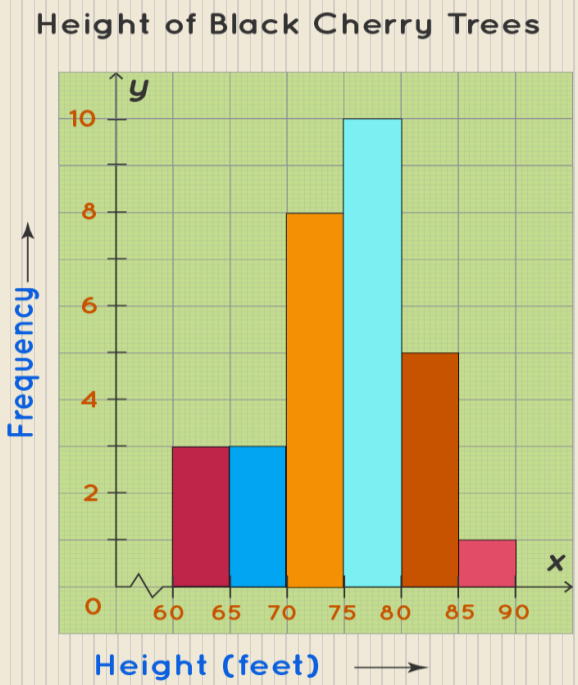
Frequency Distribution

- Is a representation, either in a graphical or tabular format, that displays the number of observations within a given interval. The frequency is how often a value occurs in an interval while the distribution is the pattern of frequency of the variable.

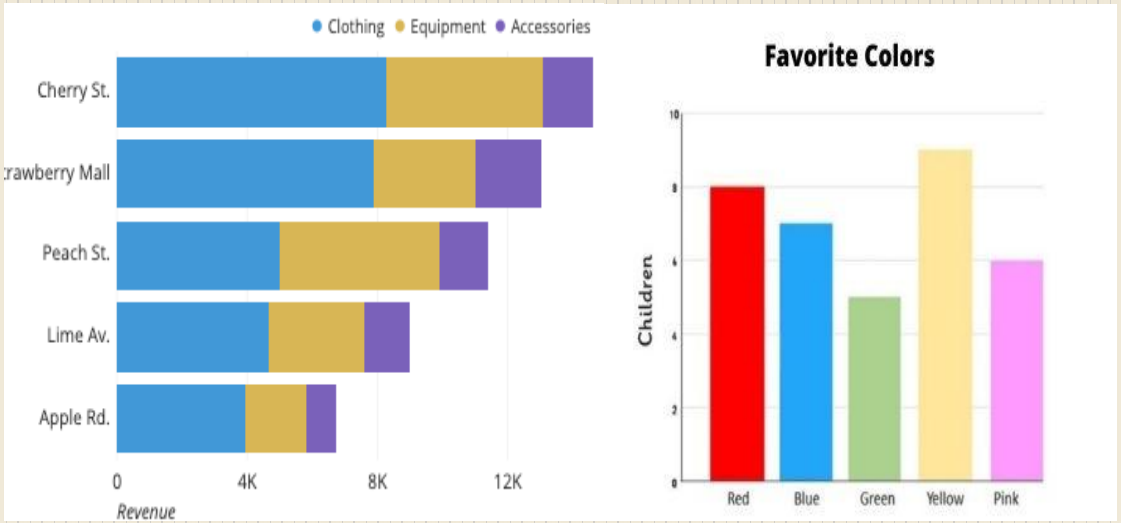
Particle Size (μ)	Frequency	CF	CRF (%)
60-64	2	2	3.17
65-69	7	9	14.29
70-74	11	20	31.75
75-79	15	35	55.56
80-84	10	45	71.43
85-89	9	54	85.71
90-94	6	60	95.24
95-99	3	63	100

