SYLLABUS

Subject –Business Statistics

1.	**Statistics**: Meaning, Definition ,Significance ,Scope and Limitations of Statistical investigation ,Process of data collection ,primary and secondary Data ,Methods of sampling, preparation of Questionnaire ,Classification and Tabulation of data, preparation of of statistical Series and its types,
2.	**Measurementof Central Tendency**- Mean, Mode, Median,Partition Value, Geometric Mean and Harmonic Mean.
3.	**Dispersion and Skewness**- Range,Lorenz Curve, Quartile Deviation, Mean Deviation, Standard Deviation . Coefficient of Variation, Variance. **Correlation**- Meaning, Definition, Types and Degree of Correlation, Coefficient of Correlation Methods.
4.	**Regression Analysis**–Meaning, Uses, Difference between Correlation and Regression, Rregression Equations, calculation of Coefficient of Regression **Analysis of Time Series**- Meaning, Importance ,Components, Measurement of long term trends. Measurement of of cyclical and Irregular fluctuations.
5.	**Index Number**- Meaning, Characteristics, Importance and uses, construction of Index number, Cos of living Index ,Fisher's ideal Index number, Diagrammatic and Graphical presentation of data. **Association of Attribute** (only two variable),Meaning, Types, Characteristics, Methods of determining Association of Attribute

UNIT — I
STATISTICS

The word "Statistics" of English language has either been derived from the Latin word status or Italian word statistics and meaning of this term is "An organised political state.

Meaning: The science of collecting, analysing and interpreting such data or Numerical data relating to an aggregate of individuals.
E.g:- Statistics of National Income, Statistics of Automobile Accidents, Production Statistics, etc.
Definition: - "The classified facts relating the condition of the people in a state specially those facts which can be stated in members or in tables of members or in any tabular or classified arrangements."

-Webster

"Statistics may be regarded as (i) the study of population (ii) The study of variation (iii) The study of method of reduction of data"

-R.A. Fisher.

The science of statistics is the method of judging collective, natural or social phenomena from the results obtained by the analysis of an enumeration or collection of estimates.

-W I king

Nature /Features /Characteristics of statistics

- It is an aggregate of facts.
- It is numerically expressed.
- It is estimated according to reasonable standard of accuracy.
- It is collected for pre-determined purpose.

It is collected in a systematic manner

Nature /Features /Characteristics of statistics

- It is an aggregate of facts
- It is estimated according to reasonable standard of accuracy
- It is numerically expressed
- Analysis of multiplicity of causes.
- It is collected for pre-determined purpose
- It is collected in a systematic manner

Statistical Methods

- By this method we mean methods specially adapted to the elucidation of quantitative data affected by a multiplicity of causes. Few Methods are:-
- Collection of Data (2) Classification (3) Tabulation (4) Presentation (5) Analysis (6) Interpretation (7) Forecasting.

Theoretical

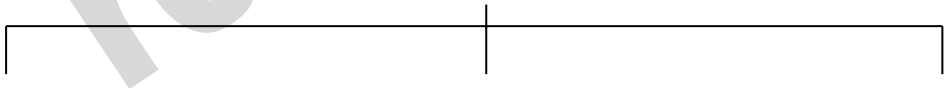
- : Mathematical theory which is the basis of the science of statistics is called theoretical statistics.

Division of Statistics

Applied

- It deals with the application of rules and principles developed for specific problem in different disciplines. Eg: - Time series, Sampling, Statistical Quality control, design of experiments.

Division of Statistics



Theoretical Statistical Methods Applied

Theoretical: Mathematical theory which is the basis of the science of statistics is called theoretical statistics.
Statistical Methods: By this method we mean methods specially adapted to the quantitative data affected by a multiplicity of causes.
Few Methods are:-
(1) Collection of Data (2) Classification (3) Tabulation (4) Presentation (5) Analysis (6) Interpretation (7) Forecasting.
Applied: - It deals with the application of rules and principles developed for specific problem in different disciplines.
Eg: - Time series, Sampling, design of experiments.

Functions of Statistics:-
- It presents facts in a definite form.
- It simplifies mass of figures
- It facilitates comparison
- It helps in prediction
- It helps in formulating suitable & policies.

Scope of Statistics:-
1. Statistics and state or govt.
2. Statistics and business or management.
   - Marketing
   - Production
   - Finance
   - Banking
   - Control
   - Research and Development
3. Statistics and Economics
   - Measures National Income
   - Money Market analysis
   - Analysis of competition, monopoly, oligopoly,
   - Analysis of Population etc.
4. Statistics and science
5. Statistics and Research

Limitations:-
(i) It is not deal with items but deals with aggregates.
(ii) Only on expert can use it
(iii) It is not the only method to analyze the problem.
(iv) It can be misused etc.

STATISTICAL INVESTIGATION

Meaning: In general it means as a statistical survey.
In brief, it is Scientific and systematic collection of data and their analysis with the help of various statistical method and their interpretation.

STAGES OF STATISTICAL INVESTIGATION:-

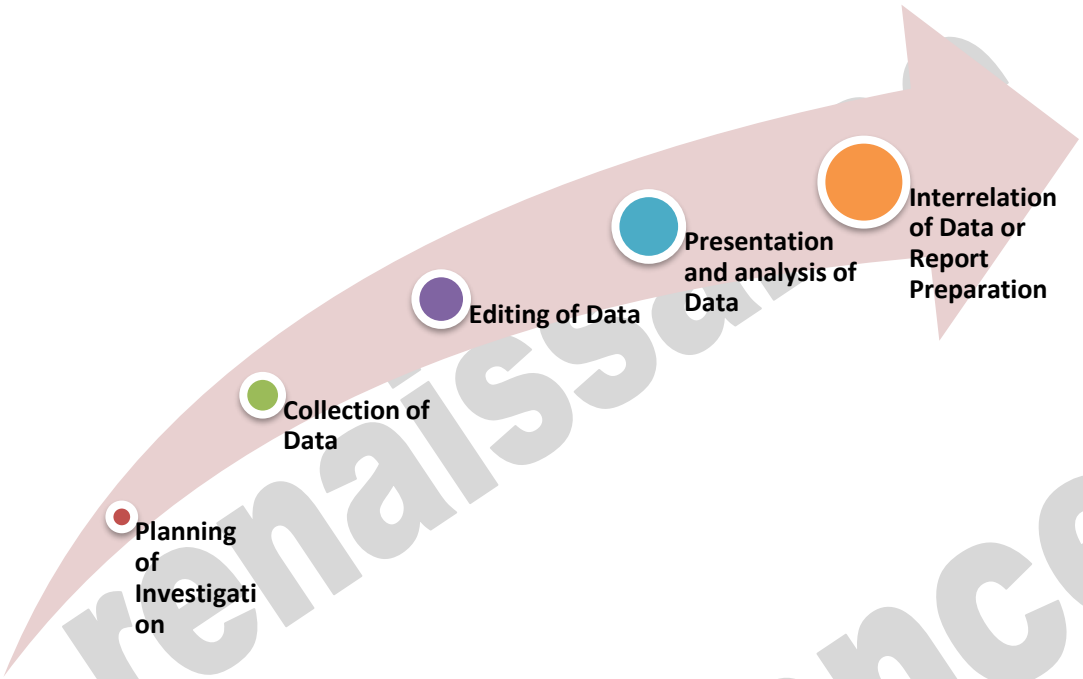
Meaning: In general it means as a statistical survey.
In brief. Scientific and systematic collection of data and their analysis with the help of various statistical method and their interpretation.

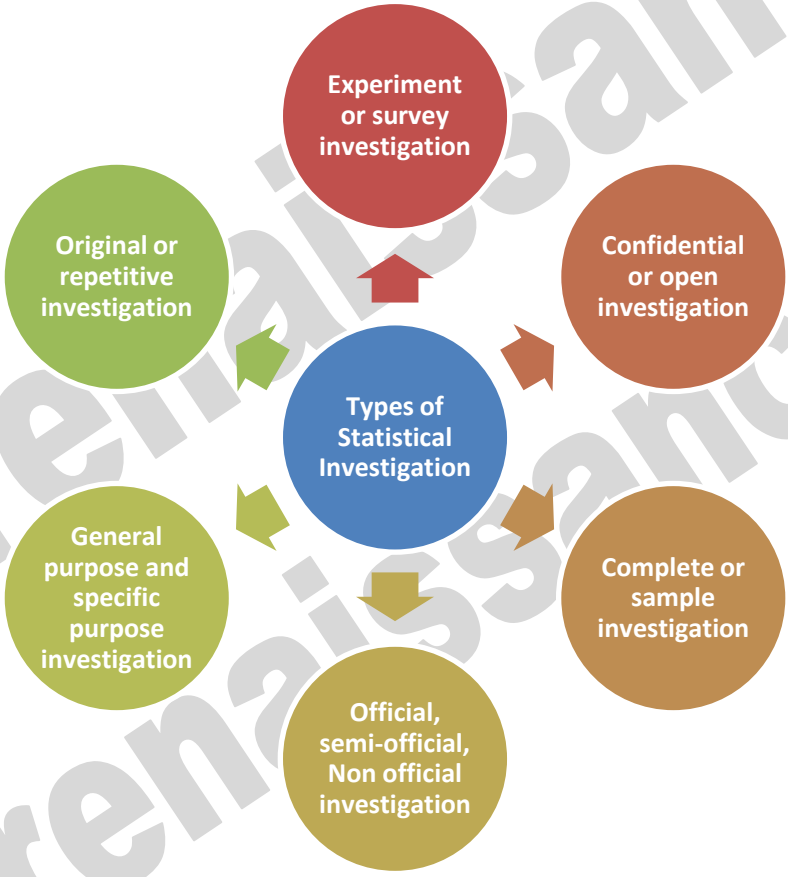
Stages of Statistical Investigation:-

→ Planning of Investigation
→ Collection of Data
→ Editing of Data
→ Presentation of Data
  (a) Classification
  (b) Tabulation
  (c) Diagrams
  (d) Graphs
→ Analysis of Data
→ Interrelation of Data or Report Preparation

Types of Statistical Investigation:-

1. Experiment or survey investigation
2. Complete or sample investigation
3. Official, semi-official, Non official investigation
4. Confidential or open investigation
5. General purpose and specific purpose investigation
6. Original or repetitive investigation.


Planning of Investigation
Collection of Data
Editing of Data
Presentation and analysis of Data
Interrelation of Data or Report Preparation


Types of Statistical Investigation
Experiment or survey investigation
Confidential or open investigation
Complete or sample investigation
Official, semi-official, Non official investigation
General purpose and specific purpose investigation
Original or repetitive investigation

PROCESS OF DATA COLLECTION

Data: - A bundle of Information or bunch of information.
Data Collection: Collecting Information for some relevant purpose & placed in relation to each other.
Types of Data:-

1. **Raw Data:-** When we collect data through schedules and questionnaires or some other method eg:- Classification, tabulation etc.
2. **Processed Data:-** When we use the above raw data for application of different methods of analysing of data. Like using correlation, Z-test, T-test on data. That will be known as processed data.

