VIDYA UNIVERSITY PRESS

TOUCH n PASS

सर्वश्रेष्ठ परीक्षा पथ-प्रदर्शक

B.Com.
First Year
(Certificate Course)
Semester-I

Business Statistics

SYLLABUS

UNIT-I

Evolution of Statistics in Indian, contribution of Indian Statistics Prof. Prasanta Chandra Mahalanobis. Introduction to Statistics : Meaning, Scope, Importance and Limitation, Statistical Investigation-Planning and organization, Statistical units, Methods of Investigation, Census and Sampling. Collection of Data-Primary and Secondary Data, Editing of Data, Classification of Data, Frequency Distribution and Statistical Series, Tabulation of Data, Diagrammatical and Graphical Presentation of Data.

UNIT-II

Measures of Central Tendency : Mean, Median, Mode, Quartile, Decile, Percentile, Geometric and Harmonic Mean; Dispersion : Range, Quartile Deviation, Mean Deviation, Standard Deviation and its Co-efficient, Co-efficient of Variation and Variance, Test of Skewness and Dispersion, Its Importance, Co-efficient of Skewness.

UNIT-III

Correlation : Meaning, application, types and degree of correlation, Methods : Scatter Diagram, Karl Pearson's Coefficient of Correlation, Spearman's Rank Coefficient of Correlation, concurrent method, Parabola Error & Standard Error.

UNIT-IV

Index Number : Meaning, Types and Uses, Methods of constructing Price Index Number, Fixed-Base Method, Chain-Base Method, Base conversion, Base shifting deflating and splicing. Consumer Price Index Number, Fisher's Ideal Index Number, Reversibility Test-Time and Factor, Analysis of Time Series : Meaning, Importance and Components of a Time Series. Decomposition of Time Series : Moving Average Method and Method of Least square.

VIDYA®
UNIVERSITY PRESS

Registered Office
Vidya Lok, Baghpat Road, T.P. Nagar,
Meerut, Uttar Pradesh (NCR) 250 002
Phone : 0121-2513177, 2513277
www.vidyauniversitypress.com

Editing & Writing
Research and Development Cell

Printer
Vidya University Press

CONTENTS

UNIT-I	:	Origin of Statistics in India	...3
UNIT-II	:	Measures of Central Tendency	...45
UNIT-III	:	Correlation	...89
UNIT-IV	:	Index Number	...121
	⊙	Model Paper	...159

UNIT-I

Origin of Statistics in India

SECTION-A VERY SHORT ANSWER TYPE QUESTIONS

Q.1. What do you know about Professor Prasanta Chandra Mahalanobis?

Ans. Professor Prasanta Chandra Mahalanobis, known as the father of Indian Statistics, was born in Calcutta (Kolkata).

Q.2. Write main objectives of Statistics.

Ans. Following are the important objectives of statistics :

1. By analysing the information included in sample data, it draws inferences about population. These inferences consists assessments of extent of uncertainty.
2. It designs the method and range of sampling from that observations make the base for making valid inferences.

Q.3. Give three differences between census and sample.

Ans. The differences between the census and sample are as follows :

Basis	Census	Sample
Meaning	When the whole population is targeted for collecting the relevant information about the members of population, it is called Census.	When a representative sample is prepared from the target population for collecting the relevant information, it is called Sampling.
Reliability of Data	Data obtained through census may be reliable and complete.	Data obtained through sampling is not fully reliable, it may have some error.
Time Money Taken	As it covers the entire population, it is time-taking process.	As it covers only few representatives of the population, it is not time-taking process.

Q.4. What are the characteristics of statistical units?

Ans. Following are some characteristics of statistical units :

1. The unit should be appropriate to the objective of the enquiry.
2. The unit should be standard.
3. The unit should be easily and clearly understandable.
4. The unit should have comparability.
5. The unit should be unambiguous and precisely defined.
6. There should be uniformity in the selected unit.
7. The unit should not be inadequate and incorrect.

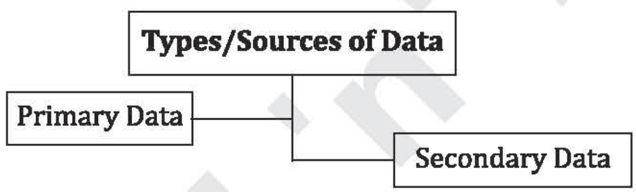
Q.5. What are the demerits of probability sampling?

Ans. Demerits of Probability Sampling are as follows :

1. **High Level Skills :** A very high level of skill and experience is required for using probability sampling.
2. **Non-utilisation of Additional Knowledge :** No additional knowledge is used in this method about how the population is organised.
3. **Less Efficient :** Other sampling methods are more statistically efficient in comparison to this method.
4. **High Costs :** In comparison to non-probability sampling, probability sampling involves higher cost.
5. **More Time Required :** For planning and execution of a probability sample a lot of time is required.

Q.6. Write the types of data.

Ans. Depending upon the sources being utilised, any statistical data can be divided into two categories, which are as follows :



Q.7. Write any two demerits of primary data.

Ans. Primary data has following limitations :

1. **Costly Affair :** Primary data collection is an expensive task. It involves different activities like selecting type of technique, preparing questions, and hiring trained professionals for collecting information or observing targets, etc. In this process, a huge amount is spent, which is why it is costly to conduct.
2. **Time Consuming :** Collecting primary data effectively takes more time. Developing research plan, deciding sources of information, and selecting the methods of data collection are time consuming activities.

Q.8. Compare Classification and Tabulation.

Ans. Following table illustrates the difference between classification and tabulation :

S.No.	Classification	Tabulation
1.	Classification is the base for tabulation of data. so it is not dependent in tabulation.	Tabulation is used for advanced analysis of data.
2.	Classification makes data simple.	Tabulation makes clear presentation of data.
3.	Classification breaks data into groups and subgroups. These groups and subgroups are made on the basis of data similarities dissimilarities.	Tabulation helps to list the data in a logical order on the basis of data and properties.

Q.9. Write any one definition of classification.

Ans. According to *R.L. Connor*, "Classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities, and giving expression to the unity of attributes that may subsist (manage to survive) amongst a diversity of individuals."

Q.10. What do you understand by Line Charts?

Ans. The line chart consists of numerous points interconnected through straight lines. Generally, line charts are used to denote the fluctuation in data over a given time period.

Q.11. Define frequency curve.

Ans. If the mid-point of the upper boundaries of the rectangles of a histogram is corrected by a smooth freehand curve, then such diagram is called a frequency curve. The curve should be start and end at the base line.

Q.12. What precautions should be taken in constructing table?

Ans. Precautions in Constructing Table are given below :

1. Do not give the much detail of content in table.
2. Be care of its simplicity.
3. Be care about columns and rows headings.
4. It should not be much lengthy.
5. One should maintain suitable approximation and figure's rounded off.
6. Measuring units must be clearly defined.
7. One can make multiple tables if the observations are large.

Q.13. What are the limitations of graphs?

Ans. Limitations of Graphs are given below :

1. Many people are not accustomed to it and they do not attach much importance to it.
2. Graphic presentation may often give misleading impressions. Much depends upon the scale taken. Two different scales may show different fluctuations in the data.
3. Accuracy is rather not possible in a graph.
4. Graphs cannot be quoted in support of some statement.
5. Only one or two characteristics can be depicted in a graph. If more features are shown, the graph becomes difficult to follow.

SECTION-B SHORT ANSWER TYPE QUESTIONS

Q.1. Explain the nature of statistics.

Ans.

Nature of Statistics

Statistics is an Art as well as a Science. It can be understand by the following topics :

Statistics as a Science

Science is a body of systematized knowledge from which specific propositions are deducted in accordance with a few general principles. Although ail sciences differ, the logical scientific methods are common to all the sciences. Using the approach of systematic doubt, scientific method is a process of discovering the truth in a systematized manner by logical considerations. There are in general four stages in a scientific inquiry :

(i) Observation, (ii) Hypothesis, (iii) Prediction, and (iv) Verification.

Since statistical methods are based on the same fundamental ideas and processes as other sciences, so statistics is said to be a science. But statistics is different from Physics, Chemistry, etc.

This is why *Croxton and Cowden* said, "Statistics is not a science, it is a scientific method." In fact, statistics is not studied for its own sake, rather it is widely employed as highly valuable in the analysis of problems in the natural and social sciences.

On the basis of the above discussion we can say that "Statistics is not only a science but is also a scientific method" and may be called a science of scientific methods. According to *Tippett*, "As science, the statistical method is a part of the general scientific method as it is based on the same fundamental ideas and processes." In the words of *Wallis* and *Roberts*, "Statistics is not a body of substantive knowledge but a body of methods for obtaining knowledge." In fact, statistics is a means not an end.

Statistics as an Art

An art is an applied knowledge and creation of beauty leading to perfection. If science is a knowledge then art is an action. Since the successful application of statistical methods depends to a considerable degree on the skill and special experience of the statistician and on his knowledge of the field of application, statistics may be called an art of applying scientific methods similar to an artist who possesses and can apply the requisite skill, experience and patience for the creation of beauty leading to perfection.

Statistics is both a Science and an Art

According to *Anderson and Bancroft*, "Statistics is the science and art of the development and application of the most effective method of collecting, tabulating, and interpreting quantitative data in such a manner that the falliability of conclusions and estimates may be assessed by means of inductive reasoning based on mathematics of probability."

According to *Tippett*, "It is both a science and an art. It is a science in that its methods are basically systematic and have general application, and an art in that their successful application depends to a considerable degree on the skill and special experience of the statistician and his knowledge of the field of application, *e.g.*, Economics."

Q.2. Discuss main limitations of statistics.

Ans.

Limitations of Statistics

Major limitations of statistics are as follows :

1. **Statistics is only One of the Methods of Studying a Problem :** In all conditions it is not necessary that statistical tools give the best solution. In a statistical problem it is important to take a country's culture, religion and philosophy into consideration.
2. **Statistical Results are True only on an Average :** Conclusions or inferences which are drawn after statistical study may not be true universally but only under certain conditions.
3. **Statistic Deals only with Quantitative Characteristics :** Statistics deals with the quantitative statements of the facts. Various characteristics cannot be expressed quantitatively and hence statistics is unable to analyse these characteristics.

Qualitative characteristics such as honesty, efficiency, intelligence, blindness and deafness cannot be studied statistically.

4. **Statistics does not deal with Individual Measurements :** The study of individual measurements lies outside the range of statistics as it deals with aggregates of facts. *For example,* marks obtained by a single student in a class have no statistical significance but the average marks received by the students of the same class are relevant to statistics.
5. **Statistics is collected with a given purpose and cannot be Indiscriminately applied to any situation :** The relevance of a statistical result in one situation does not guarantee its utility in another. Usage of secondary data leads to incorrect drawing of conclusions. It is, therefore, necessary to thoroughly scrutinise statistical data before inference.
6. **Statistical relations do not necessarily bring out the 'Cause and Effect' relationship between phenomena :** Statistics tells the relationships between variables but does not define which one is the cause and which one is the effect. It only reveals the association between the two variables.
7. **Statistics does not reveal the entire Story :** Statistics only simplifies and analyses complex data but does not give the real picture of the background about the data.
8. **Statistics can be misused :** Statistics may be misused when statistical conclusions are based on incomplete information. This may lead to incorrect conclusions.

Q.3. Write the meaning and definition of statistical investigation. Describe the preliminary steps you would take in planning of statistical investigation of statistical investigation.

Ans.

Meaning and Definition of Statistical Investigation

Investigation or *Inquiry* means a search for truth, knowledge or information. Statistical Investigation, therefore, means a search conducted by statistical methods.

In general, statistical investigation means statistical survey.

In brief, scientific and systematic collection of data and their analysis with the help of various statistical methods and their interpretation so as to throw light on some problem to fulfil a certain purpose after making a suitable plan is known as statistical investigation or inquiry or survey.

The person who carries the statistical investigation is known as investigator. The person from whom the information is obtained is known as respondent. The investigator needs the help of certain persons to carry the investigation, they are called enumerators.

Planning of Statistical Investigation

To consider various points before the actual work of statistical investigation, is termed as planning of statistical investigation. Planning of a statistical investigation is essential for its successful completion and getting the best results at the least cost and time. According to *Neter* and *Wasserman*, "The relevancy and accuracy of the data obtained in a survey depend directly upon the care with which the survey is planned."

The points that require careful consideration at the planning stage of statistical investigation are stated below :

1. Definition of the problem and to understand the object well of investigation.
2. To decide the scope (*i.e.,* field of Data Collection).

3. At what and how much time the investigation is to be completed.
4. To consider the organisation and budget of the investigation.
5. To take decision about the sources of information relevant to the investigation.
6. To decide the type of the investigation keeping in mind the object, time, nature, scope, expected cost, sources, etc.
7. To decide about the statistical unit, unit of measure and measure of accuracy.
8. Selection of the proper method of data collection.
9. Organisation of investigation (construction of questionnaire, selection of enumeration, and their training).
10. Editing, presentation, analysis, interpretation, etc. of the data collected.
11. To prepare report and publish the facts.

Q.4. What is a statistical unit? What are its different types?

Ans.

Statistical Unit

It is a unit in which the data have to be collected should be found out by the investigators. The **statistical units** are those entities for which information is collected and for which statistics are compiled. These units are used for statistical purpose.

An individual, a family, a group, etc., can be a statistical unit. Statistical units must be defined very clearly before collecting data. It avoids any confusion. If we select the wrong statistical unit, then collection of data will also be wrong. Hence, to resolve any statistical problem, only data collection is not important.

The units of data measurement also play an important role.

Types of Statistical Units

Following are the two types of statistical units :

1. **Units for Data Collection :** In these units, data collection is performed. These units either involve counting or measurement. The counting works for physical items and the measurement works for qualitative attributes. Therefore the data collection process can deal with either discrete values or measurable values. Here discrete values like number of deaths, number of persons, etc., and measurable values like kilograms, metres, litres, etc. Units of collection can be three types :
   (i) **Simple Unit :** This unit is a single unit and does not have any qualification. *For example,* worker, hour, rupee, house, etc., are the simple units.
   (ii) **Composite Unit :** This unit is a single unit and possesses some qualification. *For example,* skilled worker, machine-hour, etc., are the composite units.
   (iii) **Complex Unit :** This unit is a single unit and possesses two or more qualifications. *For example,* production per machine-hour is a complex unit.
2. **Units for Statistical Analysis and Data Collection :** The main objective of statistical data collection is comparisons between two different data sets. These comparisons can be performed on the basis of time or space. So the facility of comparisons is provided by the units of statistical analysis and interpretation These units may be of the following types :

(i) **Coefficient :** It is the unit used for the comparison of numerator and denominator. The formula is : C = Number/Total base number.

(ii) **Rate :** These are usually expressed per hundred, per thousand or per million, etc., *e.g.,* death rate is expressed in terms of per thousand.

(iii) **Ratio :** It is used to express the relative values of two homogeneous facts, *e.g.* ratio of males and females will be 4 :1 if there are 800 males and 200 females in a factory.

Q.5. Explain the census method of statistical investigation and discuss their relative advantages and disadvantages.

Ans.

Census Method

If the detailed information about every individual person or item of a given population or universe is collected then it is called 'complete enumeration' or census method. *For example*, during the Census of Population (which is done every ten years in India), the information about each individual person is collected, residing in India. This method provides information about each and every unit of population with complete accuracy.

Advantages of Census Survey Method

1. **Small Population :** In the condition when the size of universe is small then census method is appropriate technique in comparison to sampling method.
2. **No Sampling Errors :** No sampling error occurs in this method because each and every unit is included in the study.
3. **Highest Degree of Accuracy :** In this method there is no chance of leaving any element for data collection, hence complete accuracy is obtained because all the members are covered in the enquiry.

Disadvantages of Census Survey Method

1. **Destructible Population :** If the population is unstable then this technique cannot be used.
2. **No Urgent Results :** This technique cannot be used when results of any problem is required urgently.
3. **Expensive :** Small organisations and individuals cannot afford this technique because it involves high amount of money, time and energy. Generally the government follows census method of survey.
4. **More Organisational Skills :** Organisational difficulties are faced in this technique because a large team of researchers is required, having different standards of skills, willingnesses and efficiencies.

Q.6. Explain the sampling method of statistical investigation and discuss their relative advantages and disadvantages.

Ans.

Sampling Method

This method is used by taking a sample of consumer's for interview. The sampling may be random or stratified sampling.

The success in this method depends on making the correct sampling and cooperation of consumers is necessary.

By this method time is saved, more over it is economic. It is to be noted that if the universe is small, sample survey is not necessary for investigation.

Advantages of Sampling Method

Following are the advantages of sampling :

1. **Gives More Comprehensive Information :** Thorough investigation of the study is the result of small sample which provides more complete information as all the members of the population have equal chance to be included in the sample.
2. **More Accurate :** Small errors are in sampling because small data are involved in collection, tabulation, presentation, analysis and interpretation.
3. **Faster and Cheaper :** Due to small sample size the data collection, tabulation, presentation, analysis, and interpretation takes less time and it involves nominal expenses.
4. **More Effective :** Due to the size of the sample being small than that of the population, the tiredness is reduced in collecting the information and investigator works more effectively.
5. **Saves Time, Money and Effort :** The subjects involved in sampling are small in number which gives him less time to calculate, tabulate, present, analyse, and interpret and hence the researcher saves time, money and effort.

Disadvantages of Sampling Method

Following are the disadvantages of sampling :

1. **Limited Nature :** Due to small or dissimilar universe it is impossible to derive a representative sample where census study is the best possible substitute.
2. **Less Accuracy :** When the higher standard of accuracy is expected then the sampling is not suitable.
3. **Problem of Cooperation :** Due to scattered sample the subjects are uncooperative with the researcher.
4. **Specialised Knowledge Needed :** Specialised knowledge is required in the sampling method where the investigator may commit serious mistakes due to its absence.
5. **Difficulty in Selection :** It is very difficult to select a truly representative sample because a large number of factors obstructing the method of selecting good samples.
6. **Biased Selection :** Biased selection of the respondents may be included in sampling by the research worker.

