1. INTRODUCTION TO STATISTICS

ECONOMICS

According to Alfred Marshall Economics is "the study of man in the ordinary business of life".

In economics we deal with-consumer, producer, seller, employer, and employee etc.

Economic activities are undertaken for a monetary gain. This is what economists mean by ordinary business of life.

Scarcity is the root of all economic problems, no scarcity, there would have been no economic problem.

Consumption, Production and Distribution

Economics involves the study of man engaged in economic activities of various kinds. The diverse economic activities are *consumption, production and distribution*.

STATISTICS IN ECONOMICS

To analyse economic problems we know more about economic facts. Such economic facts are also known as economic data. We try to find some measures that help to solve an economic problem. In Economics, such measures are known as policies. No analysis of an economic problem would be possible without data. In short we make use of statistical tools to formulate economics data, policies, reforms, economic laws etc.

WHAT IS STATISTICS?

Statistics deals with the collection, analysis, interpretation and presentation of numerical data. It is a branch of mathematics and also used in the disciplines such as accounting, economics, management, physics, finance, psychology and sociology.

Statistics can be expressed in two senses:

Singular and Plural sense: In singular sense methods or techniques. In plural sense quantitative information or data

ECONOMIC DATA ARE QUANTITATIVE,

The data that can be measured by using a numerical number, eg: taxes, height, weight, prices of goods, income, marks in a subject etc.

ECONOMIC DATA ARE QUALITATIVE.

The data which cannot be measured in quantitative terms is called qualitative data.eg. Health, honesty, aptitude, gender, intelligence and so on

FUNCTIONS OF STATISTICS

1. It simplifies complexities
2. It presents the data in a definite form.
3. It presents the data in a precise form.
4. It helps condensation of data.
5. It enables comparison of data.
6. It helps in testing.
7. It helps in prediction.
8. It helps in formulation of policies.
9. It enlarges human experience.

LIMITATION OF STATISTICS

1. Statistics Studies Only Quantitative Data
2. Statistics does not study individual cases.
3. Statistical results are true only on an average
4. Statistics does not reveal the entire story of the problem.
5. Statistical methods can be used only by an Expert.

DISTRUST OF STATISTICS

1. By distrust we mean
2. Lack of confidence in statistical statements and statistical methods
3. Statistics can prove anything.
4. Whether statistics is good or bad depend on its use.

REASONS FOR DISTRUST OF STATISTICS

1. Lack of knowledge in statistical techniques
2. Negligence of the limitation of statistics
3. Biased thinking of the user

2. COLLECTION OF DATA

COLLECTION OF DATA

Collection of data may be defined as an investigation into a phenomenon that can be expressed in numbers For Example: Conducting a survey to finding out the number of illiterates in a village

IMPORTANT CONCEPTS

1. POPULATION or UNIVERSE

A Population or Universe refers to any collection of specific group of human being or non-human objects such as educational institution, time, geographical area, supply and demand of a commodity, etc. For Example: Suppose we want to study the average height of the students studying in a higher secondary school, we will consider the height of entire students studying in that higher secondary school. Total number of students in that School is the Population or Universe

2. SAMPLE

Sample is the representative portion of the population or Universe. For Example: Suppose we want to study the average height of the students studying in a school, then we will consider the height of a Portion of students studying in that school. That portion of students' taken as representatives from entire School is the sample

3. INVESTIGATOR

The Investigator is a person, institution, or Government that is responsible for conducting the statistical enquiry and data collection.

4. ENUMERATOR

Enumerator is the person who actually collects the data of enquiry. If investigator himself collects the data then he is both investigator and enumerator.

5. RESPONDENT

The person or an institution that provide information to the investigator or enumerator is called Respondent or Informant.

STAGES OF STATISTICAL ENQUIRY

1. PLANNING STAGE

Before starting data collection, an investigator has to:

- Define the Problem/Study.
- Define the purpose of study.
- Find out the various sources of information for the study.
- Find out the tool and techniques of data collection for study.
- Determine the units for data collection.

2. EXECUTION STAGE

In the Execution Stage,

- The investigator can collect data or he can appoint enumerators.
- The enumerator should be trained to conduct enumeration work.
- After data collection, the data collected should be systematically organised, presented, analysed and interpreted.

SOURCES OF DATA COLLECTION

1. **INTERNAL DATA**: Data that are collected within the organisation.
2. **EXTERNAL DATA**: Data that are collected outside the organisation. External sources are divided in to Primary and Secondary Sources

PRIMARY DATA SOURCE

Primary data are those data, which are collected for the first time by an investigator. They are original in character.

FOUR METHODS OF PRIMARY DATA COLLECTION

1. DIRECT PERSONAL INVESTIGATION

Under this method the investigator, collect the data personally. i.e. he approaches the object, conduct enquiry on the spot, collect information and so on.

2. INDIRECT ORAL INVESTIGATION

In this method the investigator, do not collect data directly instead he gets it through his enumerator.

3. INVESTIGATION THROUGH MAILED QUESTIONNAIRE

This method of investigation is done by the investigator sending questionnaire to the respondent.

4. INVESTIGATION THROUGH LOCAL REPORTS

In this method, data are through local agents or correspondent. They collect information in their own fashion according to their likes and dislikes.

MERITS AND DEMERITS OF THE ABOVE METHODS

	Direct Personal Investigation	Indirect Oral Investigation	Investigation Through Mailed Questionnaire	Investigation Through Local Reports
Reliability	High	Average	Low	Very Low
Objectivity	High	Average	Low	Very Low
Accuracy	High	Average	Low	Very Low
Universality	High	Average	Low	Very Low
Response	Good	Fair	Low	Very Low
Time, Energy and Money	Very High	High	Moderate	Low

SECONDARY SOURCES OF DATA

Secondary data are those data that are already collected by the others. Secondary Sources are of two types Published sources and unpublished sources

PUBLISHED SOURCES

1. Reports from International Publications
2. Reports of Commissions and Committees
3. Reports from Government Publications
4. Reports of Semi-Government Publications
5. Reports of Commercial Research
6. Reports of Educational Research
7. Reports of Trade Unions
8. Reports of Non-Governmental Organisation
9. Reports of Journals, Periodicals, Newspapers, Etc.

PRECAUTIONS TO BE TAKEN BEFORE USING SECONDARY DATA

1. The investigator should see whether the data available are suitable for enquiry.
2. The investigator should see whether the data are reliable.
3. The investigator should see whether the data collected by the investigator are recent in nature.
4. The investigator should see whether the investigation done by the investigator is prompt.

QUESTIONNAIRE AND SCHEDULES

If an investigator or enumerator himself fills the forms by asking questions to the respondents directly, the form is called Schedule and if an investigator or enumerator sent the forms to the respondents to get it filled, the form is called Questionnaire

	Schedules	Questionnaire
1	Filled by the interviewer	Filled by the respondent
2	More costly	Less costly
3	Good response	Low/poor response
4	Requires more time	Requires less time
5	used even the respondent is illiterate	used only when the respondent is literate
6	Information obtained are reliable	Information obtained are not reliable

STEPS IN THE CONSTRUCTION OF QUESTIONNAIRE AND SCHEDULES

1. Selection of types of Questions

The questions in the schedule or questionnaire can be of different types based on topics of study and the nature of data to be collected. Generally three types of questions are used

- **TWO WAY CHOICE QUESTION**

Do you like music? = Yes / No

- **MULTIPLE CHOICE QUESTION**

Which type of music do you like? = {Classical, Light Music, Western, and Rock}

- **OPEN QUESTION**

How do you go to office? = Can give any answer

2. ORDER OF QUESTIONS

- The questions should be arranged in some order.
- The first few questions should be interesting and attractive.
- The personal and intimate questions should be asked later.
- The questions should be arranged within the topic.
- There should not be a sudden jump.
- In asking questions, the following maxims should be observed.