Sources of Data Collection:-

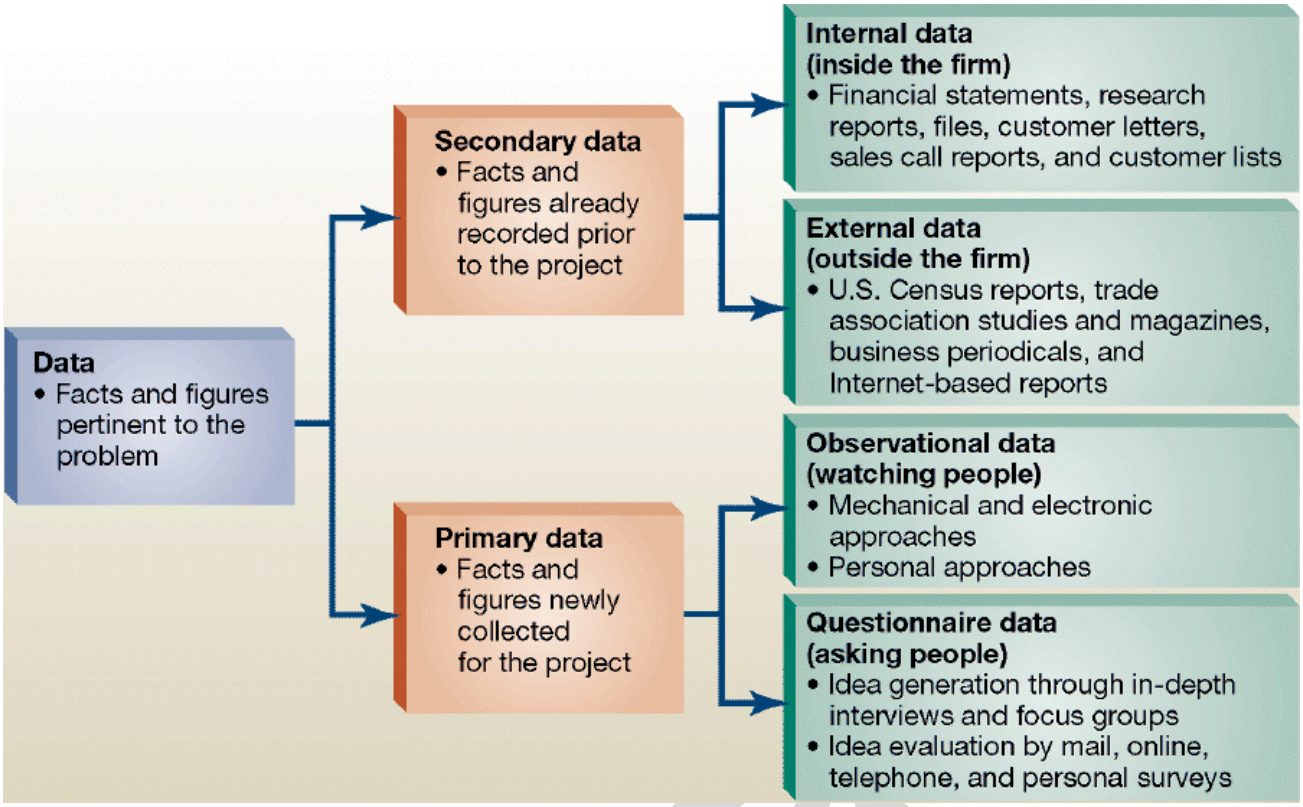
3. **Internal Data: -** When data is collected by problem the internal source for any specific purpose.
4. **External Data: -** This type of data collected by the external source.
5. **Primary Data: -** It is original and collected first time. it is like raw material and it is required large sum of money, energy and time.
6. **Secondary Data: -** Secondary data are those already in existence and which have been collected for some other purpose for answering of the question at hand.
7. **Qualitative Data: -** Which can not be measurable but only there presence and absence in a group of individual can be noted are called qualitative data.
8. **Quantitative Data: -** The characteristics which can be measured directly are known as quantitative data. It is large amount of data.

Methods of Data Collection:- (Primary Data)

- Direct Personal Interviews
- By observation
- By Survey
- By questionnaires

Difference between Primary and secondary data:-

Points	Primary Data	Secondary Data
1. Originality	Primary data are original i.e., collected first time.	Secondary data are not original, i.e.., they are already in existence and are used by the investigator.
2. Organisation	Primary data are like raw material.	Secondary data are in the from of finished product. They have passed through statistical methods.
3. Purpose	Primary data are according to the object of investigation and are used without correction.	Secondary data are collected for some other purpose and are corrected before use.
4. Expenditure	The collection of primary data require large sum, energy and time.	Secondary data are easily available from secondary sources (published or unpublished).



Collection of Data: - It means the methods that are to be employed for obtaining the required information from the units under investigations.

Methods of Data Collection:- (Primary Data)

- Direct Personal Interviews
- By observation
- By Survey
- By questionnaires

Preparation of Questionnaires:-

This method of data collection is quit popular, particularly in case of big enquires, it is adopted by individuals, research workers. Private and public organization and even by government also.
A questionnaires consists of number of question printed or type in a definite order on a form or set of forms. The respondents have to answer the question on their own.

Importance:-

i. Low cost and universal
ii. Respondents have adequate time to respond
iii. Fairly approachable

Demerits:-

(i) Low rate of return
(ii) Fill on educated respondents

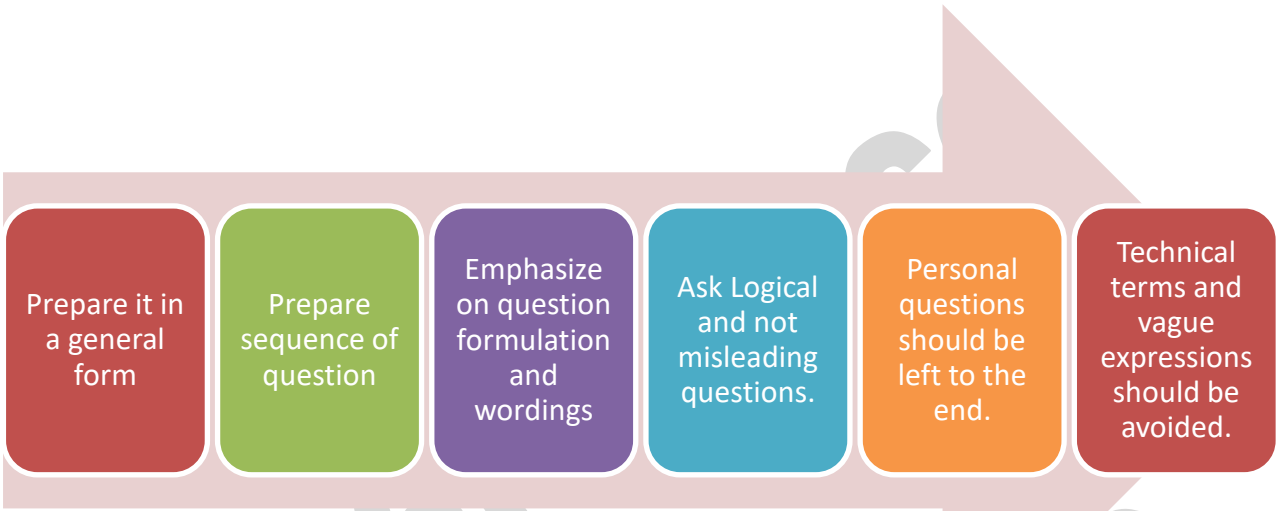
(iii) Slowest method of Response

Preparation of Questionnaires: - It is considered as the heart of a survey operation. Hence it should be very carefully constructed. If it is not properly set up and carefully constructed.

Step I	:-	Prepare it in a general form.
Step II	:-	Prepare sequence of question.
Step III	:-	Emphasize on question formulation and wordings
Step IV	:-	Ask Logical and not misleading questions.
Step V	:-	Personal questions should be left to the end.
Step VI	:-	Technical terms and vague expressions should be availed classification and Tabulation of Data

Basis of Difference	Primary Data	Secondary Data
Originality	Primary data are original i.e., collected first time.	Secondary data are not original, i.e.., they are already in existence and are used by the investigator.
Organisation	Primary data are like raw material.	Secondary data are in the from of finished product. They have passed through statistical methods.
Purpose	Primary data are according to the object of investigation and are used without correction.	Secondary data are collected for some other purpose and are corrected before use.
Expenditure	The collection of primary data require large sum, energy and time.	Secondary data are easily available from secondary sources (published or unpublished).
Precautions	Precautions are not necessary in the use of primary data.	Precautions are necessary in the use of secondary data.

Steps in construction of a questionnaire : It is considered as the heart of a survey operation. Hence it should be very carefully constructed



Example :

Classification & Tabulation of Data

After collecting and editing of data an important step towards processing that classification. It is grouping of related facts into different classes.

Types of classification:-

i. **Geographical:-** On the basis of location difference between the various items. E.g. Sugar Cave, wheat, rice, for various states.

ii. **Chronological:-** On the basis of time
e.g.-

Year	Sales
1997	1,84,408
1998	1,84,400
1999	1,05,000

iii. **Qualitative classification: -** Data classified on the basis of some attribute or quality such as, color of hair, literacy, religion etc.

iv. **Quantitative Classification: -** When data is quantify on some units like height, weight, income, sales etc.

Classification & Tabulation of Data

After collecting and editing of data an important step towards processing that classification. It is grouping of related facts into different classes.

Types of classification:-

v. **Geographical:-** On the basis of location difference between the various items. E.g. Sugar Cave, wheat, rice, for various states.

vi. **Chronological:-** On the basis of time
e.g.-

Year	Sales
1997	1,84,408
1998	1,84,400
1999	1,05,000

vii. **Qualitative classification: -** Data classified on the basis of some attribute or quality such as, colour of hair, literacy, religion etc.

Population

viii. **Quantitative Classification: -** When data is quantify on some units like height, weight, income, sales etc.

Tabulation of Data

A table is a systematic arrangement of statistical data in columns and Rows.

Part of Table:-

1. Table number
2. Title of the Table - A title is the main heading written in capital shown at the top of the table.
   Caption
3. Stub - The horizontal headings and sub heading of the row are called row captions Body of the table
4. Head note
5. Foot Note - It appears immediately below the body of the table providing the further additional explanation.

Types of Table:-

(i) Simple and Complex Table:-

(a) Simple or one-way table:-

Age	No. of Employees

25	10
30	7
35	12
40	9
45	6

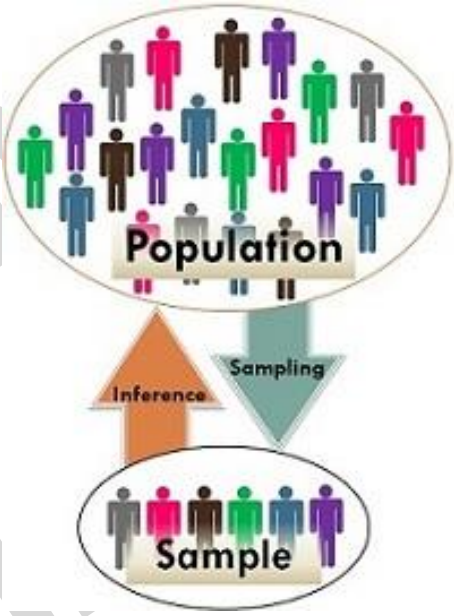
(b) Two way Table

Age	Males	Females	Total
25	25	15	40
30	20	25	45
35	24	20	44
40	18	10	28
45	10	8	18
Total	**97**	**78**	**175**

2) General Purpose and Specific Purpose Table:- General purpose table, also known as the reference table or repository tables, which provides information for general use or reference.
Special purpose are also known as summary or analytical tables which provides information for one particular discussion or specific purpose.