Q.7. What are the qualities of good sample.

Ans.

Qualities of Good Sample

1. **Approachable :** The subjects of good sample are easily accessible where the tools of research are easily conducted and easy collection of data is possible.
2. **Comprehensive :** A sample that is true representative of the population is also comprehensive in nature which is controlled by definite purpose of investigation. A sample's characteristic may be comprehensive but it may not be a good representative of the population.
3. **Accurate :** A sample is called good when it yields accurate estimates or statistics are free from errors.

4. **Free from Bias :** A good sample does not allow prejudices, pre-conceptions, and imaginations which affects its choice and it is unbiased.
5. **True Representative :** The true representative of the population and matching its properties is termed as good sample where aggregate of certain properties is the population and sample is the sub-total of the universe.
6. **Economical :** It refers that the research should not incur huge costs, time or efforts. One of the objectives of any research is to complete the research with minimum effort, time, money and resources. The researcher calculates per unit cost for each respondent. The researcher should choose that sampling design which gives minimum per respondent cost and maximum accuracy.
7. **Practical :** It means that the concepts of sample selection should be applied properly while conducting the research. The researcher should be well experienced and well instructed. The instructions which are passed to observer should be clear, complete and correct in all terms so as to avoid errors and biasness on their part. The sample should be selected on basis of the sample design. The sampling units should be representative. The sample design should be practical and feasible in nature.
8. **Goal Orientation :** Any sample which is selected by the researcher should be able to satisfy the objectives of the research. The sample should be taken in proper number. It should be customized to fit the environment under which the research is going to be conducted. If the sample is changed as per the requirement of the survey design then it can come out with better results and outcomes.
9. **Feasible :** A good sample creates the research work more feasible.
10. **Good Size :** The size of good sample is such that it yields an accurate result and the error due to probability can be estimated.

Q.8. Distinguish between primary and secondary data.

Ans.

Distinction between Primary and Secondary Data

The main difference between primary and secondary data is degree. Primary data once collected becomes secondary data for other investigation. That is the data which are primary in the hands of one may be secondary for others. In the words of *H. Secrist*, "The distinction between primary and secondary data is largely one of degree. Data which are secondary in the hands of one party may be primary in the hands of another." For example, the data relating to Indian census collected by government are primary data but this will be secondary when these data are used by other researchers. However, we may consider the following points to compare the two types of data :

S. No.	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 form 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 un published).
5.	*Precautions*	Precautions are not necessary in the use of primary data.	Precautions are necessary in the use of secondary data.

Q.9. Write the merits of primary and secondary data.

Ans.

Merits of Primary Data

Primary data is significant in research due to following reasons :

1. **Complete Control over Process :** Sometimes, organisations ask the researchers to conduct the research in specific area rather than in broader perspective. Collecting the primary data allows the researchers to collect the data of their concern and represent it in ways that can benefit the organisations. Researchers can also decide the length of study, location in which research is to be carried out, time duration, etc., as per their requirement and convenience.
2. **Wide Coverage Including Special Cases :** Primary data is applicable in many areas including some special cases. Sometimes, researchers want information regarding particular cases for which previous literature is not available. Collecting primary data is the only solution for these specific research problems or issues. In these cases, primary data is the only source of information which can be trusted for effective solution.
3. **Variety of Techniques :** Primary data can be collected through various techniques. There are numerous tools and techniques available to record and analyse primary data such as interviews, questionnaires, observation, audits, etc. It allows the researchers to explore effectively in almost every area where research is possible.
4. **Reliability :** As the primary data is collected originally by the researcher and it is current and accurate, it is more reliable than secondary data.
5. **Cost Effective Collection :** The collection of primary data is cost-effective. Many times unnecessary time and money is wasted in collecting secondary data, and the information proves to be useless. But in primary data collection, the researcher concentrates his efforts on potential sources of data which provide reliable information in optimal cost.

Merits of Secondary Data

Secondary data is significant for research in following ways :

1. **Useful in Exploratory Research :** As exploratory research is conducted with the purpose of getting better insight about an issue or phenomena, secondary data is very useful in serving its purpose by providing extensive available information from various sources. This helps the researchers carry out the research accordingly.
2. **Availability :** Secondary data is widely available and hence easily accessible. Secondary data are helpful especially when it is quite difficult to collect primary data. Majority of secondary data is available which can be utilised for a particular research.

3. **Measuring Instruments are not Required :** In collecting secondary data, there is no need to decide the tools and techniques for gathering required information, as this data is already recorded and processed by other researchers. The researchers only need to identify the relevant section of this data which is of their concern.
4. **Quality :** The quality of the secondary data is unique and rare as these are originally collected by trained professionals who have expertise in data collection.
5. **Less Time Taking :** As secondary data is already processed and compiled by other researchers, it takes very less time to collect this data. There is variety of secondary data available from various sources. Hence, researchers just need to search the data from the sources.
6. **Economic :** The secondary data is easier and cheaper to access. It is more economically collected compared to the primary data. Some of the secondary data can be obtained with absolutely no cost.

Q.10. The real and estimated monthly income of three persons *A, B, C* are given below; calculate absolute error, relative error and percentage error :

Person	*A* ₹	*B* ₹	*C* ₹
Real monthly income	1,000	600	920
Estimated monthly income	1,100	500	920

Sol.

$$\text{Absolute Error} = \text{Real Income} - \text{Estimated Income}$$

$$\text{Relative Error} = \frac{\text{Absolute Error}}{\text{Estimated Income}}$$

$$\text{Percentage Error} = \text{Relative Error} \times 100$$

	Absolute Error	Relative Error	Percentage Error
First person	$1{,}000 - 1{,}100 = -100$ negative	$\frac{-100}{1{,}00} = 0.0909$	-9
Second person	$600 - 500 = 100$ positive	$\frac{100}{500} = 0.2$	20
Third Person	$920 - 920 = 0$ zero	$\frac{0}{920} = 0$	0

Thus, *A*'s Error $= -9\%$, negative ;

B's Error $= 20\%$, positive

and *C*'s Error $= 0\%$, zero

Q.11. Write meaning and definition of classification. Discuss the main features and functions of classification.

Ans.

Meaning and Definition of Classification

Classification is a process by which data are arranged into groups or classes according to some criterion. In the words of *Connor*, "Classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblance and affinities and gives expression to the unity of attributes that they may subsist among a diversity of individuals."

The most practical example of classification can be the operation of sorting letters in a Post Office, where letters are classified according to their destination. The criterion by which the collected data are classified into classes and subclasses is called the *basis of classification*.

Main Features of Classification

The main features of classification of data are :

1. Classification of collected data is done according to purpose, field/scope and nature of the statistical investigation.
2. Classification is done according to some characteristic (attribute, variable or measurement). Objects having similar characteristics are placed in one class or group or cell.
3. Classification may be either actual or notional or imaginary.
4. Classification gives expression to the unity of attributes that they may subsist among a diversity of individuals.
5. There is no hard and fast rule for classification. In classification common sense is the chief requisite and experience the chief teacher.

Objects or Functions of Classification

The objects or functions or advantages of classification may be summarized as follows :

1. *To clarify similarity and dissimilarity* : Classification is done to bring out clearly points of similarity and dissimilarity.
2. *To facilitate comparison* : The raw data are classified to make the study easier and facilitate comparison.
3. *To condense the data and to make the data simple* : The main object of classification is to condense the raw data and present them in a simple form so that it becomes easily intelligible.
4. *To prepare the base of tabulation and other statistical techniques* : Classification is the basis of tabulation. It makes the data easier to interpret and to help the drafting of the report to derive statistical inferences.
5. *To grasp the information* : From the raw data no one can appreciate at a glance or even after a careful study the mind cannot grasp the information given there.
6. *To save a lot of time, space and energy* : Classification gives prominence to the important information and eliminates unnecessary details, so a lot of time is saved to study them.
7. *To organise the data logically and scientifically* : In the process of classification the data are classified in a logical way to depict the salient feature of them.

Q.12. Define frequency and frequency distribution.

Ans.

Frequency

Frequency is the number of occurrences of a repeating event per unit time. Or Number of times a given quantity (or group of quantities) occurs in a set of data.

For example,

1. The frequency distribution of income in a population would show how many individuals (or households) have the income of a certain level (say, 5,000 a month).
2. Consider the following data :

   Scores : 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5

   The frequency of 2 is 5.

Frequency Distribution

Frequency distribution is a series when a number of observations with similar or closely related values are put in separate bunches or groups, each group being in order of magnitude in a series. It is simply a table in which the data are grouped into classes and the numbers of cases which fall in each class are recorded. It shows the frequency of occurrence of different values of a single phenomenon.

According to *Croxton and Cowden*, "Frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups, with their corresponding frequencies side by side".

Reasons for Constructing Frequency Distribution

1. To facilitate the analysis of data.
2. To estimate frequencies of the unknown population distribution from the distribution of sample data.
3. To facilitate the computation of various statistical measures.

Q.13. Convert the cumulative frequencies into ordinary frequencies :

	Variable	**No. of Students**
(a)	less than 10	5
	less than 20	17
	less than 30	30
	less than 40	35
	less than 50	36
(b)	more than 0	36
	more than 10	31
	more than 20	19
	more than 30	6
	more than 40	1

Sol. (a) To find the frequency of a class deduct the cumulative frequency of that class from the cumulative frequency of the next (higher) class. The values less than 10, less than 20, etc. gives the upper limits of the classes :

Marks	No. of Students (c.f.)	Marks (Classes)		No. of Students (Frequency)
less than 10	5	0-10		5
less than 20	17	10-20	(17–5)	12

less than 30	30	20-30	(30–17)	13
less than 40	35	30-40	(35–30)	5
less than 50	36	40-50	(36–35)	1
				Total = 36

(b) The values more than 5, more than 10, etc. give lower limit of the classes. To find the frequency of a class, deduct the cumulative frequency of the next (higher) class from the cumulative frequency of the class :

Marks	**No. of Students (c.f.)**	**Marks (Classes)**		**No. of Students (Frequency)**
more than 0	36	0-10	(36–31)	5
more than 10	31	10-20	(31–19)	12
more than 20	19	20-30	(19–6)	13
more than 30	6	30-40	(6–1)	5
more than 40	1	40-50		1
				Total = 36

Q.14. Rearrange the following series with equal intervals and then prepare 'less than type' and 'more than type' cumulative frequency distributions :

Classes	**Frequency**
0-5	3
5-6	2
6-9	7
9-12	5
12-17	16
17-18	12
18-20	15
20-24	20
24-25	8
25-30	10
30-36	2
Total	**100**

Sol. Here the largest width of the classes is 6. Hence, we can take classes as 0–6, 6–12, 12–18, 18-24, 24-30, 30–36 with respective frequencies 3 + 2, 7 +5,16 + 12, 15 + 20, 8 + 10, 2.

Equal Class-intervals	**Frequency**
0-6	5
6-12	12
12-18	28
18-24	35
24-30	18
30-36	2
Total	**100**

The required less than type and more than type cumulative frequency distributions are :

Less than type		More than type	
Variable	**Number**	**Variable**	**Number**
Less than 6	5	More than 0	95+5=100
Less than 12	5+12=17	More than 6	83+12=95
Less than 18	17+28=45	More than 12	55+28=83
Less than 24	45+35=80	More than 18	20+35=55
Less than 30	80+18=98	More than 24	2+18=20
Less than 36	98+2=100	More than 30	2

Q.15. What is tabulation? Also give the objectives of tabulation.

Ans. Tabulation may be defined as the logical and systematic organization of statistical data in rows and columns. It is designed to simplify presentation and facilitate analysis.

According to *L.R. Connor*, "Tabulation involves the orderly and systematic presentation of numerical data in a form designed to elucidate the problem under consideration."

According to *Secrist*, "Tables are a means of recording in permanent form the analysis that is made through classification and by placing in just a position thing that are similar and should be compared".

According to *Blair*, "Tabulation in its broadest sense is an orderly arrangement of data in columns and rows."

Thus, tabulation is one of the most important and ingenious devices of presenting the data in a condensed and readily comprehensible form. It attempts to furnish the maximum information in the minimum possible space, without sacrificing the quality and usefulness of the data. However, it is an intermediate process between the collection of the data and the statistical analysis. In fact, tabulation is the final stage in collection and compilation of the data and forms the gateway to further statistical analysis and interpretation.

Following are the some objectives that are fulfilled by tabulation :

1. **Present the Purpose of Investigation :** According to *Prof. A.L. Bowley*, "The function of tabulation in the general scheme of statistical investigation is to arrange in easily accessible form the answer with which the investigation is concerned." Under the study the data which is presented in the tables makes the problems easy and clear, as well as provide the various solutions of the problems that are arisen there.
2. **Explain the Properties of Data :** Facts of data are clearly and briefly presented by the table. It removes the requirement of the word explanation of the facts. It fetches the main properties of the data.
3. **Required Less Space :** Tabular representation of data takes less space comparatively textual material. As well as the information communication through tabular form is much easier.
4. **Makes Statistical Process Easy :** It makes context of data simple as well as facilitates comparative analysis of data facts. It also speed up the interpretation of the data facts.

Q.16. Explain various types of tables.

Ans.

Types of Tables

Tables may broadly be classified into two categories :

Types of Tables

- Simple and Complex Tables
  - Simple table or one-way table
  - Two-way table
  - Higher order table
- General and Special Purpose Tables

1. **Simple and Complex Tables :** The distinction between simple and complex table is based upon the number of characteristics studied.

   (i) **Simple Table or One-way Table :** In this type of table only one characteristic is shown. This is the simple type of table. *For example,* Marks obtained by 50 students in mathematics.

Marks	Number of Students
0-10	5
10-20	12
20-30	10
30-40	18
40-50	5
Total	**50**

   (ii) **Two-way Table :** Such a table shows two characteristics and is formed when either the stub or the caption is divided into two coordinate parts. *For example,* Marks obtained by 50 students in mathematics.

Marks	Number of Students		
	Male	Female	Total
0-10	2	3	5
10-20	4	8	12
20-30	7	3	10
30-40	5	13	18
40-50	1	4	5
Total	**19**	**31**	**50**

   (iii) **Higher Order Table :** When three or more characteristics are represented in the same table such a table is called higher order table. The need for such a table arises when we are interested in presenting a number of characteristics simultaneously. While constructing such a table it is necessary to first establish an order of precedence among the attributes or characteristics sought to be classified having regard to their relative importance. For example, Marks obtained by 50 students in mathematics.

Marks obtained by 50 students in Mathematics

Marks	Number of Students								
	Male			Female			Total		
	Hoste-llers	Day-Scholars	Total	Hoste-llers	Day-Scholars	Total	Hoste-llers	Day-Scholars	Total
0-10	1	1	2	2	1	3	3	2	5
10-20	4	0	4	4	4	8	8	4	12
20-30	3	4	7	2	1	3	5	5	10
30-40	2	3	5	4	9	13	6	12	18
40-50	1	0	1	1	3	4	2	3	5
Total	**11**	**8**	**19**	**13**	**18**	**31**	**24**	**26**	**50**

It should be remembered that as the number of characteristics represented increases, the table becomes more and more confusing and as such normally not more than four characteristics should be represented in the same table. Where more than four characteristics are to be represented we can have more than one table depicting relationship between the attributes.

2. **General and Special Purpose Tables :** General purpose tables also known as reference tables or repository tables provide information for general use or reference. They usually contain detailed information and are not constructed for specific discussion. In other words, these tables serve as repository of information and are arranged for easy reference. Tables published by the governmental agencies are mostly of this kind, such as the tables contained in the Statistical Abstract of the Indian Union, detailed tables contained in the census reports, etc. Such tables tell facts which are not for particular discussion. When such tables are used by a researcher, they are usually placed in the appendix of the reports for easy reference.

Q.17. Define the parts of a table with specimen.

Ans.

Parts of a Table

1. **Table Number :** Each table should be numbered. There are different practices with regard to the place where this number is to be given. The number may be given either in the centre at the top above the title or inside of the title at the top or in the bottom of the table on the left-hand side.
2. **Title of the Table :** Every table must be given suitable title. The title is a description of the contents of the table. A complete title has to answer the question *what, where* and *when* in that sequence.
3. **Caption :** Caption refers to the column headings. It explains what the column represents. It may consist of one or more column headings. Under a column heading there may be sub-heads. The caption should be clearly defined and placed at the middle of the column.

4. **Footnotes :** Anything in a table, which the reader may find difficult to understand from the title, captions and stubs should be explained in footnotes. Footnotes are needed they are placed directly below the body of the table.
5. **Stubs :** Stubs refers to the headings of the horizontal rows and they are written on the left hand side of the rows. Whether there is need for stubs, and if yes, how many, would depend on the nature of the data.
6. **Body :** The body of the table contains the statistical data, which have to be presented. This is the most vital part of a table and the data contained in the body are arranged according to captions and the stubs.
7. **Headnote :** Statistical tables contain a headnote, which refers to the data contained in the maior part of the table, and it is placed below the title of the table. Generally written as headnote like 'in lakhs', or 'in tonnes', etc.

Following is the specimen (*i.e.*, structure or format) of a table indicating the above parts :

Table

Sub Heading	**Title**		**Headnote**	
	Caption			
	Column Heading	**Column Heading**	**Column Heading**	**Total**
Stub Entries	B	OD	Y	
Total				
Foot Notes				Source

Q.18. What is the importance of Tabulation?

Ans.