SIMPLE TO COMPLEX

NEAR TO FAR

CONCRETE TO ABSTRACT

EASY TO DIFFICULT

3. QUESTION WORDINGS

- Questions should be framed in common language of the respondent so that they can easily understand.
- The questions should not be provoking i.e. it should not hurt the feelings of the respondents.
- Words used in framing questions as far as possible familiar to the respondents.
- The terms should be clear and definite.

PRE-TESTING OR PILOT STUDY OF THE QUESTIONNAIRE

After drafting, the questionnaire it should be tested through a pilot survey .a small sample is to be selected for this. The questionnaire should be tested to them and necessary changes have to be made before the actual test is conducted.

ESSENTIALS OF A GOOD QUESTIONNAIRES AND SCHEDULE

- The questions should be self-explanatory.
- The number of questions should be kept minimum
- The questions should be simple as far as possible.
- The questions should not add provocation
- The questions should be arranged logically.
- Personal questions should be avoided.
- The questions should be easily answered in Yes or No
- Instructions to fill the questionnaire must be given

Important National Agencies that Collects Statistical Data

1. NSSO (National Sample Survey Organisation),
2. RGI (Registrar General of India),
3. DGCIS (Directorate General of Commercial Intelligence and Statistics)
4. Labour Bureaus.

3. METHODS OF DATA COLLECTION

METHODS OF DATA COLLECTION

There are two methods of data collection. They are 1.Census Method 2.Sampling Method

CENSUS METHOD

In the case of census method the entire population or universe is studied. We take every item of the population for study. The inferences are drawn from the entire universe. The main demerits of census method are large number of enumerators have to be appointed to collect data by this method. This method is expensive in terms of money time and energy. This method requires large organization to carry the investigation.

SAMPLING METHOD

Sample is the representative portion of the population or Universe. The method of selecting sample from the entire population is called sampling. Sampling methods are classified in to two. They are Non-Probability Sampling and Probability Sampling.

NON-PROBABILITY SAMPLING

In non-probability sampling, the samples are selected at the discretion of the investigator. i.e. there is place for human judgement and have no theoretical basis for estimation

PROBABILITY SAMPLING

In probability sampling, the samples are selected not at the discretion of the investigator, but by means of certain procedures, which ensure every unit of study equal chance to be included in the sample. The following are the various methods of sampling

- Random Sampling
- Stratified Random Sampling
- Systematic Sampling
- Cluster Sampling

RANDOM SAMPLING

A simple random sample is a sample selected from the population in such a way that every member of the population has an equal chance of being selected and the selection of an individual unit does not influence the selection of any other. The selection is made purely on chance. The personal bias of the investigator will not be present in the sample selected. Generally, there are two methods of random sampling: - Lottery Method and Tippets Random Number Table Method

- **LOTTERY METHOD**

In this method, each unit of the population is written in separate piece of paper, put it in a jar and mixes well. The required number of samples is selected in such a way that every member of the population has an equal chance of being selected.

- **TIPPETS RANDOM NUMBER TABLES**

Random number tables have been devised to guarantee equal probability of selection of every individual unit in the population according to their listed serial number in the sampling frame. They are available either in a published form or can be generated by using appropriate software packages.

- **Merits of Random Sampling**

1. There is no possibility of personal bias
2. This is used when the population is very large.
3. There is high accuracy in data Collection.

- **Demerits of Random Sampling**

1. It is very difficult to list of all the items of the population.
2. If the size of the sample is small, the result may not be reliable.

STRATIFIED RANDOM SAMPLING

Stratified random sampling is a method of sampling that involves the division of a population into smaller groups known as strata. In stratified random sampling, the strata are formed based on members' shared attributes or characteristics. A random sample from each stratum is taken proportional to the stratum's size

SYSTEMATIC SAMPLING

The process of systematic sampling typically involves the following steps. First selecting a fixed starting point in the larger population and then obtaining subsequent observations by using a constant interval between samples taken.

CLUSTER SAMPLING

Cluster Sampling Consists in forming suitable clusters of unit . From this Cluster samples are selected using Random Sampling

NON-PROBABILITY SAMPLING

1. Judgement Sampling
2. Convenience Sampling
3. Purposive Sampling

In all the above methods required numbers of sample are select at the convenience of the investigator.

ERRORS IN SAMPLING

- **Sampling Error**

Sampling error refers to the difference between the sample estimate and the actual value of a population characteristic. This type of error occurs when one makes an observation from the sample taken from the population. It is possible to reduce the magnitude of sampling error by taking larger sample.

- **NON-SAMPLING ERROR**

Non-sampling errors are more serious than sampling errors

These include

1. Errors in Data Acquisition
2. Non-Response Errors
3. Sampling Bias
4. Errors in Presentation
5. Errors in Calculations

4. ORGANISATION OF DATA

The data collected from primary and secondary sources are raw or unclassified. Once the data are collected, the next step is to classify them for further statistical analysis. Classification brings order in the data

Classification, is arranging or organising things into groups or classes based on some criteria. The unclassified data is called is called raw data.

DIFFERENT TYPES OF CLASSIFICATIONS

- **CHRONOLOGICAL CLASSIFICATION**

The data are classified either in ascending or in descending order with reference to time such as years, quarters, months, weeks, etc.

- **SPATIAL CLASSIFICATION**

The data are classified with reference to geographical locations such as countries, states, cities, districts, etc.

- **QUALITATIVE CLASSIFICATION**

Some attributes for example, nationality, literacy, religion, gender, marital status, etc. cannot be measured. Classification of such attributes is called Qualitative Classification.

- **QUANTITATIVE CLASSIFICATION**

Some attributes for example, height, weight, age, income, marks of students, etc., are quantitative in nature. Classification of such attributes is called Quantitative Classification.

ATTRIBUTES AND VARIABLES

- **Attributes** are those data, which cannot be expressed or measured in Terms of quantity. They are Qualitative in nature. Beauty intelligence, affection, love, sympathy, etc. are the examples of attributes.
- **Variables** are those data, which can be expressed or measured in terms of Quantity. They are quantitative in nature. Length, breadth, width, volume, weight, rainfall wind, speed are the examples of Variables.

DISCRETE AND CONTINUOUS VARIABLES

- **DISCRETE VARIABLE**

A variable is said to be Discrete when it is expressed in whole numbers. There is no fractional values are used to express this variable. For example: Number of students in a class, Number of children in a family, Number of motorcars in a showroom, etc.

- **CONTINUOUS VARIABLE**

A variable is said to be Continuous when it is expressed not only in whole numbers but also in fractional values.

For example: Length, breadth, width, volume, weight, rainfall wind, speed, etc.

FREQUENCY

- Frequency of an observation means how many times that observation occurs in the raw data.
- Frequency Curve is a graphic representation of a frequency distribution.

RANGE

Range is the difference between the largest and smallest value of the distribution. It is calculated as RANGE = HV – LV

WHAT IS A FREQUENCY DISTRIBUTION?

- A frequency distribution is a comprehensive way to classify raw data of a quantitative variable.
- It shows how different values of a variable are distributed in different classes along with their corresponding class frequencies.
- Each class in a frequency distribution table is bounded by Class Limits.
- Class limits are the two ends of a class.
- The lowest value is called the *Lower Class Limit* and the highest value the *Upper Class Limit*.
- The Class Mid-Point or Class Mark is the middle value of a class. It lies halfway between the lower class limit and the upper class limit of a class.
- Class Mid-Point or Class Mark = (Upper Class Limit + Lower Class Limit)/2

INCLUSIVE AND EXCLUSIVE METHOD

There two methods of classifying the statistical data.