Frequency Distribution Graphs

Histogram

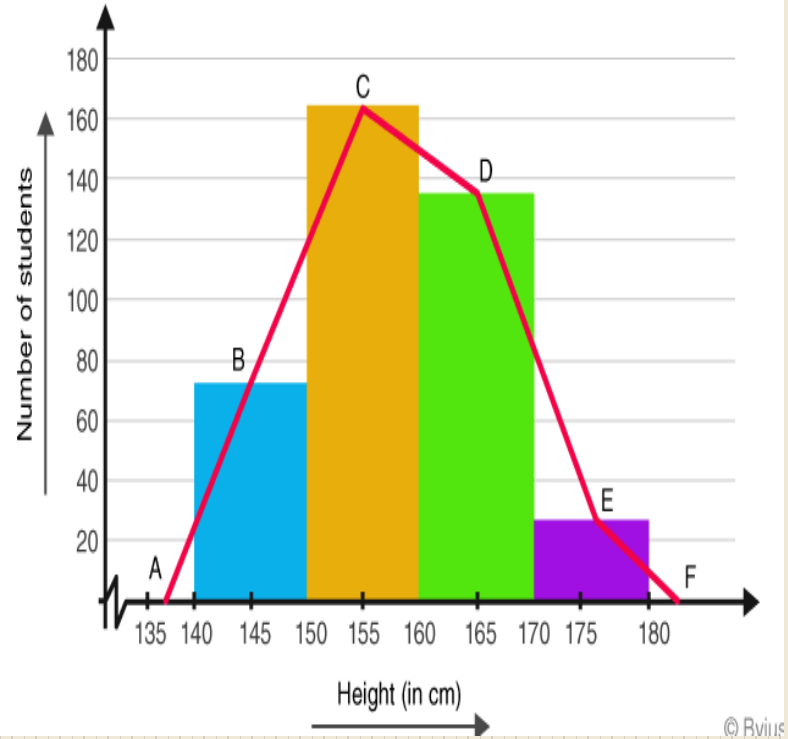


Bar Graph



Frequency Distribution Graphs

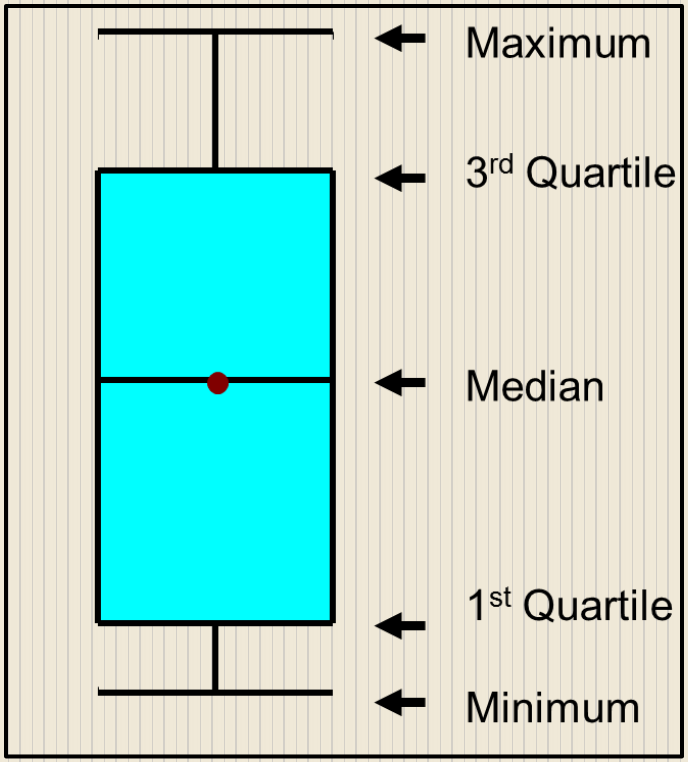
Polygon

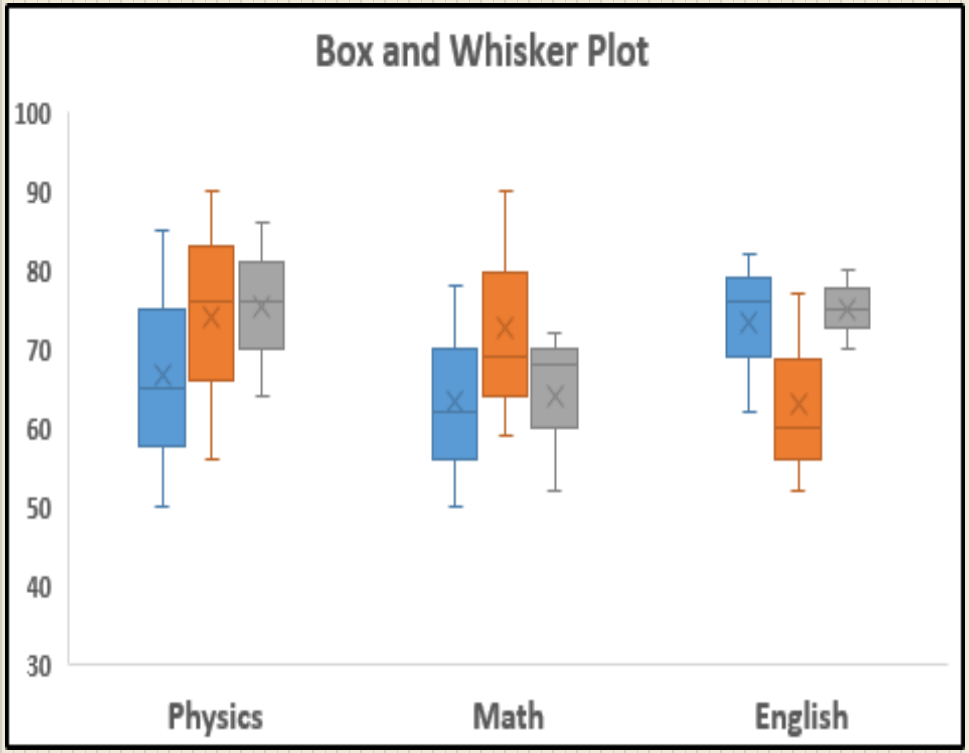


Pie Chart

Frequency Distribution Graphs

Box & Whisker Plots





Characteristics:-

- Measures of central tendency
- Measure of dispersion
- Skewness (The extent of symmetry/asymmetry)
- Kurtosis (Flatness of peakedness)

Measures of Dispersion

- It presents the scatterings in the data.
- It is also called as variability, scatter or spread.
- It helps to interpret the variation of the data from one another.

Degree to which numerical data tend to spread about an average value is called variation or dispersion of data.

Types of Measures of Dispersion:-