Importance/Significance of Tabulation

Significance of tabulation can be understood with the help of following points :

1. **Important for Data Analysis :** Tabulation is the intermediate step between data collection and analysis. Analysis cannot be performed on raw data, which makes it necessary for the data to be tabulated in a systematic and comprehensive manner. Once the data is tabulated, it is possible to proceed for further analysis.
2. **Relevant and Unambiguous Presentation :** Tabulation helps the researchers present the data correctly. It eliminates the repetition and redundancy of the data and shows the relevant information only.
3. **Helps in Comparing Data :** Since the data is presented in a very systematic manner in tables, therefore researchers can easily compare data and draw inferences simultaneously. It makes the data analysis comparatively easy to carry-out.
4. **Simplification of Complicated Data :** The biggest advantage of tabulation is that it simplifies the complicated data and presents it in a comprehensive format for the reader. Tabulated data is easy to understand and interpret. It reduces ambiguity which helps in data analysis and concluding the findings.
5. **Maximum Representation of Data :** Tabulation reduces the huge size of data by representing it in a minimum possible space. It increases the efficiency of research by facilitating the researcher to draw graphs and charts on the basis of tables.

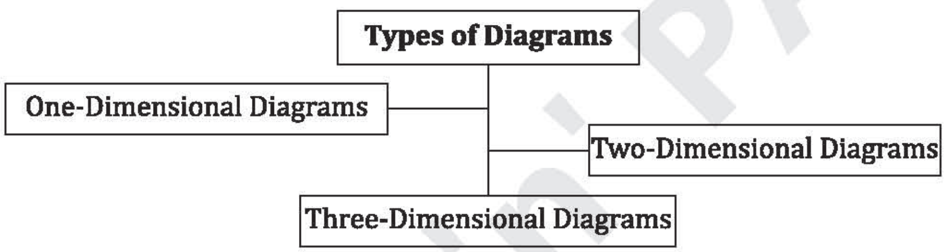
6. **Detecting Missing Data and Omissions :** Tabulation also provides a chance to detect the missing data and omissions. Thus, constructing tables bring greater accuracy in the analysis.
7. **Gives Overview :** Tabulating the data gives the reader an overview about the data without getting into the details of its collection process. It acts as a summary by providing all information in a very systematic and compact form.
8. **Gives a Bird's Eye View :** The data once arranged in a suitable form, gives the condition of the situation at a glance, or gives a bird's eye view.

Q.19. What are the different types of diagrams?

Ans.

Types of Diagrams

There are different types of diagrams that are used for representing data. Diagrams can be categorized into :



1. **One-Dimensional Diagrams :** These diagrams represent only one dimension like only the height or only the width. These diagrams are represented in the form of bar or line chart and can be divided as follows :
   (i) Line Diagram
   (ii) Simple Bar Diagram
   (iii) Multiple Bar Diagram
   (iv) Sub-Divided Bar Diagram
   (v) Percentage Bar Diagram
2. **Two-Dimensional Diagrams :** In two-dimensional diagrams, more than one dimension is taken into account. Such diagrams are called area diagrams or surface diagrams. The important types of area diagrams are :
   (i) Rectangles (ii) Squares (iii) Pie-Diagrams
3. **Three-Dimensional Diagrams :** In these diagrams, three dimensions variables like the length, breadth and width are taken into account. These diagrams are in the form of cubes, spheres, cylinders and blocks.

Q.20. Define simple bar charts with example.

Ans.

Simple Bar Charts

Bar charts are used to present the comparison of various categories in the form of horizontal and vertical bars. It is also known as bar diagrams or bar graph. Using the bar chart, one can do the comparison of various items.

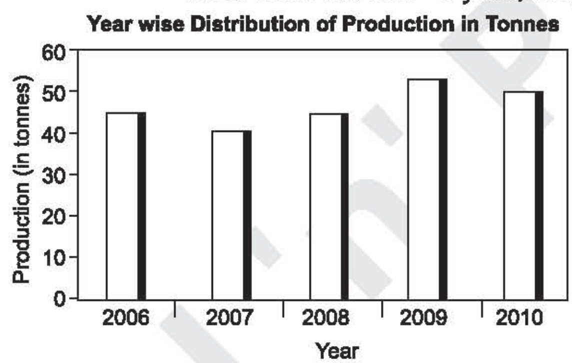
A bar graph contains the axis, scales, labels and bars. It is best way for the comparison of two or more values. It is used to show any kind of information, like number of employees in the last few years, number of boys and girls in a college, sales of a company, etc. Few bar graphs are clustered in various groups (these clusters have two or more than two bars), and some bar graphs show bars which are divided in various subparts to display the total effect.

Example : Represent the following data by a bar diagram :

Year	Production (in tonnes)
2006	45
2007	40
2008	42
2009	55
2010	50

On x-axis : 1.5 cm = 1 year; On y-axis 1 cm =10 tonnes

Sol.



Q.21. Draw a Pie diagram for the following data of production of sugar in quintals of various countries.

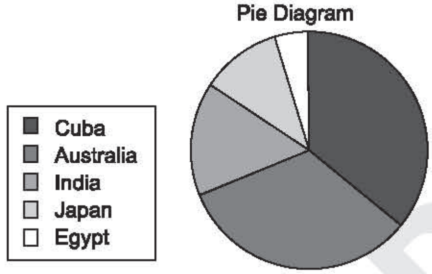
Country	Production of Sugar (in quintals)
Cuba	62
Australia	47
India	35
Japan	16
Egypt	6

Sol. The values are expressed in terms of degree as follows :

Country	Production of Sugar	
	In Quintals	In Degree
Cuba	62	$\left(\frac{62}{166} \times 360°\right) = 134$
Australia	47	$\left(\frac{47}{166} \times 360°\right) = 102$
India	35	$\left(\frac{35}{166} \times 360°\right) = 76$

Japan	16	$\left(\frac{16}{166}\times 360°\right)=35$
Egypt	6	$\left(\frac{6}{166}\times 360°\right)=13$
Total	**166**	**360**

Pie Diagram



Q.22. What are the advantages of graph?

Ans.

Advantages of Graph

1. **Certain Statistical Measures can be Ascertained with Care :** With the help of special kind of graphs, positional averages like median, quartiles, mode, etc., can be determined. Graphic presentation of statistical data is also helpful in interpolation, extrapolation and forecasting. Correlation between two sets of data can be studied.
2. **Comparison is Made Easy :** Comparisons can be made between two or more phenomena very easily with the help of a graph.
3. **Simplest Method of Presenting Data :** Graphical method is probably the simplest method of presenting statistical data. It saves time and energy of both, the statistician as well as the observer. *For example,* a busy business executive hardly has time for study of data relating to sales, purchase, cost or profit and find out their trend. If these data are presented on a graph, the basic trend can be understood in a very short time.
4. **No Knowledge of Mathematics Required :** Even a common man can understand the message of data from the graph. To understand a graph, no special knowledge of mathematics is needed.
5. **Graphs are Attractive, Interesting and Impressive :** A graph is more attractive than a table of figures.

As the features of data become visible at a glance in a graph, one can study very easily the tendency and fluctuation of data.

Q.23. Write differences between diagram and graph.

Ans.

Differences between Diagram and Graph

S.No.	Diagram	Graph
1.	Ordinary paper can be used.	Graph paper is needed.
2.	It is attractive and is easily understandable.	It needs some effort to understand.
3.	It is appropriate and effective to represent one or more variables.	It creates problem.

4.	It cannot be used for interpolation and extrapolation technique.	It is helpful in interpolation and extrapolation techniques.
5.	Median and mode cannot be estimated.	The value of median and mode can be estimated.
6.	It is used for comparison only.	It represents a mathematical relationship between the two variables.
7.	Data are represented by bars, rectangles.	Data are presented by points or lines of different kinds—dots, dashes, etc.
8.	Diagram are used for publicity as they are attractive. They give only approximate information. To a statistician or a researcher, diagrams are not helpful in analysis.	Graphs are very much useful to statisticians or researchers in analysis.

SECTION-C LONG ANSWER TYPE QUESTIONS

Q.1. Discuss the meaning and scope of statistics.

Ans.

Meaning of Statistics

Broadly speaking the word statistics means :

(i) Statistics or Statistical Data,

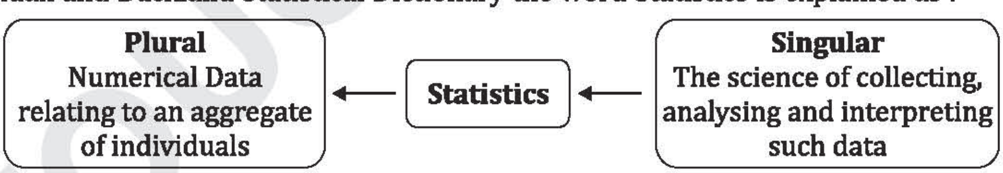
(ii) Statistical Methods or Science of Statistics,

(iii) Statistics, plural of the word Statistic.

The Common man refers the word statistics as numerical data. For example, Statistics of National Income Statistics of Automobile Accidents, Employment Statistics, Market Statistics, Production Statistics, Import and Export Statistics, etc.

In this sense the word "Statistics' is considered in plural noun form. In a wide and singular form Statistics refers as a Science of Statistics. In this sense the science of statistics is commonly known as *Statistical Methods*.

In Kendall and Buckland Statistical Dictionary the word Statistics is explained as :



The three meanings of the word 'Statistics' are contained in the statement due to Tate :
"You compute statistics by statistics from statistics."

Scope/Applications of Statistics

Statistics has affected almost all areas of life as it covers simple households to big businesses, even government also. Some of the areas where statistics has been used are as follows :

1. **Statistics and Mathematics :** Statistics can be considered as a branch of science which is conceived on the foundation of mathematics. A person should have some knowledge of mathematics for understanding the basic fundamentals of statistics.

   According to *Connor*, "Statistics is a branch of Applied Mathematics which specialises in data."

According to *W.I. King*, "Statistics may properly be considered as a branch of mathematics in as much as it attempts to formulate definite rules of procedure applicable in handling groups of data of many different varieties."

2. **Statistics and the Government :** Statistics has been used extensively since the beginning of organised society. Statistics has been used by the administrative heads and rulers of the states in the form of collecting data on different aspects for the purpose of formulating sound military and fiscal policies. This data includes figures of population, tax collection, military strength, etc. In the present times, the government is the biggest collector of data as well as the biggest user of statistics. A huge amount of data is collected and interpreted by various departments of the government for formulating efficient policies and decision-making.
3. **Statistics and Physical Science :** The use of statistical methods is continuously increasing in the field of physical sciences such as Biology, Physics, Chemistry, Astronomy, Medicine, etc. Statistical data are collected from different results of different experiments.
4. **Statistics and Economics :** Statistics is used as an important tool in economics study and research. Economics is mainly concerned with production and distribution of wealth and also savings and investments. Statistical tools are used in the following economic interest areas :
   (i) Statistical methods are used for measuring and forecasting Gross National Product (GNP).
   (ii) Statistical studies of business cycles give a clear picture about the economic stability.
   (iii) Economic policy-making depends on the statistical analysis of population growth, unemployment figures, rural or urban population shifts, etc.
   (iv) Optimum utilisation of resources is possible with the use of econometric models which uses statistical methods.
   (v) For the study of finance, banking, consumer savings and credit availability, financial statistics is necessary.
5. **Statistics and Research :** In the current scenario statistics is an essential part of research study. Improvement in knowledge has been possible because experiments are carried out with the help of statistical methods. For example, experiments about crop yields and their correlation with different types of fertilisers and different types of soil are designed and studied with the help of statistical techniques. In the current time, statistical methods are used in all types of research work including medicine and public health.
6. **Statistics and Natural Sciences :** Statistics is also very important in the study of natural sciences such as astronomy, biology, medicine, meteorology, zoology, botany, etc. *For example*, for diagnosing the exact disease of a patient, the doctor must believe on real data such as the body temperature, pulse rate, blood pressure, etc.
7. **Statistics in Education :** Statistics is used extensively in the field of education because research has become a common feature in all branches of activities. In education, statistics plays a vital role in formulation of new policies for new courses to be started,

consideration of infrastructural requirements for new courses, etc. Apart from this, there are many people involved in research work who test past knowledge and develop new knowledge with the help of statistics.

8. **Statistics in Astronomy :** Astronomers were one of the first groups of people who used statistics in the study of movement of heavenly bodies and eclipses and other such astronomical issues. Astronomers earlier relied heavily on estimation but later statistics helped to turn these estimations into accurate ideas.
9. **Statistics in Accounting and Auditing :** In accounting, exactness is an essential component but for decision-making purposes, approximation is taken into account. The current asset value is calculated on the basis of its current values and the corrected values are determined with the help of current purchasing power of money or the current value of it, while taking depreciation into consideration. This is done through the use of price indices which are based on the collection of statistics.

   For determining the trend of future profits it is required to use the study of correlation analysis between the profits and dividends. In auditing, sampling is generally used as it is not possible to examine voluminous transactions due to lack of human resources. An auditor will first co-relate the past error percentage with the current error rate after conducting a pilot audit. After this, the auditor decides on the sample size of books to be audited.
10. **Statistics in Planning :** Statistics is important for efficient planning in all modern economies, especially in the developing countries. This helps in successful planning by taking into consideration the correct analysis of complex statistical data. The plans, adopted for the economic development of a country, are also made on the basis of statistics available about the different economic problems being faced.
11. **Statistics and Commerce :** Commerce is now very much dependent on statistics for success of businesses. Any businessman cannot afford to over stock or under-stock the goods. In the initial stage it is required to do a market survey using statistical techniques to understand the changing tastes of the consumers because a number of multinational companies have also entered with new products and services. Thus, the statistical techniques help in providing the present conditions and forecast the future.
12. **Other Areas :** Statistics is used by many businesses such as insurance companies, stock brokerage houses, banks, public utility companies and so on. It also helps politicians study their winning chances from a constituency through the use of sampling techniques in random selection of voter samples and studying their preferences on issues and policies.

Q.2. What is importance of statistics?

Ans.

Universal Utility *Or* Significance/Importance of Statistics

The subject of statistics started in the form of a science of state craft whose object was to collect significant facts for the state with which help the state-head took decisions about the increase or decrease of number of army personnel, food for them, salaries of state servants, land tax etc., as per need. This is why this branch of knowledge was called Science of State craft, or Science of Kings or Plitical Arithmetic. Today statistics studies affect every person and touches several aspects of his life. *Walker* have rightly stated that, "To a very striking extent our culture has become a statistical culture." In every area and activity of society the importance of statistics is acknowledged and for this reason it can be said that statistics is no more a science of kings but has become the Science of the Human Welfare. Statistical interpretation is very essential in all subjects like, Astronomy, Sociology, Mathematics, Economics, Education, Psychology, Physics, Chemistry, Medical Science ete. There is hardly any field whether it be administration, insurance banking or planning etc., whose statistical methods are not applicable.

The science of statistics is growing in importance everyday. It is now being used in almost every field of human activity. Its application has become so wide that no branch of human knowledge from the graphic arts to astrophysics and from numerical composition to missile guidance escapes its approach. Statistics provides tools and techniques for research workers in analysis of problems in both natural and social sciences, economics, commerce, industry, trade, biology sociology, psychology, medicine.

According to *Tippett*, "For some subjects statistics provides ideas of basic importance, for some it provides methods of investigation. In one way or the other, or in both ways, statistics has an impact on most other branches of knowledge." That is why it is said that, "Sciences without statistics bear no fruit, statistics without sciences has no root. Statistics is required to understand the various problems relating to business, social policy or state crafts."

According to *Wallis and Roberts*, "Statistics is a tool which can be used in attacking problems that arise in almost every field of empirical inquiry. Statistical methods are used effectively to the most diverse subjects, ranging from minor business and personnel decision to obstruse questions of pure sciences and scholarship."

Economist depends on data for the study of National Income, Production, Purchasing Power of Money, etc. The information related with human progress : the prosperity and downfall, profit and loss, inflation and deflation, fluctuations in prices, unemployment, etc. are discussed with the help of statistics. The data is to be given to understand any law of economics. The estimates of future are made with its help. To emphasize the universal utility of statistics *Croxton and Cowden* said, "Today there is hardly a phase of endeavour, which does not find statistical devices at least ocassionally useful."

Anatolle France also said, "If they (countries) do not count, then they will not be counted in the world."

Wallis and *Roberts* opined, "Statistical methods are used effectively to the most diverse subjects, ranging from minor business and personal decision to obstruse questions of pure science and scholarship." According to Bowley, "Knowledge of statistics is like the knowledge of foreign language or algebra. It may prove of use at any time under any circumstances."

For any person or in any condition the knowledge of statistics is a must. To emphasize the importance of statistics *Tippett* wrote, "Statistics affects every body and touches life at many points."

It is obvious from the above discussion that statistics has universal utility.

To be more specific we may discuss the application/significance/importance/utility of statistics in various fields.

Q.3. Write about Statistical Investigation. Explain the organization of statistical investigation.

Ans.

Statistical Investigation

Statistical investigation is the search for knowledge on a specific area using statistical approaches on the basis of methodically collected and analysed data. It's based on mathematical principles and facts so it's a scientific study. This research is necessary because statistical analysis cannot be performed without statistical data. Statistical investigation is the study of a problem using statistical methods. The data on the problem form the basis of the investigation. Therefore, data should be collected on related problems.

The person who does statistical research is known as investigator and respondents are the people from whom investigator collected the information. Data collected in this way must be analyzed and processed using suitable statistical systems and analyzed data provide valuable information for the problem that is under investigation.

Organisation of Statistical Investigation

Statistical investigation is an extensive and lengthy process. It goes through various phases from the initial planning to the preparation of the final report. Following are the significant steps of statistical investigation :

1. **Planning a Statistical Enquiry :** Planning is the first step of statistical investigation. Planning should be based on the objectives and scope of the investigation. The researcher must decide in advance about :
   - (i) Type of Enquiry to be carried out,
   - (ii) Source of information,
   - (iii) Duration of the study and
   - (iv) Unit of study.
2. **Collection of Data :** The data collection is the next step after planning. Data can be collected using following sources :
   - (i) Primary source, and;
   - (ii) Secondary source.