- **INCLUSIVE METHOD**

Under this method, the upper limit of one class is included in that class itself. Consequently, there is a gap between the upper limit and lower limit. For example;

Class	Frequency
10-19	5
20-29	9
30-39	25
40-49	15

- **EXCLUSIVE METHOD**

In this method the class interval are fixed and the upper limit of one class is not included in that class. It is included in the next class. That is, the topmost value of a class is not included in that class but it is included in the next class. For example;

Class	Frequency
10-20	5
20-30	9
30-40	25
40-50	15

CUMULATIVE FREQUENCY TABLE

There are two types of Cumulative. They are: 1. Less than Cumulative Frequency Table and 2. Greater than Cumulative Frequency Table

- **LESS THAN CUMULATIVE FREQUENCY TABLE**

A less than Cumulative Frequency Table gives the number of items having values less than the upper limit of each Class

- **GREATER THAN CUMULATIVE FREQUENCY TABLE**

A greater than Cumulative Frequency Table gives the number of items having values greater than the lower limit of each class.

		Less than Cumulative Frequency Table		Greater than Cumulative frequency table	
Class	Frequency	Less than (Upper limit)	**Cumulative Frequency**	Greater than (Lower limit)	**Cumulative Frequency**
0-10	5	Less than 0	**0**	Greater than 0	**60**
10-20	9	Less than 10	**5**	Greater than 10	**55**
20-30	25	Less than 20	**14**	Greater than 20	**46**
30-40	15	Less than 30	**39**	Greater than 30	**21**
40-50	6	Less than 40	**54**	Greater than 40	**6**
		Less than 50	**60**	Greater than 50	**0**

UNIVARIATE DISTRIBUTION

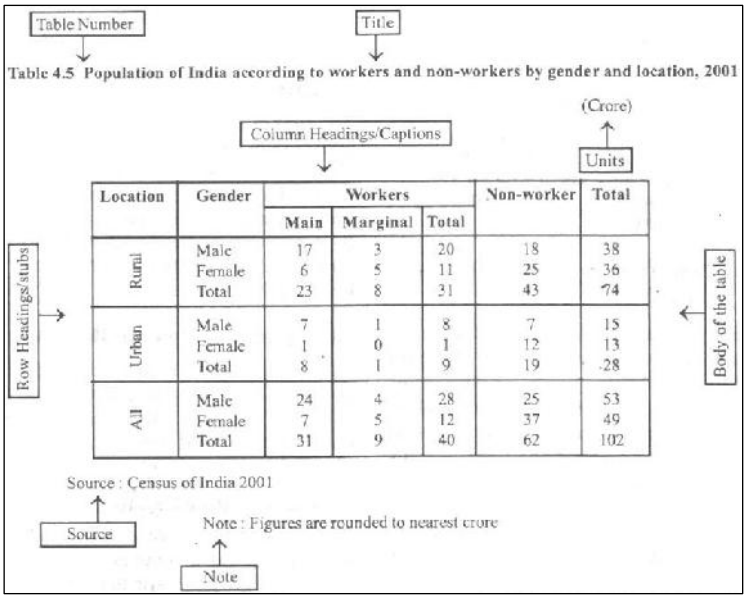
The frequency distribution of a single variable is called a univariate distribution

BIVARIATE DISTRIBUTION

A bivariate distribution is the frequency distribution of two variables.

PARTS OF A TABLE

1. Table number
2. Title
3. Caption or column heading
4. Stubs or row heading
5. Body of table
6. Unit of measurement
7. Source
8. Note



Location	Gender	Workers			Non-worker	Total
		Main	Marginal	Total		
Rural	Male	17	3	20	18	38
	Female	6	5	11	25	36
	Total	23	8	31	43	74
Urban	Male	7	1	8	7	15
	Female	1	0	1	12	13
	Total	8	1	9	19	28
All	Male	24	4	28	25	53
	Female	7	5	12	37	49
	Total	31	9	40	62	102

TYPES OF TABLE

1. SIMPLE TABLE – data are presented according to one characteristic only.
2. DOUBLE TABLE – data are presented about two interrelated characteristics of a particular variable.
3. THREE WAY TABLES – This table gives information regarding three interrelated characteristics of a particular variable.
4. MANIFOLD TABLE – This table explains more than three characteristics of the data.

5. MEASURES OF CENTRAL TENDENCY

☞ **Average:**

It is a value which is typical or representative of a set of data. Averages are also called Measures of Central Tendency.

☞ **Functions of Average:**

1. Presents complex data in a simple form.
2. Facilitates comparison.
3. Helps government to form policies.
4. Useful in Economic analysis.

☞ **Essentials of a good Average:**

1. Simple to calculate.
2. It should be easy to understand.
3. Rigidly defined.
4. Based on all items of observation.
5. Least affected by extreme values.
6. Capable of further algebraic treatment.
7. Least affected by sampling fluctuation.
8. Graphic measurement possible

☞ **Types of Average**

Generally, averages are classified into two. They are:

Mathematical Average and Positional Average

☞ **Mathematical Average:**

It includes Arithmetic Mean, Geometric Mean, and Harmonic Mean.

☞ **Positional Average:**

It includes Median, Mode, Quartile, Septiles, Deciles, Percentiles, etc.

☞ **Requisites of a Good Measure of Central Tendency**

1. It should be rigidly defined.
2. It should be capable of easy computation.
3. It should be capable easily understandable.
4. It should be based on all observations.
5. It should not be affected by extreme values.
6. It should not be affected by fluctuations in sampling.
7. It should be capable of further mathematical treatment
8. It should be capable of ascertaining graphical.

Arithmetic Mean

Arithmetic Mean is defined as the sum of the observation divided by number of observations. Let X_1, X_2, X_3 ….Xn. be n set of observations, then Arithmetic Mean

Individual Series	$AM = \bar{X} = \frac{X1 + X2 + X3 + \cdots + Xn}{N} = \frac{\sum X}{N}$
Discrete series	$AM = \bar{X} = = \frac{\sum fX}{\sum f}$
Continuous Series	$AM = \bar{X} = \frac{\sum fX}{\sum f}$
Continuous Series (Short Cut Method)	$AM = \bar{X} = A + \frac{\sum fd}{\sum f} xC$

☞ **Merits of Arithmetic Mean**

1. Arithmetic Mean is rigidly defined.
2. Arithmetic Mean is easy to compute.
3. Arithmetic Mean is easy to understand.
4. Arithmetic Mean is based on all observations
5. Arithmetic Mean is less affected by fluctuations in sampling.
6. Arithmetic Mean has several mathematical properties.

☞ **Demerits of Arithmetic Mean**

1. Arithmetic Mean is affected by extreme values.
2. Arithmetic Mean cannot be found out in open-end classes.
3. Arithmetic Mean cannot be found out in a class having unequal class interval.
4. Arithmetic Mean is not capable of ascertaining graphical

☞ **Properties of Arithmetic Mean**

1. The algebraic sum of deviations of items from arithmetic mean $\ddot{X}$ is always Zero	$\sum (X - \bar{X}) = 0$
2. The Sum of the squared deviations of the item from A.M. is minimum	$\sum (X - \bar{X})^2 < \sum (X - A)^2$

☞ **Numerical Problems**

Find the arithmetic mean (individual Series)

1. X: 22, 24, 26, 21, 20.
2. X: 7, 9, 4,6,5,3
3. X:40,50,60,25,35,55,65,75,85,90
4. X: 15, 20,25,30,10
5. X: 12,18,27,20,17,6
6. X: 7,20,19,32,38,17,4,1
7. X: 5, 8,6,7,4
8. X: 15, 20,25,24,12,31,71,52
9. X: 6,10,15,7,4
10. X: 3,8,21,37,61,130,26

Find the arithmetic mean (Discrete Series)

1.

X	1	2	3	4	5	6	7	8
f	5	6	6	10	7	5	4	3

2.

X	1	2	3	4	5	6
f	20	34	40	52	26	16

3.

X	1	2	3	4	5	6	7	8
f	5	6	6	10	7	5	4	3

4.

X	1	2	3	4	5	6	7	8	9
f	7	11	16	17	26	31	11	1	1

5.