1) Absolute Measure:- It contains the same unit as the original data set.

Examples:- Range, SD, etc

2) Relative Measure:- It obtained as a ratios or percentages of the average.

Examples:- Coefficient of range, SD, etc

Range:- It is easily calculated by subtracting the lowest scores in the series from highest. It is the difference between highest and lowest value of dispersion.

$$R = H - L$$

Coefficient of Range:- Ratio between the difference of the extreme values and the sum of the extreme values.

$$\text{Coefficient of Range} = \frac{H-L}{H+L}$$

Merits of Range

1) It is the simplest form of the measure of dispersion.
2) It is easy to calculate and understand.

Demerits of Range

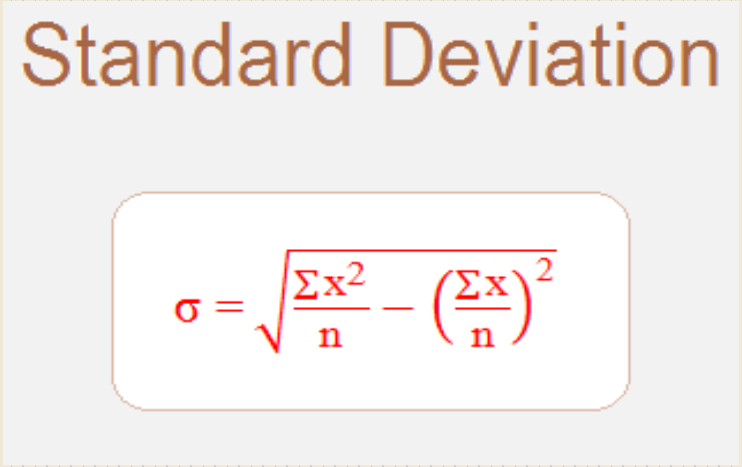
1) It is based on two extreme observations, hence get affected by fluctuations.