METHODS OF SAMPLING

Meaning: - The process of obtaining a sample and its subsequent analysis and interpretation is known as sampling and the process of obtaining the sample if the first stage of sampling.



The various methods of sampling can broadly be divided into:

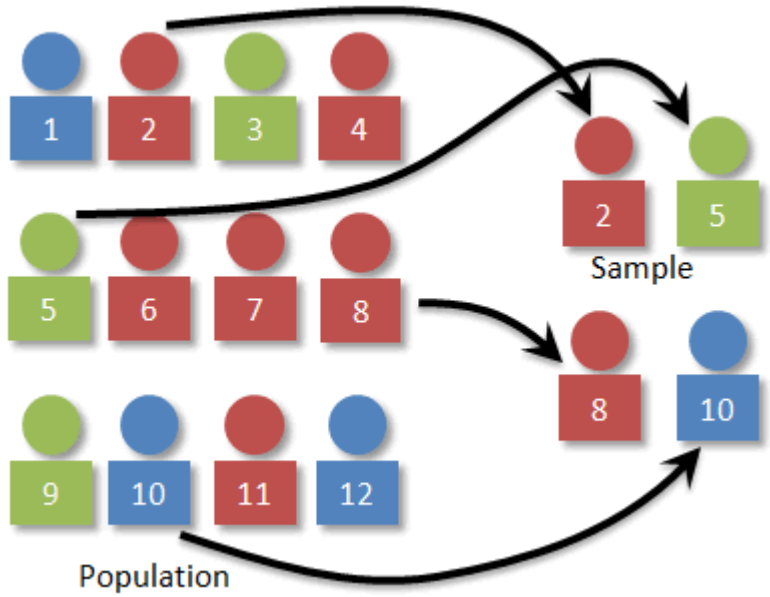
i. Random sampling method

ii. Non Random sampling method

Random Sampling Method

I Simple Random Sampling: - In this method each and every item of the population is given an equal chance of being included in the sample.

(a) Lottery Method (b) Table of Random Numbers



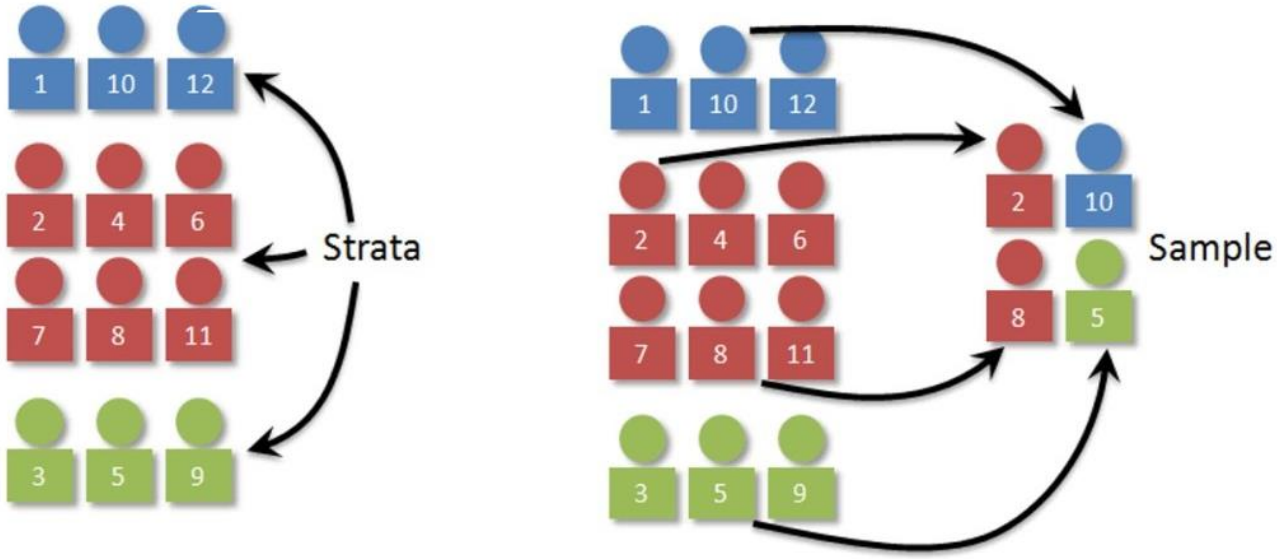
Merits:

Equal opportunity to each item.
Better way of judgment
Easy analysis and accuracy

Limitations:

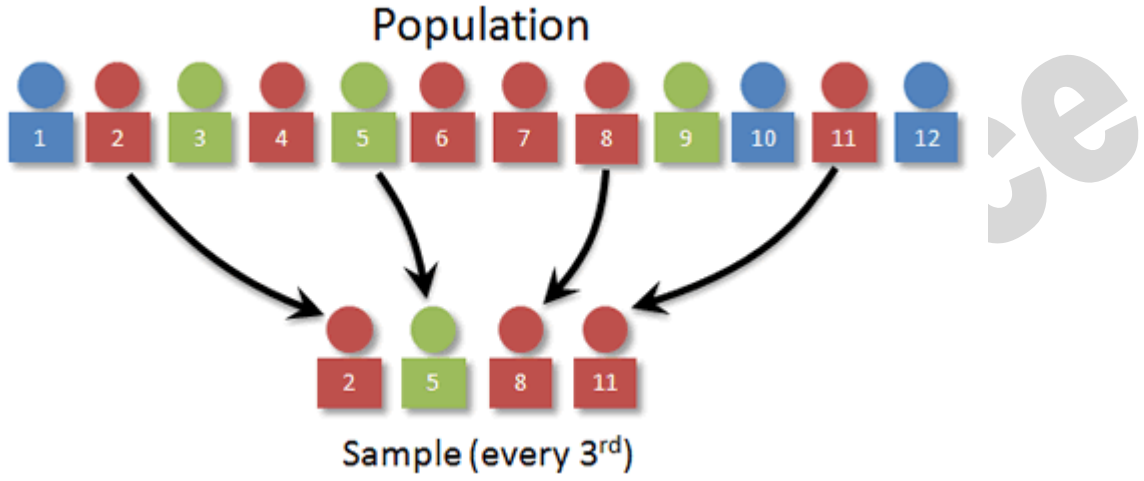
Different in investigation
Expensive and time consuming
For filed survey it is not good

II Stratified Sampling:- In this it is important to divided the population into homogeneous group called strata. Then a sample may be taken from each group by simple random method.



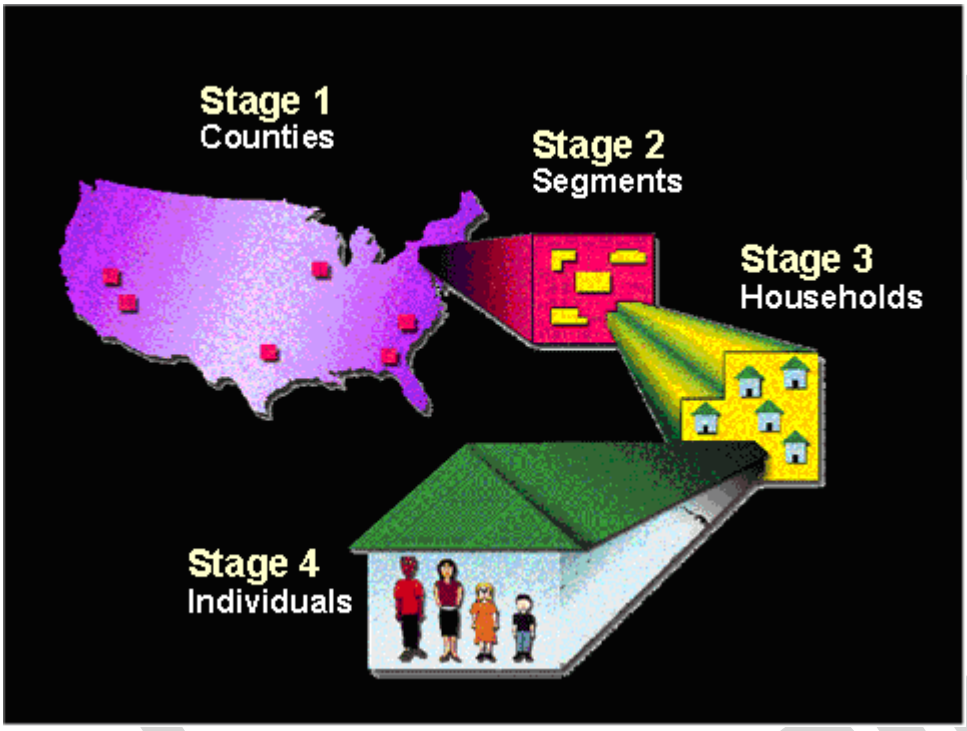
Merit:- More representative sample is used.
Grater accuracy
Geographically Concentrated
Limitations: Utmost care must be exercised due to homogeneous group deviation. In the absence of skilled supervisor sample selection will be difficult.

III Systematic Sampling:- This method is popularly used in those cases where a complete list of the population from which sampling is to be drawn is available. The method is to be select k th item from the list where k refers to the sampling interval.



Merits: - It can be more convenient.
Limitation: - Can be Baised.

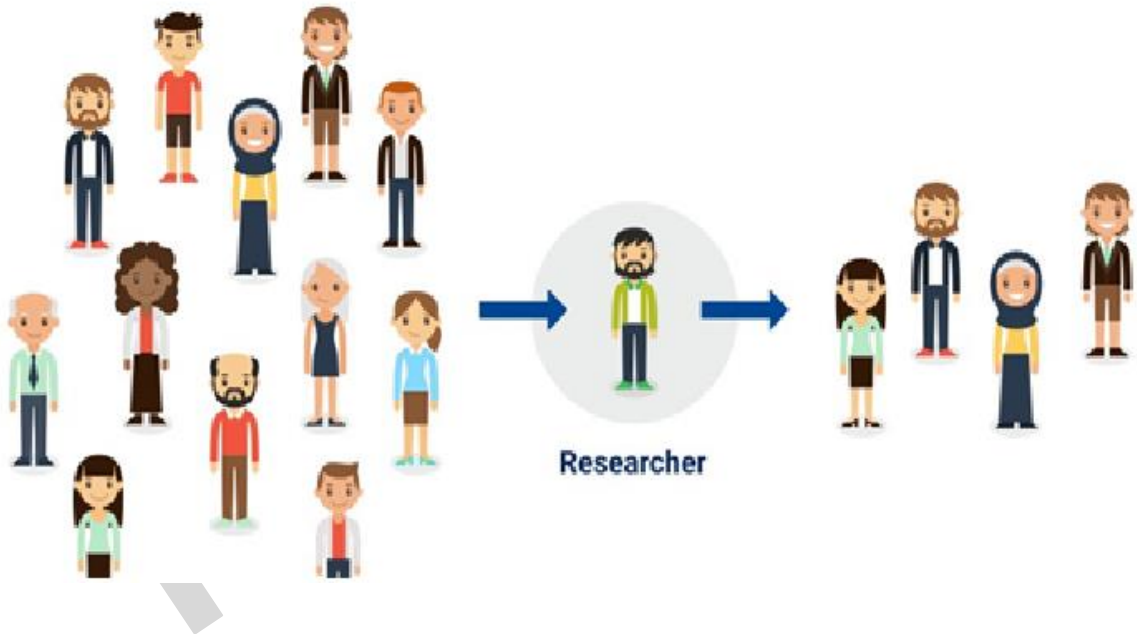
IV Multi- Stage Sampling: - This method refers to a sampling procedure which is carried out in several stages.