X	4	10	15	17	20	23
f	3	5	8	13	4	1

Find the arithmetic mean (Continuous Series)

1.

X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	8	16	22	30	26	12	6

2.

X	0-10	10-20	20-30	30-40	40-50
f	2	5	15	4	4

3.

X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	5	10	15	20	15	10	5

4.

X	40-50	50-60	60-70	70-80	80-90	90-100	100-110
f	2	5	10	15	12	5	1

5.

X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	20	25	15	12	8	12	8

6.

X	0-10	10-20	20-30	30-40	40-50	50-60
f	2	5	10	7	4	2

7.

X	0-9	10-19	20-29	30-39	40-49	50-59	60-69
f	8	10	27	45	23	4	3

☞ **Word Problems in Arithmetic Mean**

1. The mean marks scored by 50 students are 60. Later it was found that one value was wrongly entered as 84 instead of 48. Find the correct mean
2. The mean marks scored by Gokul in Six subjects 80. Later it was found that one value was wrongly entered as 53 instead of 83 Find the correct mean
3. The mean wages paid to 100 workers is ₹180.4. Later it was found that wages of two workers were wrongly entered as ₹297 and ₹165 instead of ₹197 and ₹185 Find the correct mean wages
4. The mean wages paid to 150 workers is ₹360. Later it was found that wages of three workers were wrongly entered as ₹144, ₹158 and ₹190 instead of ₹165, ₹138 and ₹109. Find the correct mean wages
5. Raja's average marks in 5 test papers are 325. later it was found that three marks were wrongly entered as 53,17 and 80 instead of 83,70and18 find the correct mean

☞ **Combined Mean**

$$\bar{X}_{12=}\frac{N1.\bar{X}1 + N2.\bar{X}2}{N1 + N2}$$

☞ **Find Combined Mean**

N1	N2	$\bar{X}1$	$\bar{X}2$
25	75	18	14

$$\bar{X}_{12=} \frac{N1.\bar{X}1 + N2.\bar{X}2}{N1 + N2}$$

$$\bar{X}_{12=} \frac{25x18 + 75x14}{25 + 75}$$

$$\bar{X}_{12=} \frac{450 + 1050}{100}$$

$$\bar{X}_{12=} \frac{1500}{10}$$

$$\bar{X}_{12}=15$$

N1	N2	$\bar{X}1$	$\bar{X}2$
50	70	85	80

1. The average rain fall from Monday to Wednesday is 4 cms. the average rainfall for the remaining days is 6 cms find the average rain fall for the entire week
2. The mean salary for a group of 40 workers is ₹5200 per month and that for a group of 60 workers is ₹6800 per month. What is the combined mean salary?
3. The mean salary for a group of 100 workers is ₹5000 per month and that for a group of 300 workers is ₹6000 per month. What is the combined mean salary?

☞ **Find Missing Frequency**

- Given AM=28, find the Missing frequency

Class	Frequency	Mid point	fx
0-10	12	5	60
10-20	18	15	270
20-30	27	25	675
30-40	X	35	35X
40-50	17	45	765
50-60	6	55	330
	80+X		**2100+35X**

$$\text{AM} = \bar{X} = \frac{\sum fX}{\sum f}$$

$$28 = \frac{2100 + 35x}{80 + x}$$

$$28(80 + x) = 2100 + 35x$$

$$2240 + 28x = 2100 + 35x$$

$$2240 - 2100 = 35x - 28x$$

$$140 = 7x$$

$$x = \frac{140}{7} = 20$$

The missing frequency is 20

- Given AM=24, find the missing frequency

Class	Frequency
0-10	5
10-20	8
20-30	x
30-40	7
40-50	4

- Given AM=29, find the missing frequency

Class	Frequency
0-10	2
10-20	5
20-30	X
30-40	7
40-50	4
50-60	2

Median

As distinct from Arithmetic Mean, Median is not based on every observation. It is a positional average. The term position means the place of value in a series. Median is the value of the middle item when the items are arranged in ascending or descending order. Median divides the series in to

individual series	Discrete series	Continuous series
$\frac{n+1}{2}$	$\frac{n+1}{2}$	$\frac{n}{2}$
$\frac{n+1}{2}^{th}$ item	$\frac{n+1}{2}^{th}$ value	$Median = l1 + \frac{l2-l1}{f1}(m-c)$ *where l1 is the lower limit of median class* *l2 is the Upper limit of median class* *fi the frequency of median class* $m = n/2$ *c is the cumulative frequency of the class preceding to median class*

☞ **Merits of Median**

1. Median is rigidly defined.
2. Median is easy to compute.
3. Median is easy to understand.
4. Median is less affected by extreme values.
5. Median is capable of ascertaining graphical.
6. Median can be found out in open-end classes.
7. Median can be found out in a class having unequal class interval

☞ **Demerits of Median**

1. Median is not based on all observations
2. Median is affected by fluctuations in sampling
3. Median has no mathematical properties.

☞ **Quartiles**

Quartiles divide the series into four equal parts. The value of item, which divides the first half of the series, is called lower Quartiles or first Quartiles or Q1, which is less than the Median. The value of item, which divides the second half of the series in to two equal parts, is called upper Quartiles or third Quartiles or Q3, which is greater than Median The value of item, which divides the whole series in to two equal halves is called middle Quartiles or second Quartiles or Q2, which is equal to Median.

Procedure for finding out Quartiles, Deciles and Percentiles

Quartiles	Individual / Discrete	Continuous
Q1	¼ (N+1) th item	N/4
Q2	½ (N+1) th item	N/2
Q3	¾ (N+1) th item	3N/4
Deciles	**Individual / Discrete**	**Continuous**
D3	3/10 (N+1) th item	3N/10
D5	5/10(N+1) th item	5N/10
D8	8/10(N+1) th item	8N/10
Percentiles	**Individual / Discrete**	**Continuous**
P38	38/100 (N+1) th item	38N/100
P55	55/100(N+1) th item	55N/100
P69	69/100(N+1) th item	693N/100

(Use same procedure of finding median in all cases after locating Q1, Q2 Q3, Deciles and Percentile Item/class

☞ **Find Median (individual Series)**

1. 7 10 14 21 28 30 3 9 12
2. 67 53 59 56 35 20 56 55 87
3. 94 33 66 68 32 80 48 70
4. 12 16 17 21 28 19 30 32
5. 20 5 10 25 15 5 20
6. 2 4 5 7 3 2 1

☞ **Find Median (Discrete Series)**

1.

X	4	5	6	7	8	9	10
F	2	4	5	7	3	2	1

2.

X	3	4	5	6	7	8	9
F	3	4	8	4	8	3	2

3.

X	4	10	15	17	20	23
F	3	5	8	13	4	1

4.

X	5	10	15	20	25	30	35
F	2	3	4	6	10	5	2

5.

X	4	5	6	7	8	9	10	11
F	40	48	52	56	60	63	57	55

☞ **Find Median (Continuous Series)**

X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
F	8	16	22	30	24	12	6

X	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
F	5	8	7	12	28	20	10	10

X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
F	20	25	15	12	8	12	8

X	0-9	10-19	20-29	30-39	40-49	50-59	60-69
F	2	3	5	10	15	8	7

Mode

It is the value which occurs the most frequently in a series. Mode is represented by the letter 'Z'

☞ **Calculation of Mode**

Individual Series: By observation identify the value that occurs most frequently in a series or by conversion into discrete series and then identify the value corresponding to which there is highest frequency.

Discrete Series: By Inspection Method or by Grouping Method: By preparing Grouping Table and then preparing Analysis table.

Continuous Series: Determination of Modal class by Inspection Method or Grouping table and Analysis table. Then applying the formula

$$Z = l1 + \frac{f1 - f0}{2f1 - f0 - f2}(l2 - l1)$$

Where

l_1 is the lower limit of modal class

l_2 is the Upper limit of modal class

f_1 frequency of modal class

f_0 frequency of the class previous to modal class

f_2 frequency of the class following to modal class

☞ **Merits of Mode**

1. It is easy to understand and simple to calculate.
2. Not affected by extreme values.
3. Can be located graphically.
4. Easily calculated in case of open-end classes.