Standard Deviation (σ):-

It measures the amount of variation or how spread out numbers in a set of values.

Low value of SD means it is close to the mean, it is also called as **expected value** of data set.

High value of SD means values are spread out over a wide range.



Relative Standard Deviation:-

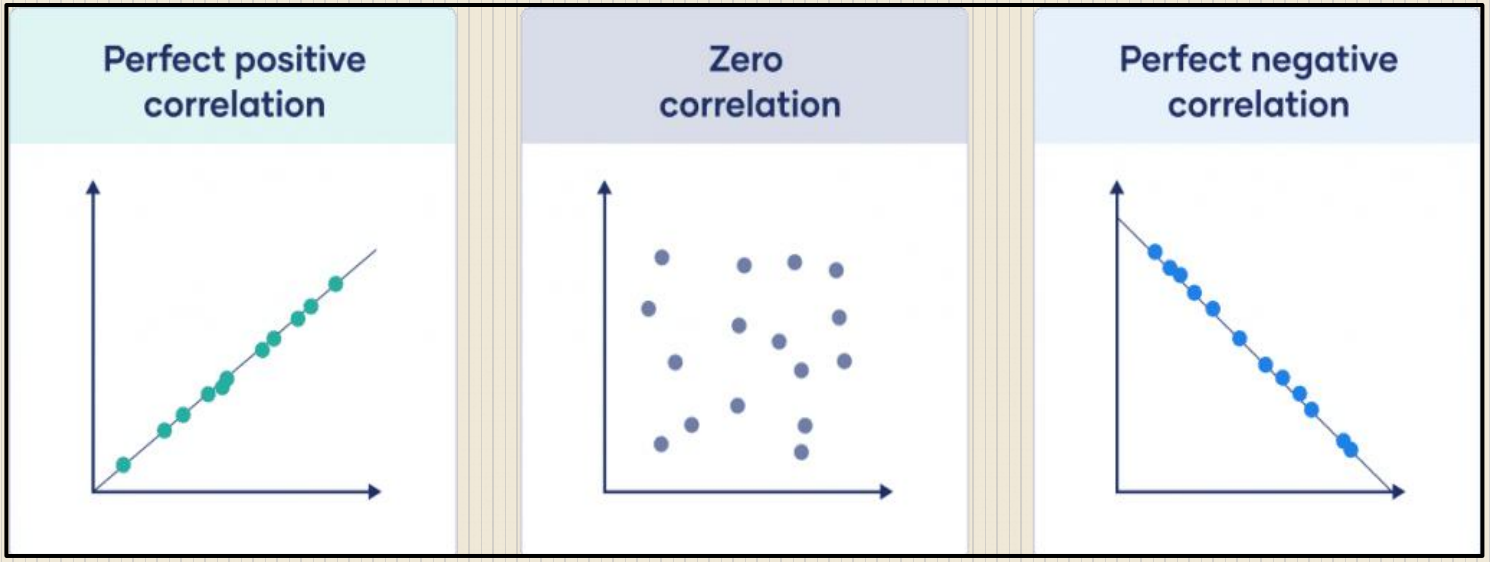
$$RSD = \frac{(S \times 100)}{\bar{x}}$$

Characteristics of SD:-

1) It includes algebraic signs & it is less affected by sampling fluctuations.
2) Small SD has higher probability of getting a value close to the mean.

Correlation

- Means the relation between two variables.
- If the data has only two series, then it is known as bivariate frequency distribution.
- More than two series data is labeled as multi-variate frequency distribution.



Coefficient of Correlation

- Outcome of correlation is known as coefficient of correlation.
- It is reported as **R** or **r** & its value is in between **-1** to **+1**
- +1 indicates strong positive relationship, -1 indicates strong negative relationship & when it is 0 there exist no relationship at all.