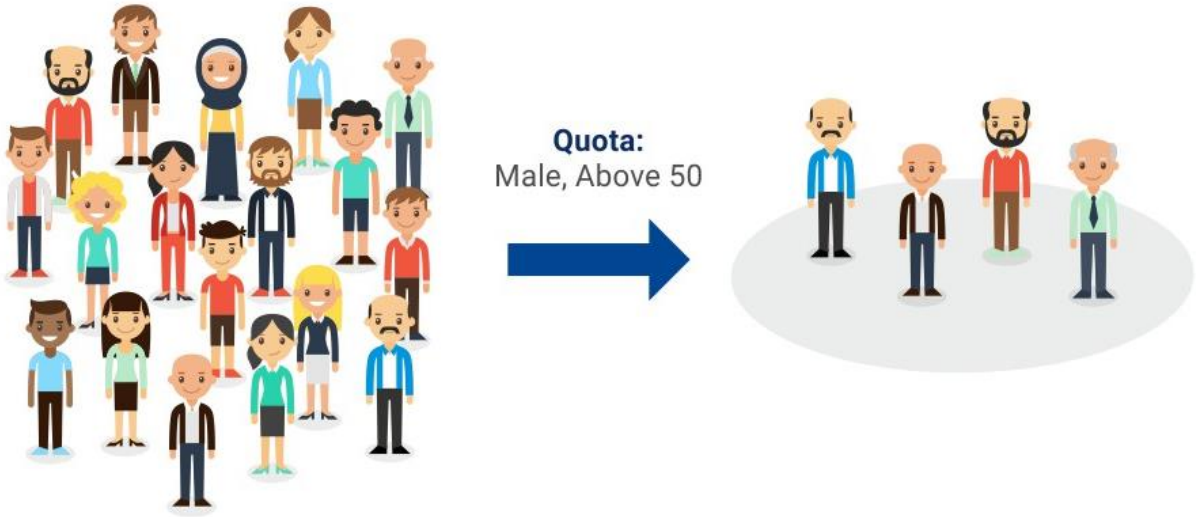
Merit: - It gives flexibility in Sampling
Limitation: - It is difficult and less accurate

Non Random Sampling Method:-

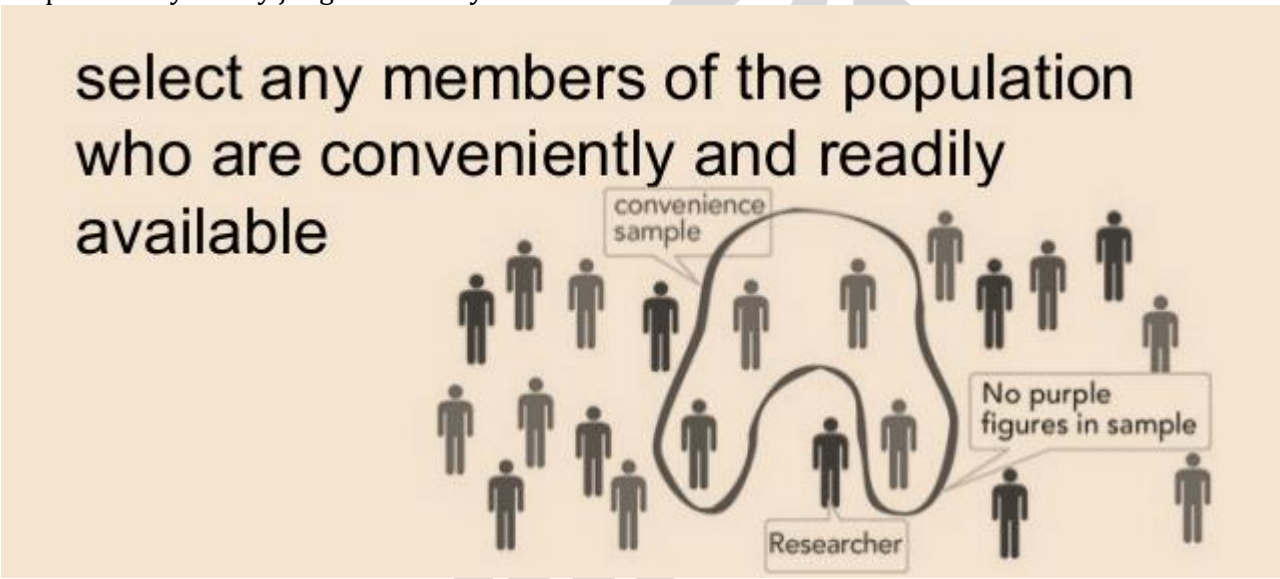
I. **Judgment Sampling: -** The choice of sample items depends exclusively on the judgment of the investigator or the investigator exercises his judgement in the choice of sample items. This is an simple method of sampling.



II. **Quota Sampling: -** Quotas are set up according to given criteria, but, within the quotas the selection of sample items depends on personal judgment.



III. **Convenience Sampling: -** It is also known as chunk. A chunk is a fraction of one population taken for investigation because of its convenient availability. That is why a chunk is selected neither by probability nor by judgment but by convenience.



Size of Sample:- It depends upon the following things:-
Cost aspects.
The degree of accuracy desired.
Time, etc.
Normally it is 5% or 10% of the total population.

Limitation of overall sampling Method:-

Some time result may be inaccurate and misleading due to wrong sampling.
Its always needs superiors and experts to analyze the sample.
It may not give information about the overall defects. In production or any study.
It Becomes Biased due to following reason:-
(a) Faulty process of selection
(b) Faulty work during the collection of information
(c) Faulty methods of analysis etc.

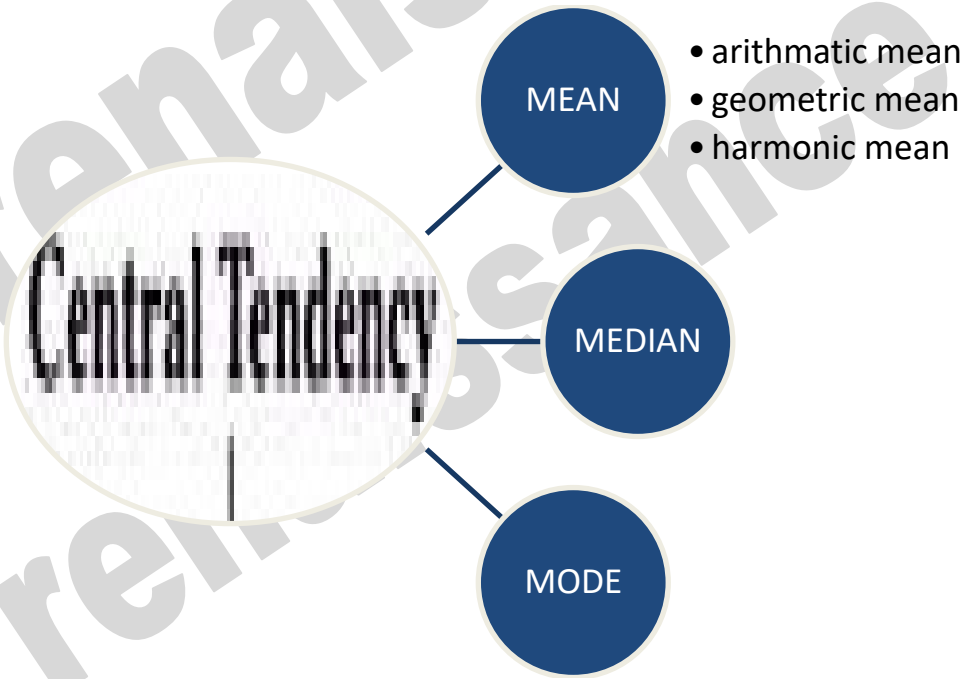
UNIT-II
MEASURES OF CENTRAL TENDENCY

The point around which the observations concentrate in general in the central part of the data is called central value of the data and the tendency of the observations to concentrate around a central point is known as Central Tendency.

Objects of Statistical Average:
- To get a single value that describes the characteristics of the entire group
- To facilitate comparison

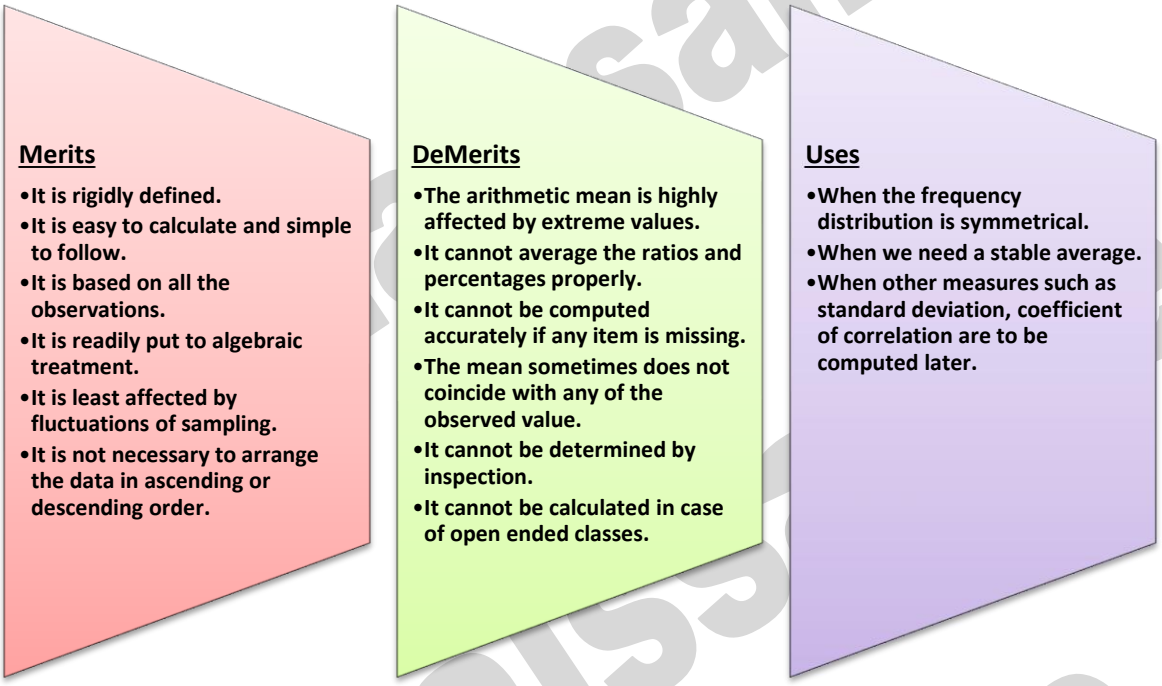
Functions of Statistical Average:
- Gives information about the whole group
- Becomes the basis of future planning and actions
- Provides a basis for analysis
- Traces mathematical relationships
- Helps in decision making

Requisites of an Ideal Average:
- Simple and rigid definition
- Easy to understand
- Simple and easy to compute
- Based on all observations
- Least affected by extreme values
- Least affected by fluctuations of sampling
- Capable of further algebric treatment



ARITHMETIC MEAN ($\bar{X}$)

Arithmetic Mean of a group of observations is the quotient obtained by dividing the sum of all observations by their number. It is the most commonly used average or measure of the central tendency applicable only in case of quantitative data. Arithmetic mean is also simply called "mean". Arithmetic mean is denoted by $\bar{X}$.