☞ **Demerits of Mode**

1. Not rigidly defined.
2. If mode is ill defined, mathematical calculation is complicated.
3. Not based on all items.
4. Not suited to algebraic treatment.

Relationship between Mean Median and Mode

1. In case of symmetrical distribution: Mean = Median = Mode

2. In case of asymmetrical distribution: Mode = 3 Median – 2 Mean (empirical formula)

Empirical Formula: Mode = 3 Median – 2 Mean

Numerical Problems – Mode

☞ **Find Mode (Individual Series)**

1. 61 64 63 61 63 64 60 65 64 64 63 64 62 50 18 30
2. 15 18 19 20 18 22 25 18 26 30
3. 8 9 11 16 23 21 27 35 21 7 21
4. 5 14 30 8 6 2
5. 20 24 32 25 26 24 32

☞ **Find Mode (Discrete Series)**

1.	X	4	5	6	7	8	9	10
	F	2	4	5	7	3	2	1
2.	X	3	4	5	6	7	8	9
	F	3	4	10	4	8	3	2
3.	X	4	10	15	17	20	23	
	F	3	5	8	13	4	1	
4.	X	5	10	15	20	25	30	35
	F	2	3	4	16	10	5	2
5.	X	4	5	6	7	8	9	10	11
	F	40	48	52	56	7	63	57	55

☞ **Find Mode (Continuous Series)**

1.	X	10-20	20-30	30-40	40-50	50-60	60-70	
	F	16	14	19	17	7	13	
2.	X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
	F	8	16	22	30	24	12	6
3.	X	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
	F	5	8	7	12	28	20	10	10
4.	X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
	F	20	25	15	12	8	12	8
5.	X	0-9	10-19	20-29	30-39	40-49	50-59	60-69
	F	2	3	5	10	15	8	7

6. MEASURES OF DISPERSION

MEASURES OF DISPERSION:-

Dispersion refers to the variation of the items around an average. According to Dr Bowley, "Dispersion is the measure of variations of items." To quote CONNOR: - "Dispersion is a measure of the extent to which the individual items vary."

☞ **Objectives of Dispersion:**

1. To determine the reliability of an average
2. To compare the variability of two or more series
3. It serves the basis of other statistical measures such as correlation etc.
4. It serves the basis of statistical quality control.

☞ **Properties of a good measure of dispersion:**

1. It should be easy to understand.
2. It should be simple to calculate
3. It should be uniquely defined.
4. It should be based on all observations.
5. It should not be unduly affected by extreme items.

☞ **MEASURES OF DISPERSION MAY BE EITHER ABSOLUTE OR RELATIVE.**

Absolute measures of dispersion are expressed in the some units in which data of the series are expressed i.e., rupees kgs, tons etc. whereas relative measures of dispersion are independent of the units of measurement. They are expressed in percentage these are used to compare two or more series which are expressed in different units.

- **Absolute measures of dispersion are:-**

1. Range
2. Quartile Deviation
3. Mean Deviation
4. Standard deviation and variance.

- **Relative measures of dispersion are:-**

1. Coefficient of Range.
2. Coefficient of Quartile Deviation
3. Coefficient of mean Deviation
4. Coefficient of standard Deviation
5. Coefficient of variation

- **Graphic method of studying dispersion, known as Lorenz curve**

RANGE

Range is the simplest measure of dispersion: - It is the difference between the largest and smallest value of the distribution. Range is calculated as

Range $= HV - LV$

Coefficient of Range $= \frac{HV - LV}{HV + LV}$

- **Merits of Range:-**

1. It is simple to understand and easy to calculate
2. It is widely used in statistical quality control.

- **Demerits of Range:-**

1. It is affected by extreme values in the series.
2. It cannot be calculated in case of open-ended series.
3. It is not based on all the items of the series.

INTER QUARTILE RANGE AND QUARTILE DEVIATION

Another measure of dispersion is Inter-quartile range is the difference between the upper quartile Q3 and lower quartile Q1. Quartile deviation or Semi inter quartile Range is half of the difference between the upper quartile and lower quartile i.e. half of the inter-quartile range.

Interquartile Range	$= Q_3 - Q_1$
Semi interquartile Range or Quartile Deviation	$= \frac{Q_3 - Q_1}{2}$
Co. efficient of Quartile Deviation	$= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

- **Merits of Quartile Deviation:-**

1. It is easy to compute.
2. It is less affected by extreme items
3. It can be computed in open-ended series.

- **Demerits of Quartile Deviation:-**

1. It ignores half i.e. 50% of the items.
2. It is useful only for rough study.
3. It is not based on all observations.

MEAN DEVIATION:-

It is defined as the arithmetic average of the absolute deviations (ignoring signs) of the various items from a measure of central tendency; i.e. mean or median. Generally, mean deviation is calculated from median because the sum of the absolute deviations taken from median is minimum or least.

- **MEAN DEVIATION FROM MEAN (INDIVIDUAL SERIES /UNGROUPED DATA)**

$$= \frac{\sum |X - \bar{X}|}{N}$$

- **MEAN DEVIATION FROM MEAN (DISCRETE/CONTINUOUS SERIES)**

$$= \frac{\sum f\,|X - \bar{X}|}{\sum f}$$

- **COEFFICIENT OF MEAN DEVIATION FROM MEAN**

$$= \frac{Mean\ Deviation\ from\ Mean}{\bar{X}}$$

- **Merits of mean Deviation:-**

1. It is based on all observations.
2. It is least affected by extreme items.
3. It is simple to understand and easy to calculate.

- **Demerits of mean Deviation:-**

1. It ignores ± signs in deviations
2. It cannot be computed with open - ended series.
3. It is not well defined measure because it is calculated from different averages (Mean, Median & Mode)
4. It is difficult to compute when mean or Median or Mode comes in fractions.

STANDARD DEVIATION

It is the most widely used measure of dispersion. It is defined as the positive square root of the arithmetic average of the squares of deviations taken from the mean. Variance is another measure of dispersion. It is represented by Greek letter 'σ' (small Sigma)

- **STANDARD DEVIATION (INDIVIDUAL SERIES/UN-GROUPED DATA)**

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

- **STANDARD DEVIATION (DISCRETE AND CONTINUOUS SERIES-DIRECT METHOD)**

$$\sigma = \sqrt{\frac{\sum f(X - \overline{X})^2}{\sum f}}$$

- **COEFFICIENT OF STANDARD DEVIATION**

$$= \frac{\sigma}{\overline{X}} x100$$

- **Merits of standard Deviation:-**

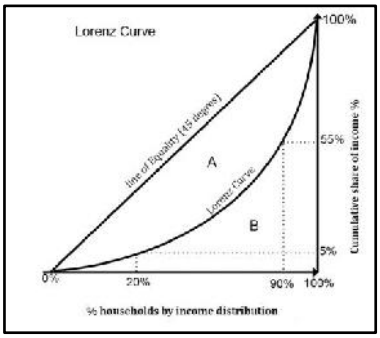
1. It is rigidly defined.
2. It is based on all observations whereas range and quartile- deviations are not based on all items.
3. It takes algebraic signs in consideration where as these are ignored in mean- Deviation.
4. It can be algebraically manipulated, i.e. we can find the combined S.D. of two or more series.
5. It serves the basis of other measures like correlation etc.

- **Demerits of standard deviation:-**

1. As compared to range and quartile deviation, it is difficult to understand and compute.
2. It gives more importance to extreme items.

LORENZ CURVE

It is a graphical method of measuring dispersion. It has great utility in the study of degree of inequality in the distribution of income and wealth between the countries. It is also useful for comparing the distribution of wages, profits etc. over different business groups. It is a cumulative percentage curve in which the percentage of frequency (persons or workers) is combined with the percentage of other items such as income, profits, wages etc.