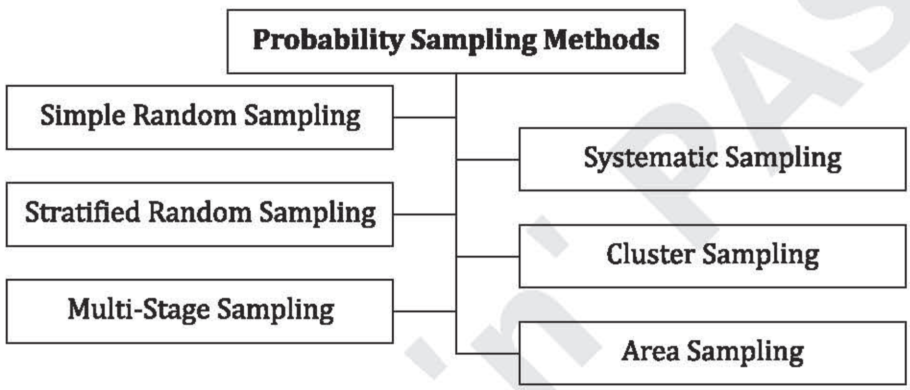
   Collected data must be correct and precise. There should be no bias in data collection.
3. **Editing of Data :** Organizing data means arranging data systematically. The collected data should be refined from unnecessary data. Then collected data must be classified based on certain common characteristics.
4. **Presentation of Data :** Organized data can be presented in the form of charts and graphs. It makes it easy to understand the properties and nature of the data.
5. **Analysis of Data :** The tabulated data should be analysed to draw a conclusion. Several statistical methods such as mean, dispersion, correlation, regression, index, time series, etc. are used for analysing the data.
6. **Interpretation of Data :** The final step in statistical research is data interpretation. To interpret means to draw conclusions. If the interpretation is not done correctly, it can lead to wrong conclusions. Hence, one must be very careful in interpreting the data in order to draw a valid conclusion.

Q.4. What do you mean by probability sampling method? Explain the different probability sampling methods.

Ans.

Probability Sampling Methods

The method, in which all units of the universe are given equal chance of being selected in the sample, is known as Probability Sampling. There is an assurance of the results in terms of probability that are obtained through probability or random sampling. The significance of the results lies in measuring the errors of estimation obtained from a random sample which brings predominance of the random sampling designs over the intentional sample design. The various probability sampling methods are as follows :



1. **Simple Random Sampling :** This is the most famous and simple method of sampling where each unit of the population is equally probable of getting included in the sample. Let us consider that the size of the population is '*N*' from which n units are to be selected at random for a sample such that $^{N}C_n$ sample has the probability of being selected equally. Simple random sampling says that : There is an equal chance for each element of the population to be included in the sample and the choices are independent to each other. Each possible sample combination has an equal chance of being chosen.

   Methods of Simple Random Sampling—Some of the common methods of drawing simple random samples are :

   (i) **Lottery Method :** This method includes the following steps :

   (a) Let *N* be the size of population and a sample of size *n* is to be drawn.

   (b) These are mixed carefully in a bag; bowl or some other container and *n* items are selected either by replacement or without replacement in one stroke.

   (c) The number of units in the population bearing the numbers on the items drawn in step (c) constitutes the desired random sample.

   (d) The application of this method is simple and easy but as the size of the population increases to infinity this method is unfeasible.

   (ii) **By Using Random Numbers :** The random number table that helps in selecting a sample has been constructed by some experts. The most popular of all the existing tables, **Tippett's Tables** is very popular and in use. Out of the given population, random selection of the samples is done through these numbers. This table ranges from 0 to 9 having an equal chance of appearing in any position of the table. The actual utilisation of the random numbers is for large population where the random number table consists of several columns and rows out of which one is selected at random and then the samples are continuously selected in sequence for the desired size of the samples. Without any bias it provides a set of random numbers.

2. **Systematic Sampling :** After the selection of one unit at random from the universe the other units are selected systematically at a specified interval of time. This method is applicable when the size of the population is finite and on the basis of any system the units of the universe are arranged such as alphabetic arrangement, numerical arrangement, or geographical arrangements.
3. **Stratified Random Sampling :** In the Stratified Random Sampling, the sample is selected from different homogeneous strata or parts of a universe instead of heterogeneous universe as a whole. The summary of this sampling procedure is as follows :
   (i) The sampled universe is divided (or stratified) into groups that are mutually exclusive and include all items in the universe.
   (ii) A simple random sample is then chosen independently from each group or stratum.

   There is a difference in the process of stratified random sampling from that of simple random sampling. In simple random sampling the sample items are chosen at random from the whole universe whereas in stratified sampling the design of the sample is the selection of the separate random sample from each stratum. The distribution of the sample among strata is based on chance in simple random sampling.

   Formally, divide the population into non-overlapping groups (*i.e.,* strata)

   $$N_1, N_2, ..., N_i$$

   Such that $N_1 + N_2 + ... + N_i = N$

   Then randomly choosing a sample of $f = \frac{n}{N}$ in each stratum, where f is the sampling fraction.
4. **Cluster Sampling :** According to this method there is further noticeable sub-division of the universe into clusters. Simple random sampling is performed and clusters are drawn accordingly constituting a sample of all the units belonging to the selected clusters. *For example,* if we have to conduct a survey in the city of Mumbai, then the city may be divided into, say, 40 blocks and out of these 40 blocks, 5 blocks can be picked up by random sampling and the people in these five blocks are interviewed to give their opinion on a particular issue. The clusters chosen should be small in size, *i.e.,* more or less the same number of sample units should be there in each cluster. This method is used in the collection of data about some common traits of the population.
5. **Multi-Stage Sampling :** Modification of cluster sampling is multi-stage sampling where in cluster sampling, a sample is constituted by all the units selected in a cluster

but in multi-stage sampling the selection of the sample units is in two, three, four stages.

Firstly, the universe is divided into first-stage sample units and then further sub-divided into second stage units out of which another sample is selected and similarly third and fourth stage sampling is performed in the same way.

For example, in an urban survey first stage sampling will be of selection of towns and then for each selected towns households will be taken as sub-sample and depending upon the need, individuals will be selected as samples of third-stage sampling.

6. **Area Sampling :** The area sampling is a form of multi-stage sampling in which maps instead of lists or registers is used as 'sampling frame'. It is commonly used in those countries which do not have proper sampling frame like a population list. For geographic sub-divisions, 'clusters sampling' is the other name of 'area sampling'. The cluster of units based on geographical area is the primary sampling units known as 'cluster designs' which are famous as area sampling. The positive and negative features of cluster sampling are also applicable to area sampling.

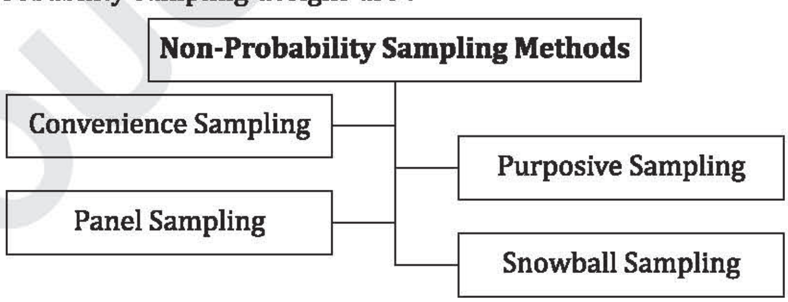
Q.5. What is Non-Probability sampling method? Define its various types.

Ans.

Non-Probability Sampling Methods

Non-probability sampling is that type of sampling procedure which does not have any ground for estimating the probability that whether or not each item in the population has been included in the sample. There are different names of non-probability sampling such as deliberate sampling, purposive sampling and judgment sampling. In this type of sampling, the researcher deliberately selects items for the sample and the choice of researcher regarding the item is provided more weightage.

In other words, under non-probability sampling the organiser of the inquiry purposively chooses specific units of the universe to constitute a sample on the basis that the small portion selected by him, out of a huge one is typical or representative of the whole universe. The various non- probability sampling designs are :



1. **Convenience Sampling :** On the basis of convenience and approachability, the choice of the sampling units by the researcher, is known as 'convenience sampling'. Samples that are selected accidentally are called 'accidental samples'. Because of the selection procedure (units are selected from their actual place) it is called as 'sample of the man in the street'. Due to their accessibility, samples units are selected. *For example*, by adding the new product in the nearby suitable shops, the potential of the product is tested. This is accomplished by observing the purchasing and selling report of the product.

2. **Purposive Sampling :** A non-probability sample which follows certain norms is called purposive sampling. Purposive sampling is basically of two types :
   (i) **Judgement Sampling :** The study which is based on the parameters of population, where the units are selected by a researcher or some other expert on his/her judgement, is called 'judgement sampling'. This technique of sampling is appropriate in the situation where the study of the population is difficult to locate or there are members who are comparatively better than others for an interview in terms of knowledge or interest.
   (ii) **Quota Sampling :** Quota sampling is the most commonly used non-probability sample designs, which is most comprehensively used in consumer surveys. Principle of stratification is also used by this sampling method. In stratified random sampling the researcher begins by building strata. The common bases for stratification in consumer surveys are demographic, *e.g.,* age, gender, income and so on. Compound stratification is generally used, *e.g.,* gender-wise age groups.
3. **Panel Sampling :** In panel sampling a group of participants are selected initially by random sampling method and the same group is asked for the same information repeated number of times during that period of time. This sample is semi-permanent where members are included repeatedly for iterative studies. In this sampling, there is a facility of selecting and contacting samples that fit in getting high response rate, even by mail.
4. **Snowball Sampling :** When the characteristic of the desired sample is limited then the special non-probability method is applicable. In this method, it is difficult to locate the respondents because it will be very costly. Depending on the referrals of the initial subjects snowball sampling generates additional subjects. Though this technique is biased and unable to represent a good cross-section from the population but dramatically it reduces the search cost.

Q.6. What is meant by primary data? Explain in brief the various methods to collect primary data.

Ans.

Primary Data

The data collected by the researcher himself for finding the solution of a particular problem or situation, is known as 'primary data'. This type of data is characterised by its originality as it is freshly collected. Various organisations conduct surveys, observations, interviews, etc, and as a result generate primary data. Although secondary data provides a basic understanding to the research problems, but sometime, it becomes necessary to collect primary data as the previously generated secondary data may not serve the purpose. Just like secondary data, researchers should also take additional care while collecting primary data such that it is accurate, reliable, and unbiased. For collecting primary data, researchers need to take many decisions regarding proper selection of relevant sources, sampling techniques, research tools, etc. To conduct any research effectively and produce valid results, researchers should collect primary data as it contains current and exact information about the incident or event.

One of the major benefits of primary data is that its validity and reliability can be verified by other experts. There are many ways to collect primary data such as observation, interviews, group discussions, case studies, etc.

Methods of Primary Data Collection

Major methods for collecting primary data are as follows :

1. **Survey :** This technique is one of the most common and widely used techniques for collecting primary data. Survey can be conducted using various methods such as using mails, telephones, internet, face-to-face, etc. The selection of survey method relies upon various factors such as the nature of population to be studied, size of sample, allotted time, allocated budget, etc.
2. **Interview/Direct Personal Investigation :** This method is also known as the direct personal investigation. Interview is the exchange of ideas which takes place between two or more people with the purpose of getting information from the respondent. The responses of the interviewee are recorded and compiled to get a better insight into the research problem. Interview can be conducted through various methods such as personal interview, telephonic interview, mail interview, panel interview, etc.
3. **Observation :** Another technique for gathering primary data is observation. When the researcher records information about a person, organisation, or situation, without making any personal contact, it is known as "observation method". In this, the researcher or the field executive observes the activity of the concerned person or organisation, to draw a pattern of behaviour or response to a particular incident. Sometimes, an artificial environment is created to collect the actual responses of the participants.
4. **Experimentation :** An important method to collect primary data is experimentation. In experimentation, the causal relationship is determined and analysed between variables.
5. **Questionnaire :** Questionnaire is a method of primary data collection that comprises of different sets of questions to get information from the respondents for the problem identified by the researcher. It is an orderly arrangement of questions that helps in generating the required primary data which can then be analysed and interpreted to solve a research problem. Using questionnaires, the researchers can ask direct questions to a number of people. The success of a questionnaire depends on the skills of researcher in framing it appropriately.
6. **Schedule :** Schedule is a technique used for collection of data by the respondents during the interview. Schedules comprise of relevant queries with blank spaces/slabs that have to be answered by the respondents so as to collect the needed information. Schedule refers to the group of questions regarding a particular topic which are being asked to the respondents in person by the interviewer.
7. **Other Methods :** Other methods for collection of data are described below :
   (i) **Warranty Cards :** Warranty cards are generally used by the dealers of consumer durables to get the feedback of products from their consumers. These are the postal sized cards placed within the package of product. These cards contain various questions regarding the performance of product and to know the needs of consumers. Customers are requested to fill and mail it back. It helps in new product development for the manufacturer.

(ii) **Auditing :** Auditing is a technique for assessing the performance and current position of any department or the organisation. Sometimes, it is also used for understanding the market and buying behaviour of customers. Distributors or manufacturers use this tool for gaining the competitive advantage and satisfying the need of customers. It is also used by the researchers for inspecting the products, services or food purchased by consumers, also known as pantry audit.

(iii) **Simulation :** Simulation is a quantitative technique for data collection. It is the creation of an artificial environment resembling a real life situation. This real life situation is simulated by using various mathematical equations and variables. Researchers can determine the relation between different variables by altering one of the variables and finding its effect on the others.

Q.7. What are secondary data? Discuss in brief the various methods to collect secondary data.

Ans.

Secondary Data

When a researcher uses data which are previously collected by some other researchers, institutions, or agencies for their own purposes are called secondary data. The researchers collect secondary data either from an internal source of an organisation, or from the published sources like reports and journals. The purposes of data may vary from that of the current study.

Methods of Collecting Secondary Data

The secondary data can be collected in following manner :

1. **Internal Secondary Data :** Secondary data generated within the research conducting organisation is known as internal secondary data.

   Data generated within the organisation can be either formal or informal. Formal data are generated periodically in a structured layout such as reports of various departments, half yearly reports, etc.

   On the other hand, informal data are not periodically generated such as conceptual booklets, new policy frameworks, etc. Formal internal data can be collected from following major sources :

   (i) **Sales Analysis :** Sales analysis reports generated within the organisation are important internal source of secondary data. These reports contain the information about the sales pattern and fluctuations in market position. These can be very useful in drawing the solution of related problems.

   (ii) **Invoice Analysis :** The invoices of an organisation also act as a secondary data source. These invoices help in understanding the sale and purchase pattern of the organisation in different situations or scenarios. The information collected through the invoices may be summarised carefully to reach a particular solution. Various data related to customer can be obtained with the help of invoices, such as, name of customer, type of product, location of product delivered, etc.

   (iii) **Financial Data :** Researchers can get a lot of financial data recorded within the organisation. These records may contain the information regarding production cost, storage cost, transportation cost, sales cost, etc. These data are very useful data for marketing research. These financial data are periodically generated from time to time, and hence are updated.

(iv) **Transportation Data :** The transportation data regarding the routes, vehicles, loads, etc., provide a lot of information regarding the transportation activities. These data allow the researchers to analyse the trade-offs between various costs and determine the ways to get maximum financial benefits.

(v) **Storage Data :** Various costs associated with storage, such as, handling cost, maintenance cost, etc., are the important data that are generated within the organisation. These data help the researchers in analysing various pros and cons related to the storage of materials and therefore suggesting suitable methods to be adopted.

2. **External Secondary Data :** Sometimes, important secondary data is not found within the organisation. The secondary data derived from different sources outside the organisation is known as "external secondary data". Some important sources of secondary data are as follows :

(i) **Libraries :** Library is one of the external secondary data sources, which the researcher may use to collect the necessary information for the research. Different kinds of libraries provide a range of data for the research. Books related to research topics, journals, magazines, research papers, etc., are available in various libraries, maintained by different organisations and institutions.

(ii) **Literature :** A variety of literature is available on different subjects and issues. These literatures are the result of extensive research practices. There is plenty of valuable information in such kind of sources, which can be utilised for the resolution of current research problems.

(iii) **Periodicals :** Business periodicals are published at regular time intervals, viz., yearly, half-yearly, bi-monthly, fortnightly, quarterly, etc. The secondary data are published by various government and non-government agencies regarding finance, trade, transport, industry, labour, etc. These periodical contain various trends, future prospects, opportunities in market, etc., which can be used by researchers in their current research problem.

(iv) **References and Bibliography :** The references and bibliography of a particular research or journal can be a useful resource for deriving secondary data related to specific issue. Researchers can take a huge amount of data which can then be analysed to get deeper insight

(v) **Census and Registration Data :** Data collected through census and different registration programs may become very useful in deriving secondary data. As this data is collected through extensive effort and field work, it contains the appropriate information about various issues like, agriculture, trade, transport, banking, etc.

(vi) **Government Departments :** The information available from government departments may be utilised as secondary data in research process. Government departments can provide various information regarding position and growth of different sectors of an economy like finance, banking, trade and transport, agriculture, etc.

(vii) **Private Sources :** There are many organisations which publish the statistically processed data for further use. These are the private institutions which perform primary research about particular events or situations and compile the final facts and figures. Some of the examples of such sources are Economic Times, Financial Express, Indian Marketing Association, etc. The researchers engaged in current marketing research can utilise the information available from these institutions by purchasing journals, magazines, newspapers, etc., which are publicly available.

(viii) **Commercial Research Institutions :** Some institutions in the market deal in purchasing and selling of different kind of data or information, which are collected through research. Many market research institutions are in the business of providing statistically processed data or information, by taking help of secondary data or by conducting fresh surveys.

(ix) **International Organisations :** Several international organisations like World Health Organisation, World Bank, International Monetary Fund, International Labour Organisation, Asian Development Bank, etc. are helpful in deriving required information or data about a particular research. These organisations have plenty of information or resources to provide data about issues like population, inflation, agriculture, education, labour problems, child problems, women development, trade and transport, etc.

Q.8. What is data editing? Discuss the essentials of editing and stages of editing.

Ans.