- **Rules of Including to Excluding And Unequal to Equal for Mean, Median and Mode**

Particulars	Mean	Median/Partition Values	Mode
Including to Excluding	No	Yes	Yes
Unequal to Equal	No	No	Yes

MEAN

1. ARITHMETIC MEAN ($\bar{X}$):

A) Individual Series:-

(i) Direct Method:-

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

(ii) Short-cut Method (Assumed Mean Method):-

$$\bar{X} = a + \frac{\sum dx}{N}$$

(iii) Step-deviation Method:-

$$\bar{X} = a + \frac{\sum ds}{N} \times i$$

x	$dx = (x - a)$	$ds = (dx / i)$
. a . .	. . . .	. . . .
$\sum x$	$\sum dx$	$\sum ds$

B) Discrete and Continuous Series:

(i) Direct Method:-

$$\bar{X} = \frac{\sum fx}{\sum f}$$

(ii) Short-cut Method (Assumed Mean Method):-

$$\bar{X} = a + \frac{\sum fdx}{\sum f}$$

(iii) Step-deviation Method:-

$$\bar{X} = a + \frac{\sum fds}{\sum f} \times i$$

For Discrete Series:

x	f	fx	$dx = (x - a)$	fdx	$ds = (dx / i)$	fds
. a . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .
	$\sum f$	$\sum fx$		$\sum fdx$		$\sum fds$

For Continuous Series:

Class-Interval	f	$x =$ $(L_1+L_2)/2$	fx	$dx = (x - a)$	fdx	$ds = (dx / i)$	fds
.	.	.	.	.	.	.	.
.	.	a	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
	$\sum f$		$\sum fx$		$\sum fdx$		$\sum fds$

2. COMBINED MEAN ($\bar{X}_{123..}$):

$$\bar{X}_{123...} = \frac{\bar{X}_1 N_1 + \bar{X}_2 N_2 + X_3 N_{3.....}}{N_1 + N_2 + N_{3.....}}$$

3. WEIGHTED MEAN ($\bar{X}_w$):

$$\bar{X}_w = \frac{\sum wx}{\sum w}$$

x	w	wx
.	.	.
.	.	.
.	.	.
.	.	.
	$\sum w$	$\sum wx$

Here, N = No. of observations, L_1 = Lower limit, L_2 = Upper limit,
a = Assumed Mean, $d_x = x - a$, i = Common Factor, $d_x' = d_x / i$, w = Weight.

MEDIAN (M)

The median is that value of the variable which divides the group into two equal parts, one part comprising of all values greater and other of all values less than the median. For calculation of median the data has to be arranged in either ascending or descending order. Median is denoted by **M**.

Merits

- It is easily understood and easy to calculate.
- It is rigidly defined.
- It can sometimes be located by simple inspection and can also be computed graphically.
- It is positional average therefore not affected at all by extreme observations.
- It is only average to be used while dealing with qualitative data like intelligence, honesty etc.
- It is especially useful in case of open end classes since only the position and not the value of items must be known.
- It is not affected by extreme values.

DeMerits

- For calculation, it is necessary to arrange data in ascending or descending order.
- Since it is a positional average, its value is not determined by each and every observation.
- It is not suitable for further algebric treatment.
- It is not accurate for large data.
- The value of median is more affected by sampling fluctuations than the value of the arithmetic mean.

Uses

- When there are open-ended classes provided it does not fall in those classes.
- When exceptionally large or small values occur at the ends of the frequency distribution.
- When the observation cannot be measured numerically but can be ranked in order.
- To determine the typical value in the problems concerning distribution of wealth etc.

MEDIAN

A) Individual Series:-

Arrange the data in ascending / descending order.

$$m = \left(\frac{N+1}{2}\right)$$

- If m is integer:
  then, $M = m^{th}$ term
- If m is fractional i.e. $m = a.b$

  then, $M = a^{th}\text{term} + b\left\{(a+1)^{th}\text{ term} - a^{th}\text{term}\right\}$

Here, N = No. of observation

B) Discrete Series:

1. Arrange the data in ascending / descending order.
2. Calculate cumulative frequency.

$$m = \left(\frac{N+1}{2}\right)$$

The value for which the cumulative frequency includes m^{th} value will be taken as MEDIAN (M).

Here, N = Sum of frequency ($\sum f$).

x	f	cf (Cumulative frequency)
.	.	.
.	.	.
.	.	.
.	.	.
	$\sum f$	

C) Continuous Series:

$$m = \left(\frac{N}{2}\right)^{th} \text{ term}$$

$$M = L_1 + \frac{m - pcf}{f} \times i$$

Here, N = Sum of frequency ($\sum f$), L_1 = Lower limit, L_2 = Upper limit,
pcf = Cumulative frequency of previous class, f =Frequency of median group,
$i = L_2 - L_1$.

Class-Interval	f	cf
.	.	.
.	.	.
.	.	.
.	.	.
	$\sum f$	

PARTITION VALUES

QUARTILES:

A) Individual Series:-

Arrange the data in ascending / descending order.

First quartile:

$$q_1 = \left(\frac{N+1}{4}\right)$$

- If q_1 is integer:

  then, $Q_1 = q_1^{th}$ term

- If q_1 is fractional i.e. $q_1 = a.b$

  then, $Q_1 = a^{th}\text{term} + b\{(a+1)^{th}\text{ term} - a^{th}\text{term}\}$

Here, N = No. of observation

Third quartile:

$$q_3 = 3\left(\frac{N+1}{4}\right)$$

- If q_3 is integer:

  then, $Q_3 = q_3^{th}$ term

- If q_3 is fractional i.e. $q_3 = a.b$

  then, $Q_3 = a^{th}\text{term} + b\{(a+1)^{th}\text{ term} - a^{th}\text{term}\}$

Here, N = No. of observation

B) Discrete Series:

1. Arrange the data in ascending / descending order.
2. Calculate cumulative frequency.

$$q_1 = \left(\frac{N+1}{4}\right) \quad \& \quad q_3 = 3\left(\frac{N+1}{4}\right)$$

The value for which the cumulative frequency includes q_1^{th} & q_3^{th} value will be taken as Q_1 & Q_3.

Here, N = Sum of frequency ($\sum f$).

x	f	cf (Cumulative frequency)
.	.	.
.	.	.
.	.	.
.	.	.
	$\sum f$	

C) Continuous Series:

$$q_1 = \left(\frac{N}{4}\right)^{th} \text{ term} \quad \& \quad q^3 = 3\left(\frac{N}{4}\right)^{th} \text{ term}$$

$$Q_1 = L_1 + \frac{q_1 - pcf}{f} \times i \quad \& \quad Q_3 = L_1 + \frac{q_3 - pcf}{f} \times i$$

Here, N = Sum of frequency ($\sum f$), L_1 = Lower limit, L_2 = Upper limit,

pcf = Cumulative frequency of previous class, f =Frequency of quartile group,

$i = L_2 - L_1$.

Class-Interval	f	cf
. . . .	. . . .	. . . .
	$\sum f$	

SEPTILES:- Septiles are S_1 to S_8

OCTILES:- Octiles are O_1 to O_7

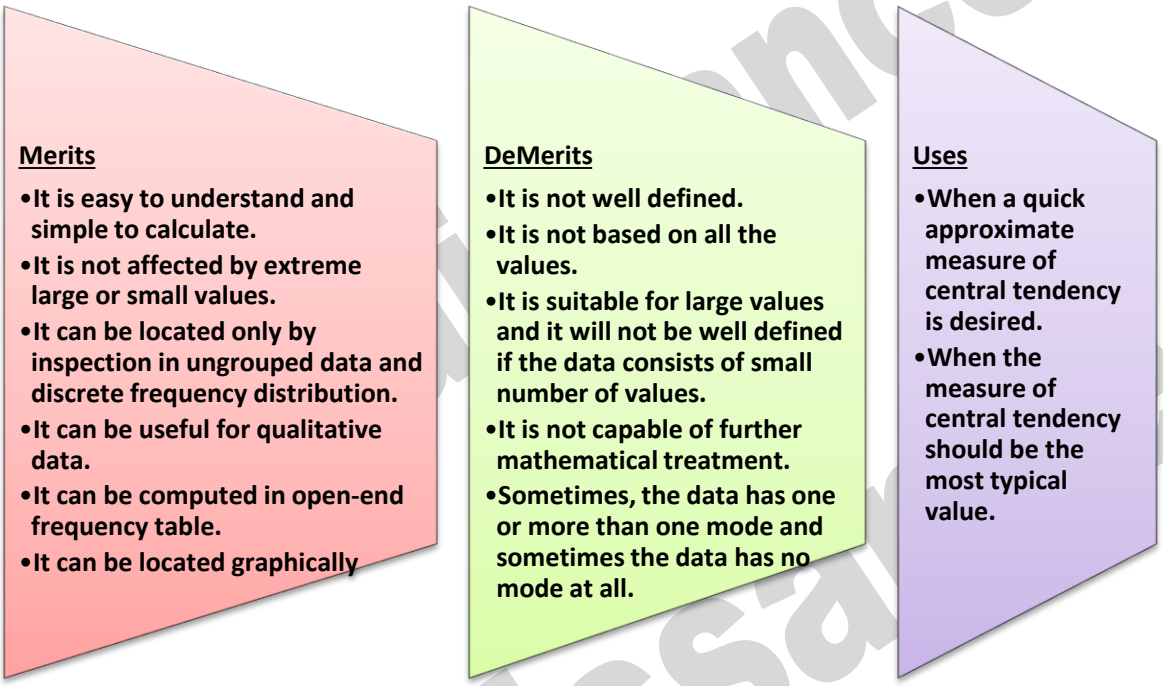
DECILES:- Deciles are D_1 to D_9

PERCENTILES:- Percentiles are P_1 to P_{99}.

Septiles, Octiles, Deciles and Percentiles formulas are same as quartiles.

MODE (Z)

Mode is the value which occurs the greatest number of times in the data. The word mode has been derived from the French word **'La Mode'** which implies fashion. The Mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values. Mode is denoted by **Z.**



MODE

A) Individual Series:-

1. By Inspection Method:- When the number of observation is small, mode is obtained by looking which one of the observation occurs most frequently.

2. By Discrete Series:- When the number of observation is large, convert the individual series into discrete series and locate mode accordingly.

B) Discrete Series:

1. Inspection Method:- When there is a regularity and homogeneity in the series, then there is a mode which can be located by looking into the frequency column for having maximum frequency.