RANGE AND COEFFICIENT OF RANGE

From the following data, calculate range and coefficient of range.(individual Series)

1.	X:	4	7	15	8	7	2	6	
2.	X:	20	34	40	52	26	16		
3.	X:	5	22	31	43	31	40	35	15	3
4.	X:	1200	700	200	250	350	2700		
5.	X:	5	15	28	24	17	10	1	

From the following data, calculate range and coefficient of range.(Discrete/ Continuous Series)

1	x	5	8	12	16	20	25	40
	f	7	8	10	15	6	9	5
2	x	1000	2000	3000	4000	5000	6000	7000
	f	12	16	2	10	15	7	3
3	x	60	61	62	63	64	65	66	67
	f	0	2	15	29	25	8	4	3
4	x	2-4	4-6	6-8	8-10	10-12	12-14		
	f	4	6	8	12	5	3		
5	x	4-8	8-12	12-16	16-20	20-24	24-28		
	f	5	18	35	46	72	17		
6	x	120-129	130-139	140-149	150-159	160-169	170-179		
	f	5	18	53	46	33	21		

QD AND COEFFICIENT OF QD

From the following data, calculate QD and coefficient of QD.(individual Series)

1.	X:	4	7	15	8	7	2	6		
2.	X:	20	34	40	52	26	16			
3.	X:	5	22	31	43	31	40	35	15	3
4.	X:	1200	700	200	250	350	2700			
5.	X:	5	15	28	24	17	10	1		

From the following data, calculate QD and coefficient of QD. (Discrete/ Continuous Series)

x	60	61	62	63	64	65	66	67
f	0	2	15	29	25	8	4	3

x	2-4	4-6	6-8	8-10	10-12	12-14
f	4	6	8	12	5	3

x	120-129	130-139	140-149	150-159	160-169	170-179
f	5	18	53	46	33	21

MEAN DEVIATION FROM MEAN

From the following data, calculate M.D from Mean and coefficient of M.D from Mean (Individual Series)

1.	X:	2	8	17	30	3		
2.	X:	5	10	12	15	20	4	
3.	X:	12	15	14	10	7	3	2
4.	X:	3	7	12	20	26	4	
5.	X:	6	10	26	28	30	32	

Calculate M.D from Mean and coefficient of M.D from Mean. (Discrete/ Continuous Series)

1	x	10	20	30	40	50	60	70
	f	8	15	30	36	14	5	3

2	x	50	55	60	65	70
	f	15	20	25	30	10

3	x	10	11	12	13	14
	f	3	12	18	12	3

4	x	0-20	20-40	40-60	60-80	80-100
	f	10	16	30	32	12

5	x	0-10	10-20	20-30	30-40	40-50
	f	4	6	10	8	2

STANDARD DEVIATION

From the following data, calculate Standard Deviation and coefficient of Variation (C V) (Individual Series)

1.	X:	40	44	54	60	62	64	70	80	90	96
2.	X:	2	4	7	8	6	3				
3.	X:	7	15	9	9						
4.	X:	2	8	10	7	3					
5.	X:	50	45	60	42	40	60	65	50	56	70

Calculate Standard Deviation and coefficient of Variation (C V).(Discrete/ Continuous Series)

1	x	6	7	8	9	10	11	12
	f	3	6	9	13	8	5	4
2	x	50	55	60	65	70		
	f	15	20	25	30	10		
3	x	10	11	12	13	14		
	f	3	12	18	12	3		

DRAW A LORENZ CURVE FROM THE DATA GIVEN BELOW

Income	100	200	300	400	500
No. of persons	80	75	50	30	20

WORD PROBLEMS

From the following table of marks obtained by two students Ram and Shyam. Find out who is more intelligent and consistent?

Ram	25	50	45	30	70	42	36	48	34	60
Shyam	10	70	50	20	95	55	42	60	48	80

Ram(X)	$X-\bar{X}$	$(X-\bar{X})^2$	Shyam (Y)	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
25	-19	361	10	-43	1849
50	6	36	70	17	289
45	1	1	50	-3	9
30	-14	196	20	-33	1089
70	26	676	95	42	1764
42	-2	4	55	2	4
36	-8	64	42	-11	121
48	4	16	60	7	49
34	-10	100	48	-5	25
60	16	256	80	27	729
440	0	1710	530	0	5928

$$\bar{X} = \frac{440}{10} = 44$$

$$\bar{Y} = \frac{530}{10} = 53$$

$$\sigma x = \sqrt{\frac{1710}{10}} = 13.07$$

$$CVx = \frac{13.07}{44} x100 = 29.7\%$$

$$\sigma y = \sqrt{\frac{5928}{10}} = 24.34$$

$$CVx = \frac{24.34}{53} x100 = 45.92\%$$

	Ram	Shyam	
AM	44	53	Shyam is more intelligent because AM is more for Shyam
CV	29.7%	45.92%	Ram is More Consistent because CV is less for Ram

2. Two Cricketers scored in 8 test matches as follows Find out who is having high run rate and consistent?

Player X	50	100	90	60	120	140	60	92
Player Y	20	140	160	100	40	120	80	90

7. CORRELATION ANALYSIS

- **MEANING OF CORRELATION:**

It studies and measures the intensity of relationship between two or more variables. If the two variables, X and Y change (vary) in such a way that with a change in value of one variable the values of the other variable also change, then they are said to be correlated.

- **SIGNIFICANCE OF CORRELATION:**

Correlation has immense utility in statistics.

1. It helps in determining the degree of relationship between variables.

2. We can estimate the value of one variable on the basis of the value of another variable correlation serves the basis of regression.

3. Correlation is useful for economists. An economist specifies the relationship between different variables like demand and supply, money supply and price level by way of the correlation.

TYPES OF CORRELATION

- **POSITIVE AND NEGATIVE CORRELATION:-**

Correlation is classified into positive and negative correlation when two variables move in the same direction, i.e. if the value of Y increases (or decreases) with an increase (or decrease) in the value of X, they are said to be variable 'X' increase (or decrease) with the decrease or increase in the value of Y variable, they one said to be negatively correlated.

- **LINEAR AND NON- LINEAR CORRELATION:-**

Correlation may be linear or non-linear. If the amount of change in one variable tends to have a constant relation with the amount of change in the other variable then the correlation is said to be liner. It is represented by a straight line. On the other hand if the amount of change in one variable does not have constant proportional relationship to the amount of change in the other variable, then the correlation is said to be non-linear or curvilinear.

- **SIMPLE, MULTIPLE AND PARTIAL CORRELATION:-**

Correlation may also be simple, multiple and partial correlation. When two variables are studied to determine correlation, it is called simple correlation on the other hand when more than two variables are studied to determine the correlation it is called multiple correlation. When correlation of only two variables is studied keeping other variables constant, it is called partial correlation.