Data Editing

Raw data have various errors and omissions that can arise during the process of data collection. Editing corrects these errors so that readers are not confused or misled. Editing changes the raw data into a displayable format so that further analysis and interpretations can be carried out efficiently.

Editing is the process of reviewing, correcting, and changing disorganized data so that data becomes more relevant and meaningful. This process reduces the confusion that can arise from incorrect and inaccurate data. It also confirms that there are no omissions or errors and data is in readable form. This process also maintained flow of information throughout the entire survey of research process.

Essentials of Editing

The editing should be done taking into account the following points :

1. **Completeness :** For effective editing, it must be taken into account that no omissions may be made. It is, therefore, imperative that all questions are asked and that the corresponding answers are recorded accordingly. If the data is missing, the researcher can derive the missing data from other data in the questionnaire or supplement the data by retrieval.
2. **Accuracy :** The editing of the data must take into account the accuracy of the recorded data. Researchers have to check the reliability of the answers themselves when collecting the data, which is not always possible. The accuracy of the answers can be estimated using "check questions" in the questionnaire specifically for important data. The "check questions" can directly assess the ambiguity of the answer or help the researcher to derive the correct answers. Researchers can also complete the answer with other related questions on the questionnaire.

3. **Consistency :** Maintaining the consistency of answers or responses is the one of the important considerations when researchers editing the data. It must be verified that the answers are given in the same way as the questions were asked. This means that all answers to a question must be answered in the same way by all respondents. Inappropriate and inadequate answers lead to confusion and misinterpretation; therefore, researchers need to ensure that the answers lead to reasonable conclusions.

Stages of Editing

The data processing can be done in two steps :

1. **Field Editing :** This editing processing is carried out during the collection of data. All collected answers or responses are checked for errors and omissions. By collecting the data, researchers instantly verify the consistency, accuracy, and completeness of the responses. Following are the two ways to perform field editing :

   (i) **By the Researcher :** For shortage of time, the interviewers record the answers in the form of symbols or short notes during the interviews.

   After completing the interview, the researcher reviews the answers, if required corrects them, and fills out the questionnaire specifying the answers to each question.

   (ii) **By the Supervisor :** This form field editing is carried out by the supervisor of the interview team, whose task is collecting data from a sample of respondents. The supervisor is responsible for the quality of the data. To achieve this task, he always ensures that the all investigators perform their jobs with honesty. This is achieved by reviewing the interview answer and correct if any errors is found in the first step.

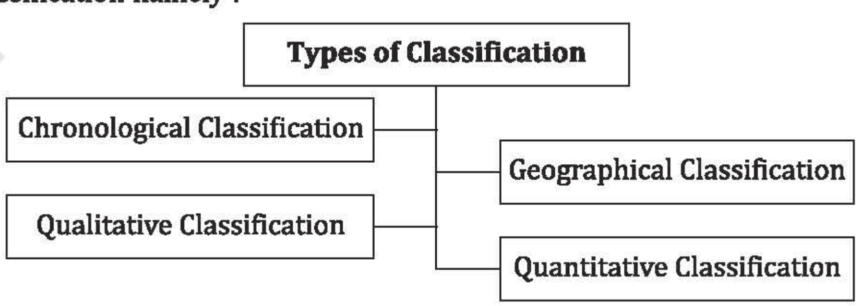
2. **Office/Central Editing :** When completed forms have arrived at the main office, a person or a team will process these forms. This process is known as the "Office Editing" or "Central Editing". The Office editing is much more accurate than the Field editing. It is best suited for mail surveys because no field survey is performed in this case.

Q.9. Discuss the different types of Classification.

Ans.

Types of Classification

Statistical data are classified in respect of their characteristics. Broadly there are four basic types of classification namely :



1. **Chronological Classification :** In chronological classification the collected data are arranged according to the order of time expressed in years, months, weeks, etc. The data is generally classified in ascending order of time. *For example,* the estimates of birth rates in India during 1970-76 are :

Year	1970	1971	1972	1973	1974	1975	1976
Birth Rate	36.8	36.9	36.6	34.6	34.5	35.2	34.2

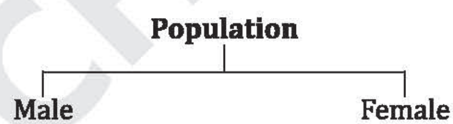
2. **Geographical Classification :** In this type of classification the data are classified according to geographical region or place. *For example,*
   (i) The production of paddy in different states in India, and
   (ii) Production of wheat in different countries, etc.

   For example, consider the data given in the table below.

Country	America	China	Denmark	France	India
Yield of Wheat in (kg/acre)	1925	893	225	439	862

3. **Qualitative Classification :** In this type of classification data are classified on the basis of some attributes or quality like gender, literacy, religion, employment etc. Such attributes cannot be measured along with a scale. For example, if the population to be classified in respect to one attribute, say gender then we can classify them into two classes namely males and females. This classification is known as **dichotomous classification.**

   Similarly, they can also be classified into 'employed' or 'unemployed' on the basis of another attribute 'employment'. *For example,* a simple qualitative classification may be shown as under :

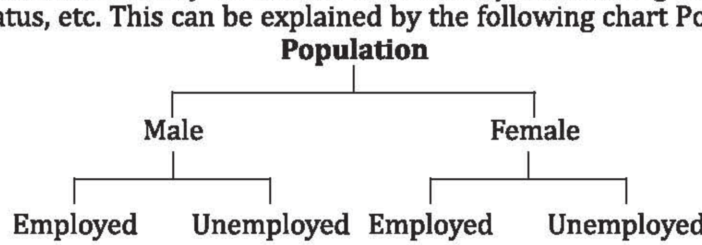


The classification, where two or more attributes are considered and several classes are formed, is called a manifold classification. *For example,* if we classify population simultaneously with respect to two attributes, *e.g.,* gender and employment, then population are first classified with respect to 'gender' into 'males' and 'females'.

Each of these classes may then be further classified into 'employment' and 'unemployment' on the basis of attribute 'employment' and as such Population are classified into four classes namely.

(i) Male employed, (ii) Male unemployed, (iii) Female employed, (iv) Female unemployed

Still the classification may be further extended by considering other attributes like marital status, etc. This can be explained by the following chart Population :



4. **Quantitative Classification :** Quantitative classification refers to the classification of data according to some characteristics that can be measured such as height, weight, etc. For example, the students of a college may be classified according to weight as given below :

Weight (in lbs)	No. of Students
90-100	50
100-110	200
110-120	260
120-130	360
130-140	90
140-150	40
Total	**1000**

In this type of classification there are two elements, namely :

(i) The variable *i.e.,* the weight, and

(ii) The frequency *i.e.,* the number of students in each class.

There are 50 students having weights ranging from 90 to 100 lb, 200 students having weight ranging between 100 to 110 lb and so on.

Q.10. What is statistical series? Discuss its various types.

Ans.

Statistical Series

Statistical series are created for properly management of the collected and classified data. *For example,* if the data consisting of the height of 10 students are systematically collected then it is known a statistical series.

According to *Secrist,* "A series as used statistically may be defined as things or attributes of things arranged according to some logical order".

According to *L.R. Conner,* "If two variable quantities can be arranged side by side so that the measurable differences in the one correspond to the measurable differences in the other, the result is said to form a statistical series".

Types of Statistical Series

I. On the basis of General Characteristic : According to the general character there are three types of statistical series :

1. Time series or Historical series, 2. Spatial series, 3. Condition series

1. **Time series :** When the data are classified and arranged with regard to time, the series is called a time series or historical series. The important factor in such series is chronology. So they are also called chronological series. *For example,* population of India :

Year	1991	1992	1993
Population (crore)	43.9	34.0	68.4

2. **Spatial series :** When the data are presented with reference to space (*i.e.,* geographical division), the series is referred to as spatial and is also called as geographical series. *For example,* the number of Industries Statewise :

State	U.P.	M.P.	Rajasthan
No. of Industries	75	40	35

3. **Condition series :** When the data are presented with reference to any criterion (other than time, place or area), the series is called a condition series. *For example,* marks obtained by students in an examination, series of index numbers of various heads, data pertaining to physical conditions such as height, weight, age, etc. *For example,*

Receipts of Governments of India for Taxes (2008-09)

	(₹ in crore)
Import Duty	465.00
Nigam Tax	342.00
Central Excise Tax	1,814.40

II. On the basis of Shape (Construction) : On the basis of their construction the statistical series are of three types :

1. Individual series, 2. Discrete series, 3. Grouped series.

1. **Individual series :** A series of individual observations where the items are listed singly is known as an individual series. *For example,* the marks obtained by four students :

Name of the student	*A*	*B*	*C*	*D*
Marks obtained	20	30	31	24

The names of the students may be arranged in alphabetical order.

An individual series arranged in an *ascending* or *descending* order of magnitude is called an *array*. The process of arranging the data in an ascending or descending order is known as *arraying of data*. The array is a useful device for studying the main features of the data when the number of items is small. However, arraying does not reduce the volume of the data.

The marks obtained by four students in ascending order are 20, 24, 30, 31 and in descending order are 31, 30, 24, 20.

In ascending order, the smallest value is written in the first place, then next larger and so on the largest value is written in the end. In the descending order, the largest value is written in the first place, then next smaller and so on the smallest value is written in the end.

2. **Discrete series :** When the distinct values of a discrete variable or a continuous variable are presented in an ascending order or descending order along with their frequencies, the series is called *discrete series*. The number of items a value of a variable occurs is called its *frequency*. According to *Prof. Boddington*, "Discrete series is one where the individual values differ from each other by definite amounts." However, one may say, "Any series described by discrete variable is a discrete series." *For example,*

(i) Marks obtained by 23 students :

Marks (Variable *X*)	15	17	18	21
No. of Students (Frequency)	3	8	5	7

(ii) The weight (in kg) of 20 children.

Weight (Variable X)	2.5	2.8	3.3	3.4
No. of Children (Frequency)	3	7	2	8

Remark : Discrete series is also called Discrete frequency distribution.

3. **Grouped series :** When the observations or measurements are arranged in ascending or descending order in groups along with their frequencies (expressed in class-intervals i.e., with a certain limit), the arrangement is called a grouped series. If the continuity of the groups is not broken i.e., the point of which a group ends, the next begins, the arrangement is called a continuous series. For example, weight of 42 students in kg :

Weight	60-65	65-70	70-75	75-80
No. of Students (Frequency)	4	10	20	18

Remark : Grouped series is also called Grouped frequency distribution.

Q.11. Define various terms related with a grouped frequency distribution.

Ans. Following are the various terms related with a grouped frequency distribution. These specific terms used in classification according to class intervals are called its *terminology*.

1. **Class :** Various groups formed in a quantitative classification are called classes.
2. **Class-interval :** The various intervals (symbols for the class) are called class-intervals $(L_1 - L_2)$.

   Although class-interval is a symbol for a class, the terms 'class' and 'class-interval' are often used in place of each other.
3. **Class frequency :** The number of observations falling within a particular class or class-interval is known as class frequency or frequency of that class.
4. **Class-limits :** The pairs of the numbers that represent the class or class-interval are called limits. In a specified class the greatest value that a variable can take, is called the upper class-limit (L_2) and the least value it can take in that class is called the lower class-limit (L_1). However, the true class-limits are called class-boundaries and the observed limits are known as apparent class-limits.
5. **Magnitude of the class-interval or width of the class :** The difference between the two limits (upper class-limit and lower class-limit) of a class is called the length or magnitude or width of the class (or class-interval). Thus,

   $$\text{Width of the class, } i = L_2 - L_1$$

   In case of inclusive method, the width of the class is equal to the difference between the lower (or upper) class-limits of the two consecutive classes.
6. **Mid-value or Mid-point of a class :** The value being half way between the lower and upper class-limits of a class-interval is called its mid-value or central value or mid-point. Thus,

   $$\text{Mid point} = \frac{\text{Lower limit} + \text{Upper limit}}{2} \text{ or } \frac{L_1 + L_2}{2}$$

7. Various forms of class-intervals :

(i) **Exclusive classes (or class-intervals) :** When the upper limit of a class-interval is equal to the lower limit of the next higher class-interval, then the class-intervals are called exclusive class-intervals. *For example,* 0-10, 10—20, 20—30, ...

To avoid the confusion, the exclusive method (by convention) is used with the explanation that upper limit of a class is not included in that class (*i.e.,* excluded from that class) but it is included in the next higher class. Thus, the class-intervals 0-10, 10-20, 20-30, 30-40, 40-50 can be written as :

I	or	II
0 but less than 10		0-9.99
10 but less than 20		10-19.99
20 but less than 30		20-29.99
30 but less than 40		30-39.99
40 but less than 50		40-49.99

(ii) **Inclusive classes or class-intervals :** When both lower limit and upper limit of a class-interval are included in it, then it is known as inclusive class-interval. *For example,*

Wages (in ₹)	0-9	10-19	20-29	30-39	40-49

Note that here width of the class is 10 and not 9.

To change an inclusive class-interval into exclusive form, find the difference of lower limit of a class and upper limit of the previous class; add half of this difference in the upper limit and subtract half of this difference from the lower limit. Thus,

$$10-9=1, \quad \frac{1}{2} \text{ of } 1 = 0.5$$

Hence, classes in exclusive form are :

–0.5-9.5   9.5-19.5   19.5-29.5   29.5-39.5   39.5-49.5

Remark : If the variable under study cannot take negative values, then we may take the interval as 0-9.5, 9.5-19.5, 19.5-29.5, 29.5-39.5, 39.5-49.5 or the intervals may be taken as :

0-10   10-20   20-30   30-40   40-50

(iii) **Open end class-interval :** If the lower limit or upper limit of a class-interval is not mentioned, then it is called open end class-interval. *For example,* less than 5, more than 60.

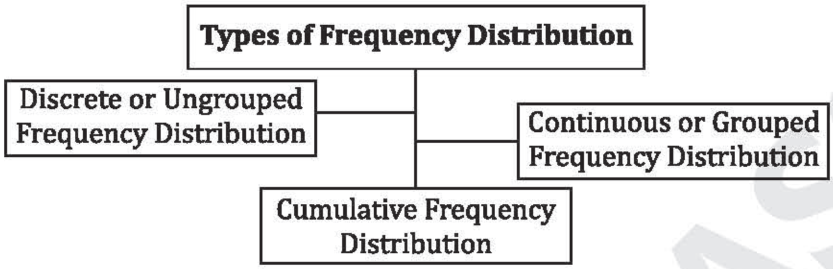
(iv) **Unequal class-intervals :** If in a frequency distribution the width of all the class-intervals are not equal, then the class-intervals are known as unequal class-intervals.

Q.12. What are the different types of frequency distribution?

Ans.

Types of Frequency Distribution

Basically there are three types of frequency distribution :



Discrete or Ungrouped Frequency Distribution

In this form of distribution, the frequency refers to discrete value. Here the data are presented in a way that exact measurement of units is clearly indicated. There is definite difference between the variables of different groups of items.

Each class is distinct and separate from the other class. Non-continuity from one class to another class exists. Data such as facts like the number of rooms in a house, the number of companies registered in a country, the number of children in a family, etc. The process of preparing this type of distribution is very simple. We have just to count the number of times a particular value is repeated, which is called the frequency of that class. In order to facilitate counting prepare a column for tally marks.

In another column, place all possible values of variable from the lowest to the highest. Then put a bar (Vertical line) opposite the particular value to which it relates. To facilitate counting, blocks of five bars are prepared and some space is left in between each block. We finally count the number of bars and get frequency.

1	0	3	2	1	5	6	2
2	1	0	3	4	2	1	6
3	2	1	5	3	3	2	4
2	2	3	0	2	1	4	5
3	3	4	4	1	2	4	5

Example : Represent the data in the form of a discrete frequency distribution.

Solution : Frequency distribution of the number of children :

Number of Children	Tally Marks	Frequency
0					3
1	𝍸 ||	7
2	𝍸 𝍸	10
3	𝍸 |||	8
4	𝍸		6
5						4
6				2
Total		**40**

Continuous or Grouped Frequency Distribution

Continuous series is one where measurements are only approximations and are expressed in class intervals, *i.e.,* within certain limits.

According to *Boddington,* "the variable which can take any intermediate value between the smallest and longest value in the distribution."

In a continuous frequency distribution the class intervals theoretically continue from the beginning of the frequency distribution to the end without break.

Cumulative Frequency Distribution

A cumulative distribution of frequencies shows the number of data items with values less than or equal to the upper class limit of each class. While a cumulative relative frequency distribution gives the proportion of the data items and a cumulative percentage frequency distribution shows the percentage of data items with values less than or equal to the upper class limit of each class.

Table : GCS Scores of Table Sharing the Cumulative Frequency

CGS Score	Frequency (Number of Patients)	Cumulative Frequency (Cumulative Number of Patients)
1	10	10
2	5	15
3	6	21
4	2	23
5	12	35
6	15	50
7	18	68
8	14	82
9	15	97
10	21	118
11	13	131
12	17	148
13	6	154

The cumulative frequency for each category tells us how many subjects there are in that category, and in all the lesser-valued categories in the table. *For example,* 35 patients had a GCS score of 5 or less.

❑

UNIT-II

Measures of Central Tendency

SECTION-A VERY SHORT ANSWER TYPE QUESTIONS

Q.1. Write the importance of measure of central tendency.

Ans. Averages occupy an important place in statistics. It represents a large group in such a way that mind can grasp simply and quickly. They are the bases of many other techniques of statistical analysis. This is the reason for **Dr. Bowley** defining statistics "as the science of average". Averages are widely used than any other statistical measure because of their applications and functions. Importance of averages in statistical analysis is obvious in the words of **Prof. R.A. Fisher**, "The inherent inability of the human mind to grasp in its entirety a large body of numerical data compels us to seek relatively few constants that will adequately describe the data."

Q.2. Write the various types of measures of central tendency.

Ans. The various averages are as given below :

(A) **Mathematical Average :** (1) Arithmetic Average or Arithmetic Mean, (2) Geometric Mean, (3) Harmonic Mean, (4) Quadratic Mean.