2. Grouping Method:- When there is a regularity and homogeneity in the frequency distribution, i.e., the frequencies increase or decrease in haphazard way or two or more frequencies are equal then it is not obvious that which one is the maximum frequency. In such case, we use the method of grouping to decide which one may be considered as maximum frequency. This method involves the following steps:

(i) Prepare grouping table,
(ii) Prepare analysis table,
(iii) Find mode.

i. Grouping Table

x	Frequency or Sum of Frequencies					
	1	2	3	4	5	6
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Column 1: Given frequencies.
Column 2: The given frequencies are added in two's.
Column 3: The given frequencies are added in two's leaving out the first frequency.
Column 4: The given frequencies are added in three's.
Column 5: The given frequencies are added in three's leaving out the first frequency.
Column 6: The given frequencies are added in three's leaving out the first two frequencies.

After making these columns, the maximum frequency or sum of the frequencies in each column is written.

ii. Analysis Table

Column No.	Size of item containing maximum frequency					
1 2 3 4 5 6						
No. of Items						

iii. Location of mode:- The value of the variable which occurs maximum number of times in the analysis table is mode.

C) Continuous Series:-

The process of computing mode in case of a grouped frequency distribution with the help of a formula involves the following steps:

1. Determine the modal class. The class having maximum frequency is called modal class. This is done either by inspection or by grouping method (Same as discrete series).
2. Determine the value of mode by applying the formula:

$$Z = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

3. When modal class is determined with the help of grouping method, we may find mode outside the modal class (i.e., modal class may not have maximum frequency). In such situation to determine the mode in the modal class, we use the formula:

$$Z = L_1 + \frac{f_2}{f_0 + f_2} \times i$$

Here, Z = Mode, L_1 = Lower limit of the modal class, L_2 = Upper limit of the modal class,
f_1 = Frequency of the modal class, f_0 = Frequency of the pre-modal class,
f_2 = Frequency of the class succeeding the modal class, $i = L_1 - L_2$

4. When mode can not be calculated then,

$$Z = 3Median - 2Mean$$

GEOMETRIC MEAN (G.M)

The geometric mean also called geometric average is the nth root of the product of n non-negative quantities. Geometric Mean is denoted by **G.M**.

Properties of Geometric Mean:

- The geometric mean is less than arithmetic mean, G.M<A.M
- The product of the items remains unchanged if each item is replaced by the geometric mean.
- The geometric mean of the ratio of corresponding observations in two series is equal to the ratios their geometric means.
- The geometric mean of the products of corresponding items in two series.

Merits of Geometric Mean:

- It is rigidly defined and its value is a precise figure.
- It is based on all observations.
- It is capable of further algebraic treatment.
- It is not much affected by fluctuation of sampling.
- It is not affected by extreme values.

Demerits of Geometric Mean:

- It cannot be calculated if any of the observation is zero or negative.
- Its calculation is rather difficult.
- It is not easy to understand.
- It may not coincide with any of the observations.

Uses of Geometric Mean:

- Geometric Mean is appropriate when:

- Large observations are to be given less weight.
- We find the relative changes such as the average rate of population growth, the average rate of intrest etc.
- Where some of the observations are too small and/or too large.

- Also used for construction of Index Numbers.

GEOMETRIC MEAN (G.M.)

A) Individual Series:-

$$G.M. = Anti\log\left(\frac{\sum \log x}{N}\right)$$

x	$\log x$
.	.
.	.
.	.
.	.
	$\sum \log x$

B) Discrete and Continuous Series:

$$G.M. = Anti\log\left(\frac{\sum f \log x}{\sum f}\right)$$

For Discrete Series:

x	f	$\log x$	$f \log x$
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
	$\sum f$		$\sum f \log x$

For Continuous Series:

Class-Interval	f	$x =$ $(L_1+L_2)/2$	$\log x$	$f \log x$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
	$\sum f$			$\sum f \log x$

C) Calculation of Average Rate:

$$\frac{r}{100} = AL.\left[\frac{\log P_n - \log P_0}{n}\right] - 1$$

Here, r = Rate, P_n = Value at the end of period, P_0 = Value at the beginning, n = Number of years

HARMONIC MEAN (H.M)

Harmonic mean is another measure of central tendency. Harmonic mean is also useful for quantitative data. Harmonic mean is quotient of "number of the given values" and "sum of the reciprocals of the given values". It is denoted by **H.M.**

Merits of Harmonic Mean:

- It is based on all observations.
- It not much affected by the fluctuation of sampling.
- It is capable of algebraic treatment.
- It is an appropriate average for averaging ratios and rates.
- It does not give much weight to the large items and gives greater importance to small items.

Demerits of Harmonic Mean:

- Its calculation is difficult.
- It gives high weight-age to the small items.
- It cannot be calculated if any one of the items is zero.
- It is usually a value which does not exist in the given data.

Uses of Harmonic Mean:

- Harmonic mean is better in computation of average speed, average price etc. under certain conditions.

HARMONIC MEAN (H.M.)

A) Individual Series:-

$$H.M. = \frac{N}{\sum(1/x)}$$

x	$1/x$
.	.
.	.
.	.
.	.
	$\sum(1/x)$

B) Discrete and Continuous Series:

$$H.M. = \frac{\sum f}{\sum(f/x)}$$

For Discrete Series:

x	f	f/x
.	.	.
.	.	.
.	.	.
.	.	.
	$\sum f$	$\sum(f/x)$

For Continuous Series:

Class-Interval	f	$x =$ $(L_1+L_2)/2$	f/x
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
	$\sum f$		$\sum(f/x)$

C) WEIGHTED HARMONIC MEAN (W.H.M.):

$$W.H.M. = \frac{\sum w}{\sum(w/x)}$$

Here, x = Speed / Rate, w= Distance / Quantity, $\text{Speed} = \frac{\text{Distance}}{\text{Time}}$

UNIT-III

DISPERSION

The Dispersion (Known as Scatter, spread or variations) measures the extent to which the items vary from some central value. The measures of dispersion is also called the average of second order (Central tendency is called average of first order).

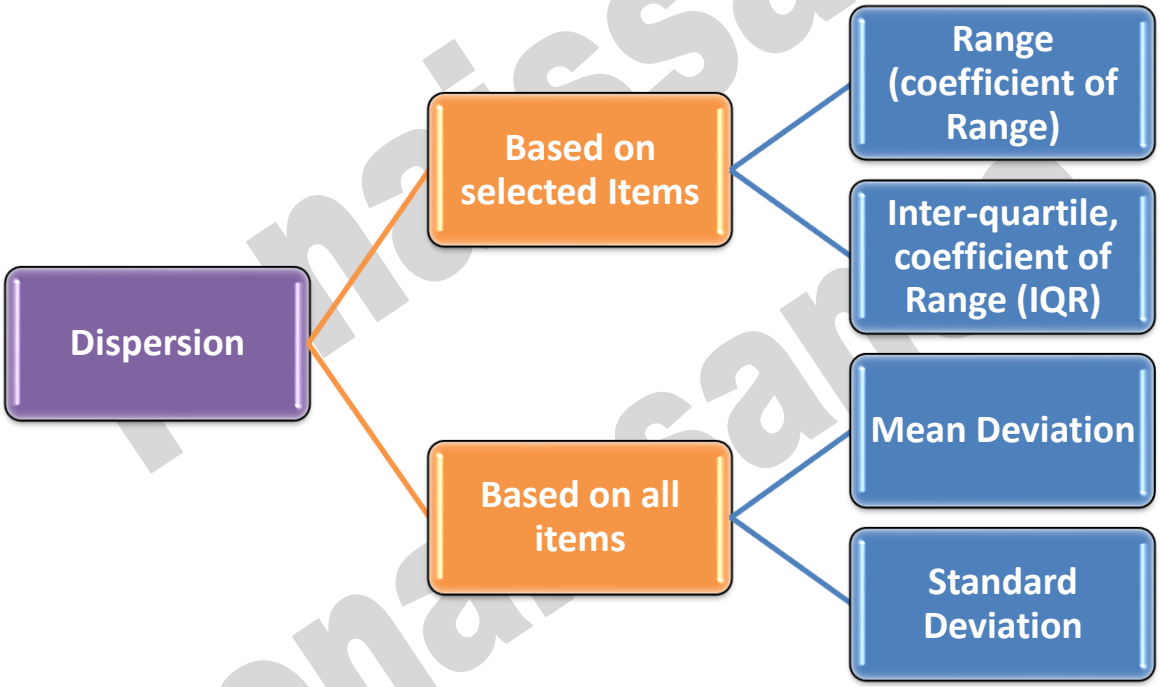
The two distributions of statistical data may be symmetrical and have common means, median or mode, yet they may differ widely in the scatter or their values about the measures of central tendency.

Significance/ objectives of Dispersion-

- To judge the reliability of average
- To compare the two an more series
- To facilitate control
- To facilitate the use of other statistical measures.

Properties of good Measure of Dispersion

- Simple to understand
- Easy to calculate
- Rigidly defined
- Based on all items
- Sampling stability
- Not unduly affected by extreme items.
- Good for further algebraic treatment



1. **Range: -** Range (R) is defined as the difference between the value of largest item and value of smallest item included in the distributions. Only two extreme of values are taken into considerations. It also does not consider the frequency at all series.
2. **Quartile Deviation: -** Quartile Deviation is half of the difference between upper quartile (Q3) and lower quartile (Q1). It is very much affected by sampling distribution.
3. **Mean Deviation: -** Mean Deviation or Average Deviation (δAlpha) is arithmetic average of deviation of all the values taken from a statistical average (Mean, Median, and Mode) of the series. In taking deviation of values, algebraic sign + and – are also treated as positive deviations. This is also known as first absolute moment.
4. **Standard Deviation:-** The standard deviation is the positive root of the arithmetic mean of the squared deviation of various values from their arithmetic mean. The S.D. is denoted as σ Sigma.

Distinction between mean deviation and standard deviation

Base	Mean Deviation	Standard Deviation
Algebric Sign	Actual +, - Signs are ignored and all deviation are taken as positive	Actual signs +, - are not ignored whereas they are squared logically to be ignored.
Use of Measure	Mean deviation can be computed from mean, median, mode	Standard deviation is computed through mean only
Further algebraic Treatment	It is not capable of further algebraic treatment.	It is capable of further algebraic treatment
Simplicity	M.D is simple to understand and easy to calculate	S.D is somewhat complex than mean deviation.
Based	It is based on simple average of sum of absolute deviation	It is based on square root of the average of the squared deviation

Variance

The square of the standard deviation is called variance. In other words the arithmetic mean of the squares of the deviation from arithmetic mean of various values is called variance and is denoted as σ2. Variance is also known as second movement from mean. In other way, the positive root of the variance is called S.D.

Coefficient of Variations- To compare the dispersion between two and more series we define coefficient of S.D. The expression is $\frac{\sigma}{\bar{x}}$ x 100 = known as coefficient of variations.

Interpretation of Coefficient of Variance-

Value of variance	Interpretation
Smaller the value of σ2	Lesser the variability or greater the uniformity/ stable/ homogenous of population
Larger the value of σ2	Greater the variability or lesser the uniformity/ consistency of the population

DISPERSION

RANGE = R

Individual Series	Discrete Series	Continuous Series
Range = L-S Where L=Largest, S=Smallest Observation	$R = L - S$	$R = L - S$
Coefficient of Range $\frac{L-S}{L+S}$	$\frac{L-S}{L+S}$	$\frac{L-S}{L+S}$

QUARTILE DEVIATION - Q.D.