METHODS OF STUDYING CORRELATION:-

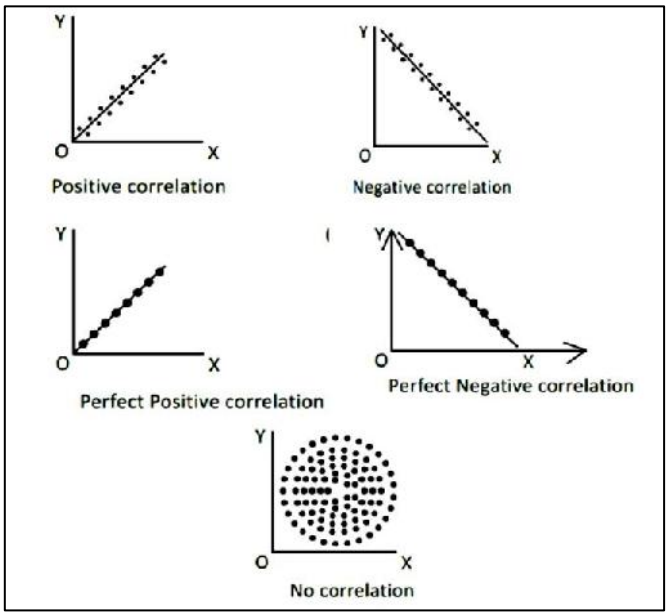
The correlation between the two variables can be determined by the following three methods:-

1. Scatter diagram
2. Karl Pearson's method of correlation coefficient
3. Spearman's method of Rank correlation.

- **SCATTER DIAGRAM:**

It is a graphic (or visual) method of studying correlation. To construct a scatter diagram, x. variable is taken on X axis and Y Variable is taken on Y-axis. The cluster of points so plotted is referred to as a scatter diagram. In a scatter diagram, the degree of closeness of scatter points and their overall direction gives us an idea of the nature of the relationship:-

1. If the dots move from left to the right upwards, correlation is said to be positive whereas the movements of dots from left to right downward indicates negative correlation.
2. Dots in a straight line indicate perfect correlation.
3. Scattered dots indicate no-correlation
4. Dots falling close to each other in a straight line indicate high degree of correlation.



DEGREES OF CORRELATION

Degree	Positive	Negative
Perfect	+1	-1
Very High	+.9 to +1	-.9 to -1
Fairly High	+.75 to+ .9	-.75 to - .9
Moderate	+.5 to +.75	-.5 to -.75
Low	+.25 to +.5	-.25 to -.5
Very Low	Less than+ .25	Less than- .25
No Correlation	0	0

KARL PEARSON'S COEFFICIENT OF CORRELATION:-

Karl Pearson's coefficient of correlation is an important and widely used method of studying correlation. Karl Pearson has measured the degree of relationship between the two variables with help of correlation coefficient. Coefficient of correlation measures the degree of relationship between the two variables. Computation of Karl Pearson coefficient of correlation: - The various formulae used to calculate coefficient of correlation

$$r=\frac{\sum XY}{\sqrt{\sum X^2}\sqrt{\sum Y^2}}$$

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

- **PROPERTIES OF COEFFICIENT OF CORRELATION: -**

Some of the important properties of Karl- Pearson's coefficient of correlation are: -

1. The correlation coefficient is independent of the units of measurement of the variables:-
2. The value of co-relation coefficient(r) lies between +1 and -1.
3. The correlation coefficient is independent of the choice of both origin and scale of observations.
4. The correlation coefficient of the variables x and y is symmetric,

- **ADVANTAGES OF KARL PEARSON'S METHOD:-**

Karl person's method assumes a linear relationship between two variables x and y. The most important advantage of this method is that it gives an idea about co-variation of the values of two variables and also indicates the direction of such relationships.

RANK CORRELATION:-

Charles Edward spearman evolved another method of finding out correlation between different qualitative attributes of a variable. This is known, as rank correlation coefficient. When a group of individuals are arranged according to their degree of possession of a character (say, beauty, intelligence etc.), they are said to be ranked. Spearman's formula for ranks correlation coefficient in as follows:-

$$R = 1 - \frac{6\sum D^2}{(N^3 - N)}$$

NUMERICAL PROBLEMS

1. Find Karl Pearson Coefficient of Correlation

X	Y	X-X =X	$(X-X)^2$	Y-Y= Y	$(Y-Y)^2$	XY
10	13	-9	81	-5	25	45
12	18	-7	49	0	0	0
18	12	-1	1	-6	36	6
24	25	5	25	7	49	35
23	30	4	16	12	144	48
27	10	8	64	-8	64	-64
114	108	0	236	0	318	70

X =114/6 = 19, Y =108/6 =18

$$r = \frac{\sum XY}{\sqrt{\sum X^2} \cdot \sqrt{\sum Y^2}}$$

$$\frac{70}{\sqrt{236}\, x \sqrt{318}} = .2555$$

Low degree of positive correlation

☞ **Find Karl Pearson Coefficient of Correlation**

1	X	8	7	6	3	2	1	5	4
	Y	7	5	4	1	3	2	6	8
2	X	46	36	39	45	54	58	30	40
	Y	30	60	40	50	70	65	33	52
3	X	1	2	3	4	5	6	7	8	9
	Y	9	8	7	6	5	4	3	2	1
4	X	15	20	18	25	22				
	Y	2	8	60	25	5				
5	X	8	18	25	30	46				
	Y	60	50	33	20	15				

☞ **Find Rank Correlation**

X	Y	R_1	R_2	$R_1-R_2=D$	D^2
10	13	6	4	2	4
12	18	5	3	2	4
18	12	4	5	-1	1
24	25	2	2	0	0
23	30	3	1	2	4
27	10	1	6	-5	25
					38

$$R=1-\frac{6\sum D^2}{N^3-N}$$

$$1-\frac{6x38}{6^3-6}$$

=1-1.08= -.08

High degree of negative correlation

X	17	13	15	16	6	11	14	10	7	12
Y	36	46	35	24	12	18	27	22	2	8
Rank -X	1	2	3	4	5	6	7	8	9	
Rank- Y	9	8	7	6	5	4	3	2	1	
X	36	56	20	65	42	33	44	50	15	60
Y	50	35	70	25	58	75	60	45	80	38
Rank -X	1	2	4	6	7	5	3			
Rank- Y	2	1	4	5	6	7	3			

☞ **Merits of Spearman's Rank Correlation**

1. Simple and easy to calculate
2. Not affected by extreme values

☞ **Demerits of Spearman's Rank Correlation**

1. Not Suitable for grouped data
2. Not based on original values of observations

8. INDEX NUMBERS

DEFINITIONS OF INDEX NUMBERS: -

In the words of Edge worth, 'Index number shows by its variation the changes in a magnitude which is not susceptible either of a accurate measurement in itself or of direct valuation in practice.'

In the words of Tuttle, 'An index number is a single ratio (usually in percentage) which measures the combined (i.e., averaged) change of several variables between two different times, places or situations.'

☞ **Features/Characteristics of Index Numbers:-**

1. Index numbers are specialized averages.
2. Index numbers are expressed in percentages.
3. Index numbers measure the effect of changes in relation to time or place.

☞ **Types of Index Numbers:-**

1. Consumer Price Index (CPI)
2. Wholesale Price Index (WPI)
3. Index of Agricultural Production
4. Sensex (Stock Market)
5. Index of Industrial Production (IIP)

☞ **Price Index Numbers:-**

- **Wholesale Price Index Numbers**

The wholesale price index number reflects the general price level for a group of items taken as a whole. In India, it is the most popular price index used in the business industry and policy market. It acts as an indicator of the rate of inflation.

- **Retail Price Index Numbers**

The retail price index number reflects the general changes in the retail prices of various items including food, housing, clothing, and so on. The Consumer Price Index, a special type of retail price index, is the primary measure of the cost of living in a country.