(B) **Average of Position :** (1) Median, (2) Partition Values, (3) Mode.

(C) **Commercial Averages :** (1) Moving Average, (2) Progressive Average, (3) Composite Average.

Q.3. What are the advanages of mode?

Ans. Advantages of Mode :

1. Easy to understand and calculate.
2. Can be easily found out by using inspection method.
3. It is an actual value, which most frequently occurs in the series.
4. Not affected by extreme values.
5. It is simple and accurate and can be measured in an open end class-interval without determining the class limits.

Q.4. Calculate mode from the following data of income earned by 10 employees :

S.No.	Income (in ₹)	S.No.	Income (in ₹)
1	10	6	27
2	27	7	20
3	24	8	18
4	18	9	15
5	27	10	32

Sol. Mode can be calculated as follows :

Income (in ₹)	Number of Employees
10	1
15	1
18	2
20	1
24	1
27	3
32	1
Total	**10**

Since the item 27 occurs the maximum number of times, *i.e.*, 3, hence the modal income is 27.

Q.5. Write the relation among mean, median and mode.

Ans. The empirical relation among mean, median and mode is :

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}$$

Q.6. Calculate the Median when Mean and Mode of Distribution are 38.6 and 32.6 respectively.

Sol. Given, Mean $= 38.6$ and Mode $= 32.6$

$$\text{Median} = \text{Mode} + \frac{2}{3}[\text{Mean} - \text{Mode}] = 32.6 + \frac{2}{3}[38.6 - 32.6] = 32.6 + \frac{2}{3}[6] = 32.6 + 4$$

$\Rightarrow$ Median $= 36.6$

Q.7. For a given set of values of a variable, the mean and mode are 28.16 and 24 respectively, calculate the median.

Sol. Mean $= 28.16$, Mode $= 24$

$$\text{Median} = \text{Mode} + \frac{2}{3}[\text{Mean} - \text{Mode}] = 24 + \frac{2}{3}[28.16 - 24) = 24 + 2.77 = 26.77$$

$\Rightarrow$ Median $= 26.77$

Q.8. Write any two differences among mean, median and mode.

Ans. Difference among Mean, Median and Mode is a follows :

S. No.	Mean	Median	Mode
1.	It is highly affected by extreme values.	It is not affected by extreme values as in case of mean.	It is also not affected by extreme values as in case of mean.
2.	It is based on all the observations.	It is not based on all observations.	It is also not based on all observations.

Q.9. What are the disadvantages of mean deviation?

Ans. Disadvantages of Mean Deviation :

1. For open ended classes, it is not possible to calculate mean deviation.

2. Mean deviation tends to increase with the size of the sample, though not proportionately and not as rapidly as range.
3. In sociological studies, it is rarely used.
4. The ignorance of sign is mathematically unsound and illogical.

Q.10. Define coefficient of standard deviation.

Ans. Coefficient of Standard Deviation : The standard deviation is the absolute measure of dispersion. Its relative measure is called the standard coefficient of dispersion or coefficient of standard deviation. It is defined as :

$$\text{Coefficient of Standard deviation} = \frac{\sigma}{\bar{X}}$$

Q.11. From the information given below, calculate Karl Pearson's Coefficient of Skewness :

Measure	Place *A*	Place *B*
Mean	160	150
Median	142	145
Standard deviation	28	52
Third quartile	198	255
First quartile	68	78

Sol. Place $A : J = \frac{3(\bar{X} - M)}{\sigma} = \frac{3(160 - 142)}{28} = \frac{3 \times 18}{28} = 1.93$

Place $B : J = \frac{3(\bar{X} - M)}{\sigma} = \frac{3(150 - 145)}{52} = \frac{15}{52} = 0.3$

Q.12. Calculate Geometric Mean from following distribution :

X	0-15	15-30	30-45	45-60	60-75
f	8	15	20	4	3

Sol.

X	*f*	Mid value (*m*)	log *m*	*f* log *m*
0-15	8	7.5	0.8750613	7.0004901
15-30	15	22.5	1.3521825	20.282738
30-45	20	37.5	1.5740313	31.480625
45-60	4	52.5	1.7201593	6.8806372
60-75	3	67.5	1.8293038	5.4879113
	50			**$\Sigma f \log m = 71.132402$**

$$\text{G.M.} = \text{Antilog} = \frac{\Sigma f \log m}{\Sigma f} = \text{Antilog}\ \frac{71.132402}{50} = \text{Antilog}\ 1.422648 = 26.4635$$

Q.13. What are the differences between skewness and dispersion?

Ans. The main difference between Skewness and Dispersion are given as follows :

S. No.	Skewness	Dispersion
1.	Skewness gives the idea about the direction of variation.	Dispersion gives the idea about the amount of the variation.
2.	The tendency of the variation of data points into a certain direction can be understood with the use of skewness.	Dispersion is used to know the range of the data points and offset from the mean.
3.	Skewness gives an idea about the shape of the series.	Dispersion gives an idea about the composition of the series.

Q.14. Calculate range and its coefficient from the following data :

53, 46, 18, 16, 75, 84, 28

Sol. Range $R = L - S = 84 - 16 = 68$

Coefficient of range $= \frac{L-S}{L+S} = \frac{84-16}{84+16} = \frac{68}{100} = 0.68$

Q.15. Write the meaning and definition of dispersion.

Ans.

Meaning and Definition of Dispersion

Meaning : The word 'dispersion' is used in two senses in Statistics :

(i) Dispersion means the scatteredness of the values of a variable due to variation among themselves.

(ii) Dispersion means the scatteredness of the deviations from a measure of central tendency or any other fixed value.

Definition : Some definitions of dispersion are as follows :

1. According to *Dr. Bowley*, "Dispersion is the measure of variation of the items."
2. According to *Connor*, "Dispersion is a measure of the extent to which the individual vary."
3. In the words of *Brookes and Dick*, "The dispersion or spread is the degree of a scatter or of variation of the variables about a central value."
4. According to *Spriegel*, "The degree to which numerical data tend to spread about an average value is called the variation or dispersion of data."

SECTION-B SHORT ANSWER TYPE QUESTIONS

Q.1. Write advantages and disadvantages of Mean.

Ans. Advantages and disadvantages of mean are as follow :

Advantages of Mean

1. Simple to calculate and understand.
2. Some value is always determined, *i.e.,* it is never indefinite.
3. Can be used in other algebraic calculations.
4. No need of sorting or arrangement (ascending or descending order).
5. It is stable and not affected by the variation of sampling.

Disadvantages of Mean

1. Mean is greatly affected by the extreme values. *For example*, mean of 3, 7 and 200 is 70. Here, no value is present near this 70; hence, this average is of no use.
2. Sometimes mean may provide confusing impressions. *For example*, in a hospital, per day average number of patients admitted is 5.7.

   Here given information is useful but doesn't provide the actual item because some values are of no use when expressed in fraction or decimal.
3. The 'mean' cannot be predicted by just inspecting the sample item.
4. If a single value is missing, mean cannot be calculated.
5. In case of open-end classes, mean cannot be calculated.
6. Graphical representation of mean is not possible.

Q.2. Mean of 12 values is 128. While calculating this mean, one value which actually was 101 was wrongly read as 110. Find the actual mean of the values.

Sol. Given, $N = 12$, Mean (wrong) = 128

Let us assume that sum of 11 values is X.

Then
$$128 = \frac{X + 110}{12}$$
$$1536 = X + 110 \quad \text{or} \quad X = 1426$$

Now we use the actual value then find the actual mean $= \frac{1426 + 101}{12} = \frac{1527}{12} = 127.25.$

Q.3. The following table gives the heights of 350 men. Calculate the mean height of the group :

Height (in cm)	159	161	163	165	167	169	171	173
No. of Persons	1	2	9	48	131	102	40	17

Sol. **Calculation of Mean Height**

Height (X)	**Frequency (f) (No. of Persons)**	**fX**
159	1	159
161	2	322
163	9	1467
165	48	7920
167	131	21877
169	102	17238
171	40	6840
173	17	2941
Total	**$\Sigma f = 350$**	**$\Sigma fX = 58764$**

$\therefore$ Arithmetic mean (or mean height) $= \frac{\Sigma fx}{\Sigma f} = \frac{58764}{350} = 167.90$

Q.4. Calculate the geometric mean from the following data :

X :	125	133	141	173	182

Sol.

Calculation of Geometric Mean

X	log of *X*
125	2.09691
133	2.123852
141	2.149219
173	2.238046
182	2.260071
	$\Sigma \log X = 10.8681$

$$\text{G.M.} = \text{Antilog}\frac{\Sigma \log X}{N} = \text{Antilog}\frac{10.8681}{5} = \text{Antilog } 2.17362 = 149.15$$

Q.5. Calculate geometric mean from the following distribution :

Income (₹)	5	20	35	50	60	75	90	10	125
Number of Persons	15	16	18	20	13	9	8	7	6

Sol.

Calculation of Geometric Mean

Income (*X*)	Number of Persons (*f*)	log *X*	*f* log *X*
5	15	0.6989	10.4835
20	16	1.3010	20.816
35	18	1.5441	27.7938
50	20	1.6989	33.978
60	13	1.7781	23.1153
75	9	1.8751	16.8759
90	8	1.9542	15.6336
10	7	1.0000	7
125	6	2.0969	12.5814
	$\Sigma f = 112$		$\Sigma f \log X = 168.2775$

$$\text{G.M.} = \text{Antilog}\frac{\Sigma f \log X}{\Sigma f} = \text{Antilog}\frac{168.2775}{112} = \text{Antilog of } 1.5025 = ₹31.8$$

Q.6. Calculate the harmonic mean of the following data.

15, 250, 15.7, 157, 1.57, 105.7, 10.5, 1.06, 25.7, 0.257

Sol.

Calculation of Harmonic Mean

(*X*)	Reciprocals (1/*X*)
15	0.066667
250	0.004

15.7	0.063694
157	0.006369
1.57	0.636943
105.7	0.009461
10.5	0.095238
1.06	0.943396
25.7	0.038911
0.257	3.891051
	$\Sigma 1/X = 5.755729$

$$\text{H.M.} = \frac{N}{\left(\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots+\frac{1}{X_N}\right)} = \frac{N}{\Sigma\frac{1}{X}} = \frac{10}{5.755729} = 1.737$$

Q.7. Calculate H.M. of the following data :

Marks	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency	14	12	9	7	13	8	9

Sol.

Computation of H.M.

Marks	**Mid Value (m)**	**Frequency (f)**	**Reciprocal ($1/m$)**	**$f \times$ (reciprocal) ($f(1/m)$)**
20-30	25	14	0.0400	0.56
30-40	35	12	0.0285	0.342
40-50	45	9	0.0222	0.1998
50-60	55	7	0.0181	0.1267
60-70	65	13	0.0153	0.1989
70-80	75	8	0.0133	0.1064
80-90	85	9	0.0117	0.1053
		$\Sigma f = 72$		**$\Sigma f(1/m) = 1.6391$**

$$\text{H.M.} = \frac{N}{\Sigma f\frac{1}{m}} = \frac{72}{1.6391} = 43.92$$

Q.8. Using the values 20, 40, 80, verify that A.M. > H.M.

Sol. Given, $X_1 = 20, X_2 = 40$ and $X_3 = 80$

$$\text{Arithmetic Mean (A.M.)} = \frac{20+40+80}{3} = 46.67$$

$$\text{H.M.} = \frac{N}{\left(\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots+\frac{1}{X_N}\right)} = \frac{3}{\frac{1}{20}+\frac{1}{40}+\frac{1}{80}}$$

$$= \frac{3}{\left(\frac{4+2+1}{80}\right)} = \frac{3 \times 80}{7} = 34.3$$

Hence, A.M. > H.M.

Q.9. Calculate the median from the following data :

Marks (less than)	80	70	60	50	40	30	20	10
No. of Students	100	90	80	70	60	32	20	5

Ans. First we convert the cumulative series into simple series :

C.I.	(f)	Cumulative Frequency (c)
0-10	5	5
10-20	8	13
20-30	7	20
30-40	12	32
40-50	28	60
50-60	20	80
60-70	10	90
70-80	10	100
	$N = 100$	

Median number, $m = \frac{N}{2} = \frac{100}{2} = 50$. This lies in the class 40-50.

Compulative frequency of the class preceding the median class $c = 32$.

Cumulative frequency of median class $= 60$ and $f = 28$

$$M = L_1 + \frac{m-c}{f} \times (L_2 - L_1) = 40 + \frac{50-32}{28}(50-40) = 40 + \frac{18}{28} \times 10 = 46.43$$

So, required median is 46.43 and median class is (40-50).

Q.10. Find the missing frequency in the group 20-30 when the median is given to be 28.

X	0-10	10-20	20-30	30-40	40-50
f	5	8	?	16	6

Sol. Let the missing frequency is α.

X	f	(c.f.)
0-10	5	5
10-20	8	13
20-30	α	$13 + \alpha$
30-40	16	$29 + \alpha$
40-50	6	$35 + \alpha$

$$\text{Median number} = \frac{N}{2} = \frac{35+\alpha}{2} = \left(17.5 + \frac{\alpha}{2}\right)\text{th item median class is : 20-30.}$$

$\therefore\ c = 13, f = \alpha, i = 10$

$$\text{Median} = L_1 = \frac{\frac{N}{2} - C}{f} \times i = 20 + \frac{\frac{35+\alpha}{2} - 13}{\alpha} \times 10 = 20 + \frac{(5\alpha + 45)}{\alpha}$$

$$3\alpha = 45 \Rightarrow \alpha = 15$$

Q.11. Calculate the mode from the following data :

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Students	6	8	7	12	26	20	11	10

Sol. Here, the class-interval (40-50) has the maximum frequency, *i.e.*, 26, therefore, modal class is (40-50). Now we can calculate the modal value of using the formula,

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

Again we have $L_1 = 40, L_2 = 50, f_0 = 26, f_1 = 12, f_2 = 20$ and $i = 10$;

$\therefore$
$$\text{Mode} = 40 + \frac{26-12}{2 \times 26 - 12 - 20} \times 10 = 40 + \frac{140}{20} = 47$$

Q.12. Find mode for the following data :

20, 15, 18, 20, 25, 15, 16, 15, 30, 12

Sol.

Calculation of Mode

Size of Item	Number of Times it Occurs
12	1
15	3
16	1
18	1
20	2
25	1
30	1
Total	**10**

Since the item 15 occurs the maximum number of times, *i.e.*, 3, hence, the mode value is 15.

Q.13. Write the definition, merits, demerits and uses of Geometric mean.

Ans.

Geometric Mean

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

For example, (i) geometric mean of 4 and 16 is

$$\text{G.M.} = \sqrt[2]{4 \times 16} = \sqrt[2]{64} = 8$$

(ii) Geometric mean of 2, 6 and 18 is

$$\text{G.M.} = \sqrt[3]{2\times6\times18}$$
$$= \sqrt[3]{2\times2\times3\times2\times3\times3} = 2\times3 = 6$$

Merits, Demerits (Limitations) and Uses of Geometric Mean

Merits

1. It is rigidly defined.
2. It is based on all the observations.
3. It is suitable for further mathematical treatment. For example, if G_1, G_2 are the geometric means of two groups with N_1 and N_2 observations respectively then the combined geometric mean of $(N_1 + N_2)$ observations is

$$G = \text{Antilog}\left(\frac{N_1 \log G_1 + N_2 \log G_2}{N_1 + N_2}\right)$$

4. It is not much affected by fluctuations of sampling.
5. It gives comparatively more weight to small values.

Demerits (Limitations)

1. It is comparatively difficult to calculate and not simple to understand.
2. If any observation is zero, then geometric mean becomes zero.
3. It is not defined for negative values.
4. It cannot be obtained by inspection.
5. It may not be represented in the actual data.

Uses

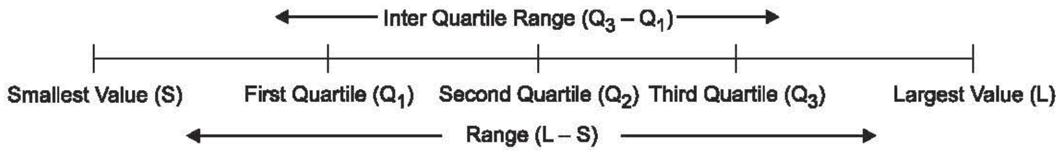
1. Geometric mean is appropriate :
   (i) When large observations are to be given less weight.
   (ii) When we find the relative changes, such as the average rate of population growth, the average rate of interest, the average rate of depreciation in machines, etc.
   (iii) Where some of the observations are too small and/or too large.
2. Geometric mean is also used in the construction of Index Numbers.

Q.14. What do you mean by Quartile Deviation? Explain it.

Ans.

Quartile Deviation

Quartile deviation is another measure of variation which gives the solution to overcome the limitation of range. In a data set, it calculates the spread over the middle half of the values. This measure of variation minimises the effect of extreme values (also known as outliers). In this method, the study of **Interquartile range** is necessary because a large amount of values in the data set lie in the central part of the frequency distribution. To calculate this value, all the data set is divided into four parts. Every part of the data set contains 25% of the observed value. In these values the highest one is known as quartile.



The half of the difference between third quartile and first quartile $(Q_3 - Q_1)$ is known as **'semi-interquartile** or **quartile deviation'**.

$$\text{Mathematically, Quartile Deviation, Q.D.} = \frac{Q_3 - Q_1}{2}$$

Where, $(Q_3 - Q_1)$ = Interquartile range.

Coefficient of Quartile Deviation

If the difference of the third and first quartiles is divided by the sum of the third and first quartiles then it is known as the **'coefficient of quartile deviation'**. In case of open-ended distribution (when the frequency distribution has the extreme class limits and no specific class limits), it is a very useful measure. Mathematically, coefficient of quartile deviation is defined as follows :

$$\text{Coefficient of Quartile Deviation (CQD)} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Where, CQD = Coefficient of Quartile Deviation; Q_3 = Third quartiloe; Q_1 = First quartile.