Individual Series	Discrete Series	Continuous Series
$Q.D. = Q_3 - Q_1$	$Q.D. = Q_3 - Q_1$	$Q.D. = Q_3 - Q_1$
Coefficient of Q.D. $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$	$= \frac{Q_3 - Q_1}{Q_3 + Q_1}$	$= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

MEAN DEVIATION - M.D. δ **("Through actual Mean, Mode, Median)**

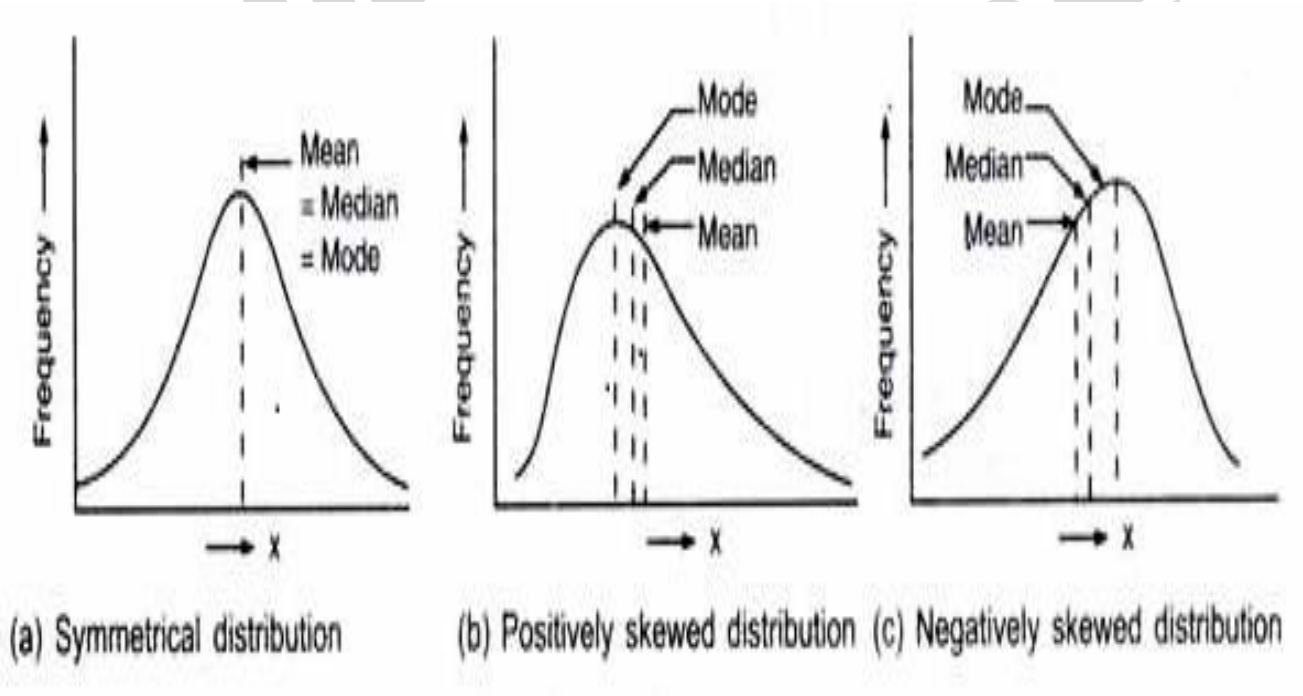
Individual Series	Discrete Series	Continuous Series
$\delta M\,(Median) = \frac{\sum dM}{N}$	$\frac{\sum fdM}{N}$	$\frac{\sum fdM}{N}$
Coefficient of $\delta = \frac{\delta}{M}$	$\delta = \frac{\delta}{M}$	$\delta = \frac{\delta}{M}$
Mean $\delta\bar{X} = \frac{\sum dx}{N}$	$\frac{\sum fd\bar{x}}{N}$	$\frac{\sum fd\bar{x}}{N}$
Coefficient of $\bar{X} = \frac{\delta}{\bar{X}}$	$\frac{\delta}{\bar{X}}$	$\frac{\delta}{\bar{X}}$
(Mode) $\delta Z = \frac{\sum dz}{N}$	$\frac{\sum fdz}{N}$	$\frac{\sum fdz}{N}$
Coefficient of δZ $\frac{\delta}{Z}$	$\frac{\delta}{Z}$	$\frac{\delta}{Z}$

Standard Deviation = σ **can be calculated through mean only**

	Individual Series	Discrete Series	Continuous Series
Direct (Through actual mean)	$\sqrt{\frac{\sum d_x 2}{N}}$	$\sqrt{\frac{\sum fd^2}{\sum f}}$	$\sqrt{\frac{\sum fd^2}{\sum f}}$
Indirect (Through assumed mean)	$\sqrt{\frac{\sum dx^2}{N} - \left(\frac{\sum dx}{N}\right)^2}$	$\sqrt{\frac{\sum fdx^2}{\sum f} - \left(\frac{\sum fdx}{\sum f}\right)^2}$	$\sqrt{\frac{\sum fdx^2}{\sum f} - \left(\frac{\sum fdx}{\sum f}\right)^2}$

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.



Skewness is positive if the tail on the right side of the distribution is longer or fatter than the tail on the left side. The mean and median of positively skewed data will be greater than the mode. Skewness is negative if the tail of the left side of the distribution is longer or fatter than the tail on the right side. The mean and median of negatively skewed data will be less than the mode. If the data graph symmetrically, the distribution has zero skewness, regardless of how long or fat the tails are.

Karl Pearson developed two methods to find skewness in a sample:

1. Pearson's Coefficient of Skewness #1 uses the mode. The formula is:

$$Sk_1 = \frac{\bar{X} - Mo}{s}$$

Where $\bar{X}$ = the mean, Mo = the mode and s = the standard deviation for the sample.

2. Pearson's Coefficient of Skewness #2 uses the median. The formula is:

$$Sk_2 = \frac{3(\bar{X} - Md)}{s}$$

Where $\bar{X}$ = the mean, Mo = the mode and s = the standard deviation for the sample.
It is generally used when you don't know the mode.

CORRELATION

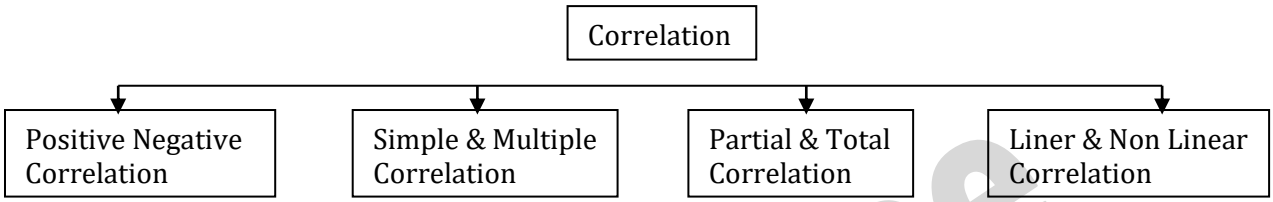
Introduction

1. Correlation is a statistical tool & it enables us to measure and analyse the degree or extent to which two or more variable fluctuate/vary/change w.e.t. to each other.
2. For example – Demand is affected by price and price in turn is also affected by demand. Therefore we can say that demand and price are affected by each other & hence are correlated. the other example of correlated variable are –
3. While studying correlation between 2 variables use should make clear that there must be cause and effect relationship between these variables. for e.g. – when price of a certain commodity is changed (↑ or ↑) its demand also changed (↑ or ↑) so there is case & effect relationship between demand and price thus correlation exists between them. Take another eg. where height of students; as well as height of tree increases, then one cannot call it a case of correlation because neither height of students is affected by height of three nor height of tree is affected by height of students, so there is no cause & effect relationship between these 2 so no correlation exists between these 2 variables.
4. In correlation both the variables may be mutually influencing each other so neither can be designated as cause and the other effect for e.g. –
   Price ↑→ Demand ↓
   Demand ↓→ Price ↑
   So, both price & demand are affected by each other therefore use cannot tell in real sense which one is cause and which one is cause and which one is effect.

DEFINITIONS OF CORRELATION

1. "If 2 or more quantities vary is sympathy, so that movements is one tend to be accompanied by corresponding movements in the other(s), then they are said to be correlated". **Connor.**
2. "Correlation means that between 2 series or groups of data there exists some casual correction". **WI King**
3. "Analysis of Correlation between 2 or more variables is usually called correlation." **A.M. Turtle**
4. "Correlation analysis attempts to determine the degree of relationship between variables. **Ya Lun chou**

TYPES OF CORRELATION

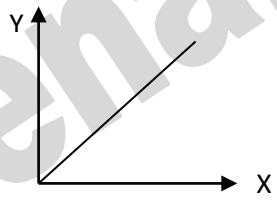
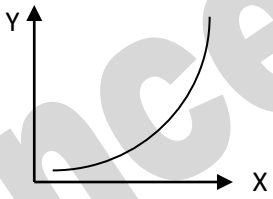
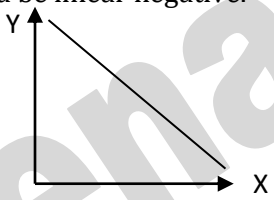
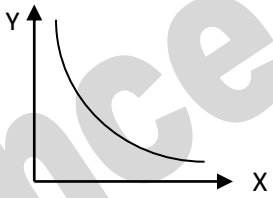


	POSITIVE CORRELATION	**NEGATIVE CORRELATION**
1	Value of 2 variables move in the same direction i.e. when increase/decrease in value of one variable will cause increase or decrease in value of other variable.	Value of 2 variables move in opposite direction i.e. when one variable increased, other variable decreases when one variable is decreased, other variable increase.
2	E.g. Supply & Price So, supply and price are .....correlated P = Price/Unit Q = quantity Supplied	E.g. Demand & Price So, Demand & Price vely correlated P = Price/Unit Q = quantity Supplied

	SIMPLE CORRELATION	**MULTIPLE CORRELATION**
1	In simple correlation, the relationship is confined to 2 variables only, i.e. the effect of only one variable is studied	The relationship between more than 2 variables is studied.
2	E.g. Demand & Price Demand depends on → Price This is case of simple correlation because relationship is confined to only one factor (that affects demand) i.e. price so we have to find correlation between demand & price. If, demand = Y If, demand – X Then, Correlation between Y & X	E.g. Demand & Price Demand depends on → Price Demand on → income This is case of multiple correlations because 2 factors (Price & Income) that affects demand are taken. We have to find correlation between demand & price. Demand & Price If, demand = Y Price = X_1 Price = X_2 Then Correlation between Y & X_1 Correlation between Y & X_2

SIMPLE CORRELATION	**MULTIPLE CORRELATION**
In partial correlation though more than 2 factors are involved but correlation is studies only between to be constant. E.g. Y → X_1, X_2 Y = Demand X_1 = Price X_2 = Income	In total correlation relationship between all the variables is studied i.e., none of item is assumed to be constant E.g. Y → X_1, X_2 Y = Demand X_1 = Price X_2 = Income
If we study correlation between Y & X1 & assume X2 to be constant it is a case of partial	If we assume that income is not constant i.e. we study the effect of both price & income on

correlation. this is what we do in law of demand – assume factors other than price as constant (Ceteris paribus – Keeping other things constant)	demand, it is a case of total correlation. In other words, cataris paribus assumption is relaxed in this case.