METHODS OF CONSTRUCTING PRICE INDEX NUMBERS

1. Unweight Index Numbers
2. Weighted Index Numbers.

Both of these methods of constructing index numbers are further classified as**:**

Aggregative Method and Average of Relatives Method

SIMPLE AGGREGATIVE OF ACTUAL PRICE METHOD

This is the simplest method of constructing index numbers. In this method, aggregate prices of all the selected commodities in the current year are divided by the aggregate prices in the base year and Multiplied by 100 to get Index. The steps in the constructions of such an index are:

1. Add up the current year prices of various commodities and denote by ΣP_1.
2. Add up the base year prices of various commodities ΣP_0.
3. Use the following formula:

$$P_{01} = \frac{\sum P1}{\sum P0} \, x \, 100$$

Where

P_{01} = Index Number of the Current Year

$\sum P1$ = Total of the Current Year's price of all commodities

$\sum P0$ = Total of the base Year's price of all commodities

- **Limitations**

1. It is influenced by the magnitude of the prices
2. Equal weights are assigned to every item
3. Prices of various commodities may be quoted in different units

Simple Average of Price Relatives Method: -

This index is an improvement over the simple aggregative price index because it is not affected by the unit in which prices are quoted. Price relative: A price relative is percentage ratio between price of commodity in the current year and that in the base year

$$\text{Price Relative } P_{01} = \frac{\text{Current Year Price } P_1}{\text{Base Year Price } P_0} \text{ x } 100$$

Price index number of current year can be found out using the formula

$$P_{01} = \frac{[\frac{P_1}{P_0} \text{x } 100]}{N}$$ Where N is the number of goods

MERITS AND DEMERITS OF AVERAGE PRICE RELATIVE INDEX

- **Merits**

1. This index has the following advantages over the simple aggregate price index:
2. The value of this index is not affected by the units in which prices of commodities are quoted. The Price relatives are pure numbers and, therefore, are independent of the original units in which they are quoted.
3. Equal importance is given to each commodity and extreme commodities do not influence the index number.

- **Demerits**

1. As it is a unweight index, each price relative is given equal importance. However, in actual practice, a few price relatives are more important than others.
2. Difficulty is faced with regard to the selection of an appropriate average.

WEIGHTED INDEX NUMBERS

Weighted index numbers can be constructed by two methods:

1. Weighted Aggregative Method; and

2. Weighted Average of Price Relatives Method.

- **Weighted Aggregative Method**

1. Laspeyre's method

2. Paasche's Method

3. Fisher's Ideal Method

Laspeyre's Method

This method was introduced by Laspeyre in 1871. In this method, weights are represented by the quantities of commodities in the base year Q_0

Formula:

$$P_{01} = \frac{\sum P1\, Q0}{\sum P0\, q0}\, x\, 100$$

P_1 = Price in the Current year
Q_1 = Quantity in the current year
P_0 = Price in the base year
Q_0 = Quantity in the base year

Paasche's Method:

This method was introduced by Paasche in 1874. In this method, weights are represented by the quantities of commodities in the Current year Q_1

Formula:

$$P_{01} = \frac{\sum P1\ Q1}{\sum P0\ q1}\ x\ 100$$

P_1 = Price in the Current year
Q_1 = Quantity in the current year
P_0 = Price in the base year
$Q_{0\ =}$ Quantity in the base year

Fisher's Method:

This method was introduced by Irving Fisher This methods Combines the techniques of Laspeyre method and Paasche method Fisher has combined the techniques of Laspeyre's and Paasche's Method. He used both base year as well as Current Year quantities (q_0 q_1) as weight.

$$P_{01} = \sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1}} \times 100$$

CONSUMER PRICE INDEX (CPI)

Meaning: -

The consumer price index numbers are also called (i) Cost of Living Index Numbers, (ii) Retail Price Index Numbers, or (iii) Price of Living Index Numbers. They are designed to measure effects of change in prices of a basket of goods and services on purchasing power of a particular section of the society during any given (current) period with respect to some fixed (base) period. Consumer Price Index reflects the average increase in the cost of the commodities consumed by a class of people so that they can maintain the same standard of living in the current year as in the base year.

In India, the consumer price Indices are constructed for the following consumer groups.

1. Industrial Workers (IW)
2. Urban- Non Manual Employees (UNME)
3. Agricultural Labourers (AL)

$$\text{Consumer Price Index Number (CPI)} = \frac{\sum P1\ Q0}{\sum P0\ q0}\ x\ 100$$

Uses of Consumer Price Index (CPI) Number

1. Consumer price index numbers helps in wage negotiations, formulation of wage policy, price policy, rent control, taxation and general economic policy formulation.
2. The government and business units use the consumer price index numbers to regulate the Dearness allowance (D.A.) to compensate them for increased cost of living due to price rise.
3. The CPI is used to measure purchasing power of the consumer in rupees.
4. Consumer Price index numbers are also used for analysing markets for particular kinds of goods and services.

INDEX OF INDUSTRIAL PRODUCTION

Index numbers of industrial production have become fairly common these days. The index number of industrial production measures the change in the level of industrial production comprising many industries.

$$\text{Index Number of Industrial Production} = \frac{\sum\left(\frac{q_1}{q_0}\right)W}{\sum W} \times 100$$

Where,
q_1 = Level of production in the current year
q_0 = Level of production in the base year
w = Weight or relative importance of industrial output

WHOLESALE PRICE INDEX NUMBERS: (WPI) -

A wholesale price index (WPI) is an index that measures and tracks the changes in the price of goods in the stages before the retail level – that is, goods that are sold in bulk and traded between entities or businesses instead of consumers.

WPI is calculated on a base year and WPI for the base year is assumed to be 100. ... In this way individual WPI values for the remaining 434 commodities are calculated and then the weighted average of individual WPI figures are found out to arrive at the overall Wholesale Price Index. The WPI is published by the Economic Adviser in the Ministry of Commerce and Industry.

☞ **Utility of Wholesale Price Index Number**

1. Indicator of Inflation

2. Forecasting Demand and Supply

3. Helps in determining real changes in aggregates

4. Useful in Cost Evaluation of various projects

Numerical Examples

Compute index number using simple aggregate method from the following data

Commodity	A	B	C	D
Price 2012	40	60	20	30
Price 2014	55	80	40	50

Commodity	2012	2014
A	40	55
B	60	80
C	20	40
D	30	50
	150	225

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

$$\frac{225}{150} x100$$

=150

Compute index number using average of price relative from the following data

Commodity	A	B	C	D	E	
Price 2012	60	50	100	105	25	
Price 2014	80	70	90	115	20	
$P01 = \frac{P1}{P0} x100$	$\frac{80}{60} x100$	$\frac{70}{50} x100$	$\frac{90}{100} x100$	$\frac{115}{105} x100$	$\frac{20}{25} x100$	
∑P01	133.3	140	90	109.52	80	552.82

$$P_{01} = \frac{\sum \left(\frac{P_1}{P_0} \times 100 \right)}{N}$$

$$\frac{552.82}{5}$$

=110.57

Compute index number using from the following data Laspeyre's, Paasche's and Fisher's ideal index number

Commodities	2012		2013					
	P0	Q0	P1	Q1	P1Q0	P0Q0	P1Q1	P0Q1
A	20	8	40	6	320	160	240	120
B	50	10	60	5	600	500	300	250
C	40	15	50	10	750	600	500	400
D	20	20	20	15	400	400	300	300
					2070	1660	1340	1070

Laspeyre's $P_{01}=\frac{\sum P_1q_0}{\sum P_0q_0}\times 100$ $\frac{2070}{1660}X100$ =124.69

Paasche's $P_{01}=\frac{\sum P_1q_1}{\sum P_0q_1}\times 100$ $\frac{1340}{1070}X100$ =125.23

Fisher's ideal index number

$P_{01}=\sqrt{\frac{\sum P_1q_0}{\sum P_0q_0}\times\frac{\sum P_1q_1}{\sum P_0q_1}}\times 100$ $\sqrt{Laspeyre's\ x\ Paashe's}$ x100 $\sqrt{124.69\ x\ 125.23}$ x100 =124.94

9. PROJECT FOR APPLICATION OF STATISTICS IN ECONOMICS

A project can be defined as a well thought out plan of action made to achieve specific objective.

Steps for developing a project:

1. **Objective**: The required data has to be collected with a clear objective of the project.
2. **Population**: To develop a project, the target population should be determined.
3. **Collection of data**: Data regarding project report should be collected using primary or secondary data.
4. Secondary data must be used with great care.
5. **Organization and presentation of data** – Report can be presented using statistical tables, graphs or diagrams.
6. **Analysis of data** – statistical measures like measures of central tendency and dispersion can be used for analysing the data.
7. **Interpretation & Conclusion** – Data collected and analysed should be correctly interpreted.