Q.15. Define variance and sample variance and also write the properties of variance.

Ans.

Variance

Absolute values are not conducive to easy manipulation; due to this reason mathematicians developed an alternative mechanism for overcoming the zero sum property of deviations from the mean. This approach utilizes the square of the deviations from the mean. The result is the variance, an important measure of variability. The variance is the average of the squared deviations about the arithmetic mean for a set of numbers. The population variance is denoted by σ^2.

Sample Variance

The sample variance is denoted by s^2. The main use for sample variances is as estimator of population variances. Because of this, computation of the sample variance differs slightly from computation of the population variance. The sample variance uses $N-1$ in the denominator instead of N because using n in the denominator of a sample variance results in a statistic that tends to underestimate the population variance. Using $N-1$ allows it to be an unbiased estimator, which is a desirable property in inferential statistics.

Sample variance, $$s^2 = \frac{\Sigma(X - \bar{X})^2}{N - 1}$$

For example, sample of six of the largest accounting firms in the United States and the number of partners associated with each firm as reported by the Public Accounting Report.

Firm	Number of Partners
Price Waterhouse	1062
McGladrey & Pullen	381
Deloitte & Touche	1719
Andersen Worldwide	1673
Coppers & Lybrand	1277
BDO Seidman	217

The sample variance and sample standard deviation can be computed as :

X	$(X-\bar{X})^2$
1062	51.41
381	454,046.87
1719	441,121.79
1673	382,134.15
1277	49,359.51
217	701,959.11
$X = 6329$	$\Sigma(X-\bar{X})^2 = 2{,}028{,}672.84$

$X = 6329$; $\Sigma(X-\bar{X})^2 = 2{,}028{,}672.84$

$$\bar{X} = \frac{6329}{6} = 1054.83$$

$$s^2 = \frac{\Sigma(X-\bar{X})^2}{N-1} = \frac{2{,}028{,}672.84}{5} = 405{,}734.57$$

The sample variance is 405,734.57.

Properties of Variance

Properties of variance are given as follows :

Property 1. If there is X and Y be the two independent random variables. Then,

$$V(X+Y) = V(X) + V(Y)$$

Property 2. In a population function if a constant 'c' is added the old and new variance are the same.

Property 3. If a constant is multiplied by the data item of a population function then the new variance is c^2 times that the old variance.

$$V(cX) = c^2 V(X)$$

Property 3. If $X_1, X_2, \ldots\ldots, X_n$ are independent trails process with $E(X_j) = \mu$ and $V(X_j) = \sigma^2$.

Let, $S_n = X_1 + X_2 + \ldots + X_n$ be the sum, and $A_n = \frac{S_n}{n}$ be the average.

$$E_{(S_n)} = n\mu, V(S_n) = n\sigma^2$$

Then,
$$E(A_n) = \mu, V(A_n) = \frac{\sigma^2}{n}$$

Q.16. What do you mean by coefficient of variation (CV)?

Ans. **Coefficient of Variation (CV)**

It is an absolute measure of dispersion. Coefficient of variation is expressed in terms of units in which actual data is collected and stated. The standard deviation of the weights of people cannot be compared with the standard deviation of heights of the people because both are given in two different units *i.e.,* weights are given in kilogram and heights are given in metres. For the purpose of comparison, standard deviation is definitely converted into relative measure (known as coefficient of variation) of dispersion. The ratio of the standard deviation to the mean in percentage is known as 'coefficient of variation'.

Mathematically,

$$\text{Coefficient of variance (C.V.)} = \frac{\sigma}{\overline{X}} \times 100$$

Example : The average daily income of 200 workers in a factory is ₹142 and their variance is 36. Find the total income of all the workers. What is the coefficient of variation in their income? Compute.

Sol. We know $\overline{X} = \frac{\Sigma X}{N}$

Here, $\overline{X} = 142$ and $N = 200 \Rightarrow 142 = \frac{\Sigma X}{200} \Rightarrow \Sigma X = 28{,}400.$ Hence, total income of all the workers is 28,400.

Now, coefficient of variance $(CV) = \frac{\sigma}{\overline{X}} \times 100$ and variance $V = \sigma^2 \Rightarrow \sigma = \sqrt{V}$.

Here, variance $= 36 \Rightarrow \sigma = \sqrt{36} = 6.$ Now, $CV = \frac{6}{142} \times 100 = 4.225$

So, coefficient of variance in their income is 4.225.

Q.17. Calculate Coefficient of Variation (C.V.) for the following data :
56, 66, 61, 68, 54, 70, 55.

Sol. Let the assumed mean is $(A) = 60$.

Calculation of Coefficient of Variation

X	$d_x = (X - A)$	(d_x^2)
56	–4	16
66	6	36
61	1	1
68	8	64
54	–6	36
70	10	100
55	–5	25
	$\Sigma d_x = 10$	$\Sigma d_x^2 = 278$

$$\overline{X} = A + \frac{\Sigma dx}{N} = 60 + \frac{(10)}{7} = 60 + 1.43 = 61.43$$

$$\text{S.D.}(\sigma)=\sqrt{\frac{\Sigma(dx)^2}{N}-\left(\frac{\Sigma dx}{N}\right)^2}=\sqrt{\frac{278}{7}-\left(\frac{10}{7}\right)^2}$$

$$=\sqrt{\frac{278}{7}-\frac{100}{49}}=\sqrt{\frac{278\times 7-100}{49}}=\sqrt{\frac{1846}{49}}=\sqrt{37.67}=6.14$$

$$\text{Coefficient of Variation}=\frac{6}{\bar{X}}\times 100=\frac{6.14}{61.43}\times 100=\frac{614}{61.43}=9.99=10$$

Q.18. Calculate coefficient of variation of the following data :

Weekly Rent (₹)	400	700	800	950	1000	1200	1450
No. of Persons Paying the Rent	11	13	34	39	18	8	2

Sol.

Calculation of Coefficient of Variation

Weekly Rent (X)	No. of Persons (f)	fX	$d_x=(X-\bar{X})$	$f(d_x)^2$
400	11	4400	– 466	2388716
700	13	9100	– 166	358228
800	34	27200	– 66	148104
950	39	37050	84	275184
1000	18	18000	134	323208
1200	8	9600	334	892448
1450	2	2900	584	682112
	$\Sigma f=125$	$\Sigma fX=108250$		$\Sigma fd_x^2=5068000$

$$\bar{X}=\frac{\Sigma fX}{\Sigma f}=\frac{108250}{125}=866$$

$$\text{Standard Deviation}(\sigma)=\sqrt{\frac{\Sigma f(d_x)^2}{\Sigma f}}=\sqrt{\frac{5068000}{125}}=\sqrt{40544}=201.3$$

$$\text{Coefficient of Variation}=\frac{\sigma}{\bar{X}}\times 100=\frac{201.36}{866}\times 100=23.24$$

Thus, $\text{CV}=23.24\%$

Q.19. Define tests of skewness. Write its importance also.

Ans.

Tests of Skewness

Following are the tests which are applied to find that a distribution is skewed or not :

1. If a distribution is skewed then the value of mean, median and mode would not coincide. The value of median generally lies between the mean and the mode. In a moderately asymmetrical distribution,

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

$$\Rightarrow \quad 3\text{ Median} = 3\text{ Mean} + \text{Mode} - \text{Mean} = 2\text{ Mean} + \text{Mode}$$

$$= 2\text{ Mean} + \text{Mode} + 2\text{ Mode} - 2\text{ Mode}$$

$$= 3 \text{ Mode} + 2 \text{ (Mean} - \text{Mode)}$$

$$\Rightarrow \quad \text{Median} = \text{Mode} + \frac{2}{3}(\text{Mean} - \text{Mode})$$

2. If a distribution is skewed then the two quartiles would not be equi-distant from the median. In other words, it can be said that $(Q_3 - M) - (M - Q_1) \neq 0$.
3. If a distribution is skewed then its graph would not give a symmetrical bell-shaped curve.
4. The sum of positive deviations and the sum of negative deviations from the median would not be equal.
5. At various points, frequencies are not equally distributed which are equidistant from the mode. In an asymmetrical distribution (Mean – Median) = 3 (Mean – Mode).

Importance/Advantages of Skewness

1. If a given distribution is normal then the skewness would be zero and it is called symmetric distribution. But generally, data points are not perfectly symmetric.
2. With the help of skewness, we know that the deviation from the mean is whether positive or negative.
3. D'Agostino's K-squared test is a goodness-of-fit normality test based on sample skewness and sample kurtosis.

Q.20. What do you mean by coefficient of skewness?

Ans.

Coefficient of Skewness

Coefficient of skewness or Measure of skewness may be subjective or relative. In a frequency distribution, absolute measures tell us about the direction and extent of asymmetry. Relative measures, generally called coefficient of skewness, provides the facility to compare two or more frequency distributions. We can find the skewness by two methods :

Coefficient/Measure of Skewness

- Karl Pearson's Coefficient of Skewness
- Bowley's Coefficient of Skewness

Karl Pearson's Coefficient of Skewness

1. When Measure Dependes upon Difference of Mean and Mode.

   Absolute Measure, $J = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}$ or $\frac{\overline{X} - z}{\sigma}$

   Where, $\overline{X}$ = Arithmetic mean; Z = Mode; and σ = standard deviation.

2. If mode is not determined, then

   Absolute Measure, Sk = 3 (mean – Median) or $3(\overline{X} - M)$

   Relative measure, coefficient $J = \frac{3\text{ (Mean} - \text{Median)}}{\text{Standard Deviation}}$ or $\frac{3(\overline{X} - M)}{\sigma}$

   Where, M = Median.

Note : Mathematically there are no limits for J, put practically.

$$J = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}} \text{ takes values between } \pm 1;$$

$$J = \frac{3\,(\text{Mean} - \text{Median})}{\text{Standard Deviation}} \text{ takes values between } \pm 3.$$

Q.21. From the following data find out the Karl Pearson's Coefficient of Skewness :

Measurement	20	21	22	23	24	25
Frequency	1	3	8	11	6	1

Sol. Let the assumed mean $(A) = 22$

Calculation of Karl Pearson's Coefficient of Skewness

Measurement (X)	**Frequency (f)**	$d_x = (X - 22)$	fd_x	d_x^2	fd_x^2
20	1	– 2	– 2	4	4
21	3	– 1	– 3	1	3
22	8	0	0	0	0
23	11	+ 1	+ 11	1	11
24	6	+ 2	+ 12	4	24
25	1	+ 3	+ 3	9	9
	$N = 30$		$\Sigma fd_x = +21$		$\Sigma f(d_x)^2 = 51$

Arithmetic Mean, $(\bar{X}) = A + \frac{\Sigma fd_x}{N} = 22 + \frac{21}{30} = 22.7$

Mode (by inspection) $(Z) = 23$

Standard Deviation $\sigma = \sqrt{\frac{\Sigma f(d_x)^2}{N} - \left(\frac{\Sigma fd_x}{N}\right)^2} = \sqrt{\frac{51}{30} - \left(\frac{21}{30}\right)^2}$

$$= \sqrt{1.7 - 0.49} = \sqrt{1.21} = 1.1$$

Absolute Measure, $\text{Sk} = \bar{X} - Z = 22.7 - 23 = -0.3$

Karl Perarson's Coefficient of Skewness,

$$(J) = \frac{\bar{X} - Z}{\sigma} = \frac{22.7 - 23}{1.1} = -\frac{0.3}{1.1} = -0.273$$

SECTION-C LONG ANSWER TYPE QUESTIONS

Q.1. What do you mean by measures of central tendency? Describe its properties, various types and importance.

Ans.

Meaning and Definition of Central Tendency

Measure of Central tendency is also known as 'measure of central value' or 'measure of location' or 'average of first order'. It is a statistical measure and calculates the location or position of central point to explain the central tendency of the whole quantity of data.

According to *Simpson and Kafka*, "A measure of location is a typical value around which other figures congregate."

According to *R.F. Fesher*, "The inherent inability of the human mind to grasp in its entirely a large body of numerical data, compels us to seek relatively few constants that will describe the data."

According to *Bowley*, "Statistical constants enable us to comprehend in a single effort the significance of the whole."

According to *M.R. Speigal*, "Average is a value which is typical or representative of a set of data."

According to *Clark and Sekkade*, "Average is an attempt to find one single figure to describe whole of figure."

Properties of Good Measures of Central Tendency

1. It should be rigidly defined.
2. Its definition should be in the form of a mathematical formula.
3. It should be easy to understand and calculate.
4. It should remain unaffected by the extreme values.
5. It should be capable of further algebraic treatment.
6. It should be based on all items in the series.
7. It should be capable of being used in further statistical computation or processing.
8. It should possess sampling stability.

Types of Average/Measures of Central Tendency

In statistics, there are various types of measures of central tendency. Some of which can be broadly classified as follows :

I. Mathematical Average.

II. Positional Average.

I. Mathematical Average

Mathematical averages are those averages which are calculated by taking into account the values of all the items in a series. When values of all the items in a series are considered in the calculation of average then it is called mathematical averages.

1. **Mean :** Mean is also known as 'Arithmetic Mean' (A.M.). To calculate the mean, summate all the observations and divide it by the total number of observations.

   Mean is denoted by $\overline{X}$.

So, $\bar{X} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$

2. **Geometric Mean :** The geometric mean of 'N' numbers is defined as the "N^{th} root of the product of 'N' numbers." If all the values of a series are multiplied and then N^{th} root of the product is extracted, then the geometric mean is obtained.

   Symbolically : G.M. $= \sqrt[N]{(X_1) \times (X_2) \times (X_3) \ldots \times (X_n)}$

3. **Harmonic Mean :** The harmonic mean is the reciprocals of averaged numbers. It is defined as the reciprocal of the arithmetic mean of the reciprocal of the individual observations. Under certain conditions, harmonic mean is a better measure of central tendency e.g., computation of average speed, average price, etc.

   Mathematically, $$\text{H.M.} = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + \ldots + \frac{1}{X_N}}$$

   Where, N = Number of observations; X_N = Value of observation.

II. Positional Average

Positional average is the average which depends on the position of the items, rather than the values of the items.

1. **Median :** When N observations are grouped and arranged in the sorting order (ascending or descending order) according to their values, then the central value of the observation is known as median. It is denoted by M or Me.
2. **Mode :** Mode is the value of the variable for which the frequency is maximum and it is denoted by Z or Mo.
3. **Quartiles :** The measure of central tendency which divides a group of data into four subgroups or parts, then it is called quartile. The three quartiles are denoted as Q_1, Q_2 and Q_3.
4. **Deciles :** If the values of the observation in a data set are arranged in ascending or descending order and divided it into ten equal parts by using nine points on the scale of observations, then it is called decile.
5. **Percentiles :** If the values of the observation in a data set are arranged in ascending or descending order and divided it into hundred equal parts by using ninety nine points on the scale of observations, then it is called percentile. It is represented by $P_1, P_2, \ldots, P_{99}$.

Importance of Central Tendency

1. **To Find Representative Value :** Measures of central tendency or averages give us one value for the distribution and this value represents the entire distribution. In this way averages convert a group of figures into one value.
2. **To Condense Data :** Collected and classified figures are vast. To condense these figures we use average. Average converts the whole set of figures into just one figure and thus helps in condensation.
3. **To Make Comparisons :** To make comparisons of two or more than two distributions, we have to find the representative values of these distributions. These representative values are found with the help of measures of the central tendency.

4. **Helpful in Further Statistical Analysis :** Many techniques of statistical analysis like Measures of Dispersion, Measures of Skewness, Measures of Correlation, and Index Numbers are based on measures of central tendency. That is why, measures of central tendency are also called as measures of the first order.

Q.2. Describe the methods of calculation of Arithmetic mean in individual series with examples.

Ans.

Calculation of Arithmetic Mean : Individual Series

1. **Direct Method :** The direct method of calculating arithmetic mean in the case of individual series involves the following steps :
   (i) Add the various given values of the variable X and find the total value which is denoted by ΣX. After that find the total No. of observation.
   (ii) Divide this total value (ΣX) by the number of observations (N) *i.e.*,

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{N} = \frac{\Sigma X}{N}$$

Where, $\bar{X}$ = mean, ΣX total of the observation, and N = Number of all terms.

Example : The weekly wage of 5 workers is as given below :

₹1,350, ₹1,400, ₹1,450, ₹1,370 and ₹1,480

Find arithmetic mean.

Sol.

Computation of Arithmetic Mean

Serial Number	Weekly Wages (in ₹)
1	1,350
2	1,400
3	1,450
4	1,370
5	1,480
$N = 5$	$\Sigma X = 7{,}050$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{7{,}050}{5} = ₹1{,}410$$

Example : Compute Arithmetic mean from the following data :

Serial Number	1	2	3	4	5
Marks of Students	8	10	20	15	7

Sol.

Computation of Arithmetic Mean

Serial Number	Marks of Students
1	8
2	10
3	20
4	15
5	7
$N = 5$	$\Sigma X = 60$

Arithmetic mean $\overline{X} = \frac{\Sigma X}{N} = \frac{60}{5} = 12$

2. **Short-Cut Method :** With the use of short cut method also, arithmetic mean can be calculated. The calculation of arithmetic mean by short cut method includes the following steps :
   (i) Any value may be taken as an assumed mean of the data. This assumed value is also known as working mean or arbitrary average (A = Assumed mean).
   (ii) Subtract assumed mean from each value of the observation ($d = X - A$).
   (iii) All the deviations are added and it is denoted by (Σd).
   (iv) Apply the formula :

$$\overline{X} = A + \frac{\Sigma d}{N}$$

Where, $\overline{X}$ = Arithmetic mean; A = Assumed mean, Σd = Sum of the derivation, N = Number of items.