	LINEAR CORRELATION	**NON-LINEAR CORRELATION**
1	In linear correlation, due to unit, change value of one variable there is constant change in the value of other variable. The graph for such a relationship is straight line. E.G. – If in a factory no of workers are doubled, the production output is also doubled, and correlation would be linear.	In non linear or curvilinear correlation, due to unit, change value of one variable, the change in the value of other variable is not constant. the graph for such a relationship is a curve. E.G. – The amount spent on advertisement will not bring the change in the amount of sales in the same ratio, it means the variation.
2	If the changed in 2 variables are in the same direction and in the constant ratio, it is linear positive correlation X: 2, 4, 6, 8 — Y: 3, 6, 9, 12 	If the change in 2 variables is in the same direction but not in constant ratio, the correlation is non linear positive. X: 50, 55, 60, 90, 100 — Y: 10, 12, 15, 30, 45 
3	If changes in 2 variables are in the opposite direction but in constant ratio, the correlation is linear negative. For eg. every 5% ↑ is price of a good is associated with 10% decrease in demand the correlation between price and demand would be linear negative. X: 2, 4, 6, 8, 10 — Y: 21, 18, 15, 12, 9 	If changes in 2 variables are in opposite direction and not in constant ratio, the correlation is non linear negative. For eg: - every 5% ↑in price of good is associated with 20% to 10%↓in demand, the correlation between price & demand would be non linear negative. X: 80, 55, 50, 90 — Y: 50, 60, 75, 130 

TYPE – 1 [BASED ON KARL PEARSON'S COFFICIENT OF CORRELATION]

Before use move to numerical, use understand the basic notions & concepts –

d_x = Deviations of xi value from mean = $(x_i - \bar{x})$

x = Mean of x value [Average of X values] = $\frac{\Sigma xi}{n}$

n = No. of observations

d_y = Deviation of y value from mean = $(y - \bar{y})$

$\bar{y}$ = Mean of y values = $\frac{\Sigma yi}{n}$

d^2_x = Square of deviation of x values = $(x_i - \bar{x})^2$

d^2_y = Square of deviation of x values = $(y_i - \bar{y})^2$

$d_x d_y$ = Product of deviations = $(x_i - \bar{x})(y_i - \bar{y})$

Covariance (x,y) $= \frac{\Sigma(xi - \bar{x})(yi - \bar{y})}{n}$

σ_x = Variance of x_i values = $\frac{\Sigma(xi - \bar{x})^2}{n}$

σ_y = Variance of y_i values = $\frac{(yi - \bar{y})^2}{n}$

r or r_{xy} = coefficient of correlation between x 7 y variables.

Direct Method for Karl Pearson's Coefficient of correlation

Deviation from actual mean method

Deviation from assumed mean method (Short Cut Method)

This method is used in the situation where mean of any series (x or y) is not in whole number, i.e. in decimal value. in this case it is advisable to take deviation from assumed mean rather than actual mean and then use the above formula.
In the above short cut method
Let, A = Assumed mean of X series
B = Assumed mean of y series
then $\Sigma d_x = \Sigma(x_i - A)$ & $\Sigma d_y = \Sigma(y_i - B)$ &
$\Sigma d_x{}^2 = \Sigma(x_i - A)^2$ & $\Sigma d_y{}^2 = \Sigma(y_i - B)^2$
$\Sigma d_x d_y = \Sigma(x_i - A)(x_i - B)$

REGRESSION ANALYSIS
The dictionary meaning of regression is "Stepping Back". The term was first used by a British Biometrician" Sir Francis Galton 1822 – 1911) is 1877. He found in his study the relationship between the heights of father & sons. In this study he described "That son deviated less on the average from the mean height of the race than their fathers, whether the father's were above or below the average, son tended to go back or regress between two or more variables in terms of the original unit of the data.

Meaning
Regression Analysis is a statistical tool to study the nature extent of functional relationship between two or more variable and to estimate the unknown values of dependent variable from the known values of independent variable.

Dependent Variables – The variable which is predicted on the basis of another variable is called dependent or explained variable (usually devoted as y)

Independent variable – The variable which is used to predict another variable called independent variable (denoted usually as X)

Definition

Statistical techniques which attempts to establish the nature of the relationship between variable and thereby provide a mechanism for prediction and forecasting is known as regression Analysis.

– Ya-lun-Chon"

Importance/uses of Regression Analysis

- Forecasting
- Utility in Economic and business area
- Indispensible for goods planning
- Useful for statistical estimates.
- Study between more than two variable possible
- Determination of the rate of change in variable
- Measurement of degree and direction of correlation
- Applicable in the problems having cause and effect relationship
- Regression Analysis is to estimate errors
- Regression Coefficient (b_{xy} & b_{yx}) facilitates to calculate of determination ® & coefficient or correlation (r)

Regression Lines

The lines of best fit expressing mutual average relationship between two variables are known as regression lines – there are two lines of regression

Why are two Regression lines –

1. While constructing the lines of regression of x on y is treated as independent variables where as 'x' is treated as treated as dependent variable. This gives most probable values of 'X' for gives values of y. the same will be there for y on x.

RELATIONSHIP BETWEEN CORRELATION & REGRESSION

1. When there is perfect correlation between two series (r = ± 1) the regression with coincide and there will be only one regression line.
2. When there is no correction (r = o)> Both the lines will cut each other at point.
3. Where there is more degree of correction, say (r = ± 70 or more the two regression line with be next to each other whereas when less degree of correction. Say (r=± 10 on less) the two regression line will be a parted from each other.

REGRESSION LINES AND DEGREE OF CORRELATION

DIFFERENCE BETWEEN CORRELATION AND REGRESSION ANALYSIS

The correlation and regression analysis, both, help us in studying the relationship between two variables yet they differ in their approach and objectives. The choice between the two depends on the purpose of analysis.

S.NO	BASE	CORRELATION	REGRESSION
1	MEANING	Correlation means relationship between two or more variables in which movement in one have corresponding movements in other	Regression means step ping back or returning to the average value, i.e., it express average relationship between two or more

			variables.
2	RELATIONSHIP	Correlation need not imply cause and effect relationship between the variables under study	Regression analysis clearly indicates the cause and effect relationship. the variable(s) constituting causes(s) is taken as independent variables(s) and the variable constituting the variable consenting the effect is taken as dependent variable.
3	OBJECT	Correlation is meant for co-variation of the two variables. the degree of their co-variation is also reflected in correlation. but correlation does not study the nature of relationship.	Regression tells use about the relative movement in the variable. We can predict the value of one variable by taking into account the value of the other variable.
4	NATURE	There may be nonsense correlation of the variable has no practical relevance	There is nothing like nonsense regression.
5	MEASURE	Correlation coefficient is a relative measure of the linear relationship between X and Y. It is a pure number lying between 1 and +1	The regression coefficient is absolute measure representing the change in the value of variable. We can obtain the value of the dependent variable.
6	APPLICATION	Correlation analysis has limited application as it is confined only to the study of linear relationship between the variables.	Regression analysis studies linear as well as non linear relationship between variables and therefore, has much wider application.

Why least square is the Best?

When data are plotted on the diagram there is no limit to the number of straight lines that could be drawn on any scatter diagram. Obviously many lines would not fit the data and disregarded. If all the points on the diagram fall on a line, that line certainly would the best fitting line but such a situation is rare and ideal. Since points are usually scatters, we need a criterion by which the best fitting line can be determined.

Methods of Drawing Regression Lines –

1. Free curve –
2. Regression equation x on y,
   X = a + by ……(1)
3. Regression equation y on x
   Y = a + bx
   Where
   'a' is that point where regression lines touches y axis (the value of dependent variable value when value or independent variable is zero)
   'b' is the slop of the said line (The amount of change in the value of the dependent variable per unit change)

Change in independent variable)

A and b constants can be calculated through –

Σ(x = a + by) (by multiplying 'Σ')

$\Sigma x = Na + b\Sigma y$ (1)

$\Sigma x (y = a + bx)$ (by multiplying Σx)
$\Sigma xy = \Sigma xa + b\Sigma x^2$ (2)

KINSDS OF REGRESSION ANALYSIS

1. Linear and Non- Linear Regression
2. Simple and Multiple Regression

FUNCTIONS OF REGRESSION LINES –

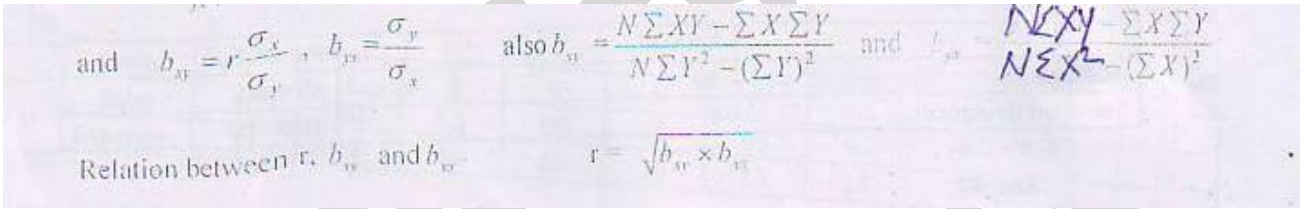
1. To make the best estimate –
2. To indicate the nature and extent of correlation

REGRESSION EQUATIONS –

The regression equation's express the regression lines, as there are two regression lines there are two regression equations –
Explanation is given in formulae –

REGRESSION LINES

1. Regression equation of x on y
   $X - X = b_{xy} (y - y)$
   Where b_{xy} = regression coefficient of X on Y
2. Regression euation of y on x
   $Y - Y = b_{xy} (x - x)$ where b_{xy} = regression coefficient of Y on X



REGRESSION ANALYSIS

Regression is based on two equations –

Equations	x on y	y on x
After elaborating them	$(x-\bar{x})=r\frac{\sigma x}{\sigma y}(y-\bar{y})$	$(y-\bar{y})=r\frac{\sigma y}{\sigma x}(x-\bar{x})$
Coefficient of Regression	$bxy=r\frac{\sigma x}{\sigma y}$	$byx=r\frac{\sigma y}{\sigma x}$
To find out coefficient of regression through actual mean	$bxy=\frac{\sum dxdy}{\sum dy^2}$	$byx=\frac{\sum dxdy}{\sum dx^2}$
through assumed mean	$bxy=\frac{\sum dxdy\times n-\sum dx\times\sum dy}{\sum dy^2\times N-(\sum dy)^2}$	$byx=\frac{\sum dxdy\times n-\sum dx\times\sum dy}{\sum dx^2\times N-(\sum dx)^2}$
$r=\sqrt{bxy\times byx}$		

REGRESSION COEFFICIENT – There are two regression coefficient like regression equation, they are (b_{xy} and b_{yx})

Properties of regression coefficients –

- Same sign – Both coefficient have the same either positive on negative
- Both cannot by greater than one – If one Regression is greater than "One" or unity. Other must be less than one.
- Independent of origin – Regression coefficient are independent of origin but not of scale.
- A.M.> 'r' – mean of regression coefficient is greater than 'r'
- R is G.M. – Correlation coefficient is geometric mean between the regression coefficient
- R, b_{xy} and b_{xy} – They all have same sign