Karl Pearson's Coefficient of Correlation (PCC)

- It shows linear relationship between two sets of data.
- It is also called as Pearson Product Moment Correlation (PPMC)
- It does not give any idea about the slope of the line, but it only tells whether there is any relationship exist or not.

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\,n\Sigma x^2 - (\Sigma x)^2\,][\,n\Sigma y^2 - (\Sigma y)^2\,]}}$$

Properties of PCC:-

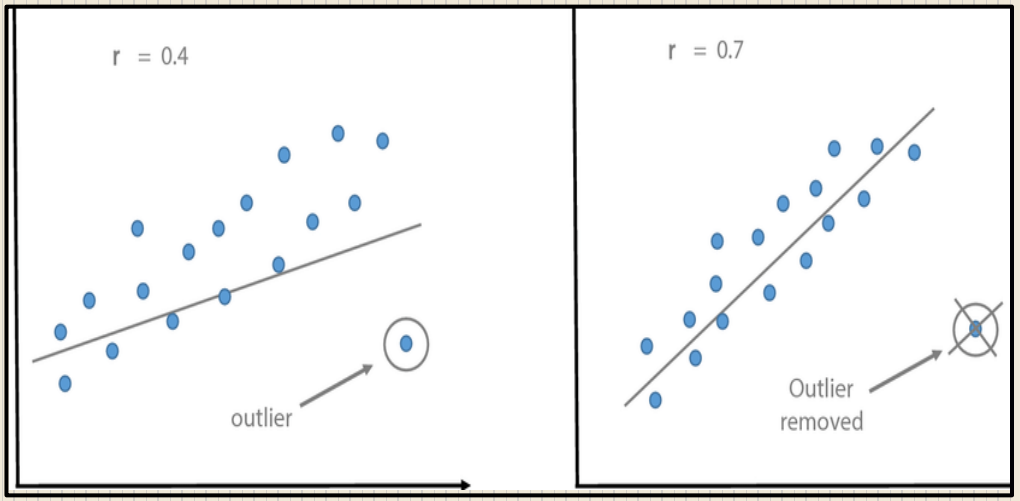
- The r is the unit-less quantity.
- The value of r always fall between +1 & -1 which determine the association between two variables.
- It treats all variables equally.
- A change of origin of the system does not affect the value of r.

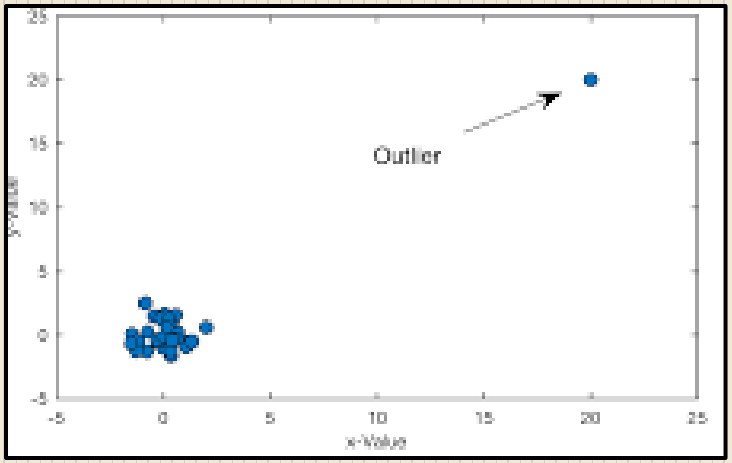
r value indicating strength of variation

Strength	Negative r	Positive r
Weak	-0.1 to -0.3	0.1 to 0.3
Average	-0.3 to -0.5	0.3 to 0.5
Strong	-0.5 to -1.0	0.5 to 1.0

Assumptions of PCC:-

- There exists a linear relationship between the two variables.
- The outliners are either kept to a minimum or remove them entirely.





Multiple Correlation

- Multiple correlations are related with the relationship between more than two variables.
- It is also known **as square root of the coefficient of determination.**
- The coefficient of multiple relations ranges between 0.00 to 1.00
- 1 value indicates predictions are correct & 0 value indicates no linear combination exist.

Thank You!!!!