Example : Calculate mean by shortcut method from the following data :

Years	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Price of Rice (in ₹)	40	50	55	80	58	60	75	35	45	52

Sol. Let the assumed mean be $A = 50$

Year	**Price of Rice (X)**	$d = (X - 50)$
2001	40	– 10
2002	50	0
2003	55	5
2004	80	30
2005	58	8
2006	60	10
2007	75	25
2008	35	– 15
2009	45	– 5
2010	52	2
$N = 10$		$\Sigma d = 50$

$$X = A + \frac{\Sigma d}{N} = 50 + \frac{50}{10} = 50 + 5 = ₹55.$$

Q.3. Describe the methods of calculation of arithmetic mean in discrete series with examples.

Ans.

Calculation of Arithmetic Mean : Discrete Series (Ungrouped Data)

In discrete series, arithmetic mean can be calculated by the following methods :

1. **Direct Method :** In the discrete series, the sum of items is determined by multiplying each value with the respective frequency. The values got after multiplication are

totalled up. To find the arithmetic mean, total value is divided by the total number of items.

Following are the steps which are involved in the calculation of mean :

(i) Variable Denoted by X and Frequency denoted by f.

(ii) Multiply each item by its frequency, denoted by (fX).

(iii) Add all the fX, denoted by (ΣfX).

(iv) ΣfX is divided by the total number of items.

$$\text{The formula is } \bar{X} = \frac{\Sigma fX}{\Sigma f}$$

Where, $\bar{X}$ = Arithmetic mean; ΣfX = The sum of products; Σf = Total number of items.

Example : Find average wages of 10 workers

Daily Wage (in ₹)	4	6	10	11	14	Total
No. of Workers	2	1	4	2	1	10

Sol.

Calculation of Arithmetic Mean

Daily Wage (X)	No. of Workers (f)	fX
4	2	8
6	1	6
10	4	40
11	2	22
14	1	14
Total	$\Sigma f = 10$	$\Sigma fX = 90$

$\therefore$ Arithmetic Mean (Average Wage) $= \frac{\Sigma fX}{\Sigma f} = \frac{90}{10} =$ ₹9.00

2. **Short-Cut Method :** For calculating the arithmetic mean by short-cut method in a discrete series the following steps are taken :

(i) Any value may be taken as an assumed mean of the data.

(ii) Subtract assumed mean from each value of the observation ($d_x = X - A$).

(iii) Each value of deviation is multiplied by its respective frequency, denoted by (fd_x).

(iv) Following is the formula which is used for the calculation of arithmetic mean :

$$\bar{X} = A + \frac{\Sigma fd_x}{\Sigma f} \quad \text{or} \quad A + \frac{\Sigma fd_x}{\Sigma N}$$

Where, $\bar{X}$ = Arithmetic Mean, A = Assumed Mean.

Σfd_x = Total of deviations multiplied with the respective frequencies.

Σf = Total of frequencies (N).

Example : From the following frequency distributions find out the mean weight of the 100 persons :

Weight (in kg)	64	65	66	67	68	69	70	71	72	73
No. of Persons	15	13	18	5	20	11	7	6	3	2

Sol. Let assumed mean $(A) = 68$;

Calculation of Arithmetic Mean

Weight in kg (X)	No. of Persons (f)	Deviation $(d_x) = (X - A)$	fd_x
64	15	– 4	– 60
65	13	– 3	– 39
66	18	– 2	– 36
67	5	– 1	– 5
68 (A)	20	0	0
69	11	1	11
70	7	2	14
71	6	3	18
72	3	4	12
73	2	5	10
	$\Sigma f = 100$		$\Sigma fd_x = -75$

$$\overline{X} = A + \frac{\Sigma fd_x}{\Sigma f} = 68 + \frac{(-75)}{100} = 68 - 0.75 = 67.25 \text{ kg}$$

Q.4. Discuss the methods of calculation of arithmetic mean in continuous series.

Ans.

Calculation of Arithmetic Mean : Continuous Series (Grouped Data)

In continuous series, arithmetic mean can be calculated by the following methods :

1. **Direct Method :** For calculating the arithmetic mean in a continuous series the following steps are taken :

   (i) First mid value of each class-interval is determined by adding the lower and upper limit of each class and dividing the total by two.

   For example, in a class interval say 0-10, the mid value is 5.

   (ii) Multiply these mid values of each class with the respective frequency of each class. In other words X will be multiplied by f.

   (iii) Add up all the products and obtain ΣfX.

   (iv) ΣfX is divided by the sum of the frequencies *i.e.,* Σf.

   Apply the formula : $\overline{X} = \frac{\Sigma fX}{\Sigma f}\left(\frac{0+10}{2} = \frac{10}{2} = 5\right)$ (This is denoted by X).

Example : From the following find out the mean profits :

Profits per Share (₹)	100-200	200-300	300-400	400-500	500-600	600-700	700-800
No. of Shares	12	20	18	30	32	26	22

Sol.

Calculation of Mean

Profits	Mid-Point (X)	No. of Shares (f)	fX
100-200	150	12	1800
200-300	250	20	5000
300-400	350	18	6300
400-500	450	30	13500
500-600	550	32	17600
600-700	650	26	16900
700-800	750	22	16500
		$\Sigma f = 160$	$\Sigma fX = 77600$

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{77600}{160} = 485.$$

The average profit is ₹485.

2. **Step-Deviation Method :** For calculating the arithmetic mean by step-deviation method in case of grouped series the following steps are taken :
   (i) Find the mid-point of each group or class, denoted by X.
   (ii) Any value may be taken as an assumed mean of the data (A = Assumed mean).
   (iii) Subtract assumed mean from the mid-point of each class, denoted by $d_x = (X - A)$.
   (iv) Deviations (d_x) are divided by their common factor (i) and step deviation is determined for each class-interval.

$$d' = \frac{d_x}{i}.$$

Thus,
$$d' = \frac{X - A}{i}$$

   (v) The step deviation of each class is multiplied by respective frequency of that class to find the total value $\Sigma fd'$.
   (vi) Find the total frequency, $N = \Sigma f$.
   (vii) Its all sum of all products divided by frequency.
   (viii) Use the formula : $\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times i$

where, $\bar{X}$ = Arithmetic mean, A = Assumed mean

$\Sigma fd'$ = Total of product of step-deviations and frequencies;

Σf = Total number of frequencies; i = Common factor in x or in d_x.

Example : Calculate the mean from the following table :

Weights of Children (in kg)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of Children	4	12	8	21	32	28	10	3	2

Sol. Let assumed mean $(A) = 55$, let $i = 10$ (Common Factor)

C.I. (Weight in kg)	Mid Value (X)	Frequency (f)	$d_x = X - A$	$d' = d_x / 10$	fd'
0-10	5	4	– 50	– 5	– 20
10-20	15	12	– 40	– 4	– 48
20-30	25	8	– 30	– 3	– 24
30-40	35	21	– 20	– 2	– 42
40-50	45	32	– 10	– 1	– 32
50-60	55	28	0	0	0
60-70	65	10	10	1	10
70-80	75	3	20	2	6
80-90	85	2	30	3	6
Total		$\Sigma f = 120$			$\Sigma fd' = -144$

$$\text{A.M.} = A + \frac{\Sigma fd'}{\Sigma f} \times i = 55 + \frac{(-144)}{120} \times 10 = 55 - 12 = 43 \text{ kg.}$$

Q.5. Describe the methods of calculation of geometric mean with examples.

Ans.

Methods of Calculating Geometric Mean

The methods of calculating geometric mean are given as follows :

I. Computation of Geometric Mean in individual Series

Let the n values be $x_1, x_2, ..., x_N$, then

$$\text{G.M.} = \sqrt[N]{x_1, x_2, ..., x_N}$$

In order to facilitate the calculations logarithms are used.

Thus,
$$\log \text{G.M.} = \frac{\Sigma \log x}{n}$$

$\Rightarrow$
$$\text{G.M.} = \text{Antilog}\left(\frac{\log x_1 + \log x_2 + \log x_3 + ... + \log x_N}{N}\right)$$

or
$$\text{Antilog}\left(\frac{\Sigma \log x}{N}\right)$$

Steps :

1. Find the logarithm of each value from log table.
2. Find the sum of all logarithms.
3. Divide this sum by the number of values.
4. Take the antilog of this quotient.

Example : The daily income of 8 families is as follows. Find geometric mean :

Daily Income (in ₹) : 70, 10, 500, 75, 8, 250, 8, 82

Sol.

x	$\log x$
70	1.8451
10	1.0000
500	2.6990
75	1.8751
8	0.9031
250	2.3979
8	0.9031
82	1.9133
Total	**13.5371**

$$\text{G.M.} = \frac{\Sigma \log x}{N} = \frac{13.5371}{8} = 1.6921$$

$$\therefore \quad G = \text{Antilog}\,(1.6921) = 49.21$$

Example : The monthly income of 10 families in a locality is as follows. Find the Geometric Mean.

Family	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Income	85	70	15	75	500	8	45	250	40	36

Sol.

Calculation of Geometric Mean

Family	**Income (in ₹) (x)**	**Logarithms ($\log x$)**
A	85	1.9294
B	70	1.8451
C	15	1.1761
D	75	1.8751
E	500	2.6990
F	8	0.9031
G	45	1.6532
H	250	2.3979
I	40	1.6021
J	36	1.5563
$N = 10$		**$\Sigma \log x = 17.6373$**

$$\text{G.M.} = \text{Antilog}\left(\frac{\Sigma \log x}{N}\right)$$

$$= \text{Antilog}\left(\frac{17.6373}{10}\right)$$

$$= \text{Antilog}\ 1.76373 = ₹\,58.03$$

II. Computation of Geometric Mean in Discrete Series

Steps :

1. Find the logarithms of the observations, *i.e.*, find $\log x$.

2. Multiply these logarithms by corresponding frequencies, *i.e.*, find $f \log x$.
3. Add these products, *i.e.*, find $\Sigma f \log x$.
4. Divide this sum by total frequency $N = \Sigma f$.
5. Take antilog of this quotient, *i.e.*, find

$$\text{G.M.} = \text{Antilog}\left(\frac{f_1 \log x_1 + f_2 \log x_2 + \ldots + f_n \log x_n}{f_1 + f_2 + \ldots + f_n}\right)$$

$$= \text{Antilog}\left(\frac{\Sigma f \log x}{\Sigma f}\right)$$

Example : Calculate Geometric Mean from the following data :

x	7.5	13	18.5	20.5	5.2	23	24	25	26	28
f	2	3	4	5	11	13	6	3	2	1

Sol.

Calculation of Geometric Mean

x	***f***	**log *x***	***f* log *x***
7.5	2	0.8751	1.7502
13	3	1.1139	3.3417
18.5	4	1.2672	5.0688
20.5	5	1.3118	6.5590
5.2	11	0.7160	7.8760
23	13	1.3617	17.7021
24	6	1.3802	8.2812
25	3	1.3979	4.1937
26	2	1.4150	2.8300
28	1	1.4472	1.4472
Total	$\Sigma f = 50$		$\Sigma f \log x = 59.0499$

$$\text{G.M.} = \text{Antilog}\left(\frac{\Sigma f \log x}{\Sigma f}\right)$$

$$= \text{Antilog}\left(\frac{59.0499}{50}\right) = \text{Antilog}\,(1.1810) = 15.17$$

III. Computation of Geometric Mean in Grouped Series

Steps :

1. Find out the mid-values of the class-intervals.
2. Find out the logarithms of the mid-values.
3. Multiply these logarithms by corresponding frequencies.
4. Add these products.
5. Divide this sum by total frequency.
6. Take antilog of this quotient, *i.e.*, find

$$\text{G.M.} = \text{Antilog}\left(\frac{\Sigma f \log x}{\Sigma f}\right), \text{ where } x = \text{mid-value}$$

Example : The following table shows the marks obtained by 20 students in Mathematics at a certain examination :

Marks	0-10	10-20	20-30	30-40	40-50
No. of Students	1	2	6	6	5

Calculate G.M. from the above table.

Sol. **Calculation of Geometric Mean**

Marks	Mid-point (x)	Frequency (f)	Logarithms ($\log x$)	$f \log x$
0-10	5	1	0.6990	0.6990
10-20	15	2	1.1761	2.3522
20-30	25	6	1.3979	8.3874
30-40	35	6	1.5441	9.2646
40-50	45	5	1.6532	8.2660
		$N = 20$		$\Sigma f \log x = 28.9692$

$$\text{G.M.} = \text{Antilog}\left(\frac{\Sigma f \log x}{N}\right)$$

$$= \text{Antilog}\left(\frac{28.9692}{20}\right)$$

$$= \text{Antilog } 1.44846 = 28.08 \text{ marks.}$$

Q.6. Describe the methods of calculation of Harmonic mean with examples.

Ans. **Methods of Calculating Harmonic Mean**

Harmonic mean calculates by the following three methods :

I. Computation of Harmonic Mean in Individual Series

Let $x_1, x_2, \ldots, x_N$ be the N values of variable X, their harmonic mean is given by

$$\text{H.M.} = \text{Reciprocal of } \frac{\Sigma \frac{1}{x}}{N} \text{ or } \frac{N}{\Sigma \frac{1}{x}}$$

i.e.,

$$\text{H.M.} = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \ldots + \frac{1}{x_N}}$$

Steps : 1. Read off the reciprocals of the observations, *i.e.*, $\frac{1}{x}$.

2. Add these reciprocals, *i.e.*, obtain $\Sigma \frac{1}{x}$.

3. Divide this sum by number of observations, *i.e.*, find $\frac{\Sigma \frac{1}{x}}{N}$.

4. Obtain the harmonic mean, H.M. = Reciprocal of $\left(\frac{\Sigma \frac{1}{x}}{N}\right) = \frac{N}{\Sigma \frac{1}{x}}$.

Example : Calculate H.M. of the following measurements :

6, 10, 15 and 20

Sol.

$$\text{H.M.} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\frac{1}{x_4}} = \frac{4}{\frac{1}{6}+\frac{1}{10}+\frac{1}{15}+\frac{1}{20}}$$

$$= \frac{4}{(0.1667+0.10+0.0667+0.05)} = \frac{4}{0.3834} = 10.43$$

or we can be solved it as follow :

Calculation of Harmonic Mean

Measurement x	*Reciprocals* $1/x$
6	0.1667
10	0.1000
15	0.0667
20	0.0500
$N=4$	$\Sigma \frac{1}{x} = 0.3834$

H.M. = Reciprocal of $\frac{\Sigma \frac{1}{x}}{N}$ = Reciprocal of $\frac{0.3834}{4}$ = Reciprocal of 0.09585 = 10.43

Example : Calculate Harmonic Mean from the following data :

2,574 475 75 5 0.8 0.08 0.005 0.0009

Sol.

Calculation of Harmonic Mean

Items x	*Reciprocals* $1/x$
2,574	0.0003885
475	0.002105
75	0.01333
5	0.2000
0.8	1.250
0.08	12.50
0.005	200.00
0.0009	1111.00
$N=8$	$\Sigma 1/x = 1324.9658235$

$$\text{H.M.} = \text{Reciprocal of } \frac{\Sigma \frac{1}{x}}{N}$$

$$= \text{Reciprocal of } \frac{1324.9658235}{8}$$

$$= \text{Reciprocal of } 165.6207154 = 0.006039$$

II. Computation of Harmonic Mean in Discrete Series

Steps :

1. Obtain the reciprocal of each observation, *i.e.*, find $\frac{1}{x}$.
2. Multiply each $\frac{1}{x}$ by corresponding frequency, *i.e.*, find $\frac{f}{x}$. However, we can find directly the value $\frac{f}{x}$ by dividing the frequency by the corresponding observation.
3. Add these products, *i.e.*, find $\Sigma \frac{f}{x}$.
4. Find total frequency $N = \Sigma f$.
5. Divide the total by total number of observation, *i.e.*, obtain $\frac{\Sigma \frac{f}{x}}{\Sigma f}$.
6. Find Harmonic Mean by H.M. = Reciprocal of $\frac{\Sigma(f/x)}{\Sigma f} = \frac{\Sigma f}{\Sigma(f/x)}$ or $\frac{N}{\Sigma f/x}$.

Example : Find the harmonic mean from the following data :

Age (in years)	50	51	52	53	54	55
No. of Persons	2	4	10	6	2	2

Sol.

Calculation of Harmonic Mean

Age in Years (x)	**No. of Persons (f)**	**Reciprocals ($1/x$)**	**Product of Cols. (2)×(3)** $\left(f \times \frac{1}{x}\right)$
1	2	3	4
50	2	0.02000	0.04000
51	4	0.01961	0.07844
52	10	0.01923	0.19230
53	6	0.01887	0.11322
54	2	0.01852	0.03704
55	2	0.01818	0.03636
	$N = 26$		$\Sigma\left(\frac{f}{x}\right) = 0.49736$

$$\text{H.M.} = \text{Reciprocal of } \frac{\Sigma(f/x)}{N}$$

$$= \text{Reciprocal of } \frac{0.49736}{26} = \text{Reciprocal of } 0.01913 = 52.28$$

III. Computation of Harmonic Mean in Grouped Series

Step : Calculate mid-values and proceed as in the case of discrete series.

Example : Calculate Harmonic Mean :

Marks	0-10	10-20	20-30	30-40	40-50
No. of Students	4	5	11	6	4

Sol.

Calculation of Harmonic Mean

Marks	Mid-point (x)	Frequency (f)	Reciprocals $\frac{1}{x}$	product of Cols 3×4 $f \times \frac{1}{x}$
1	2	3	4	5
0-10	5	4	0.20000	0.80000
10-20	15	5	0.06667	0.33335
20-30	25	11	0.04000	0.44000
30-40	35	6	0.02857	0.17142
40-50	45	4	0.02222	0.08888
		$N = 30$		$\Sigma f \times \frac{1}{x} = 1.83365$

$$\text{H.M.} = \text{Reciprocal of } \frac{\Sigma f \times \frac{1}{x}}{N}$$

$$= \text{Reciprocal of } \frac{1.83365}{30}$$

$$= \text{Reciprocal of } 0.061122 = 16.36072$$

Q.7. Explain the relationship among Arithmetic Mean, Geometric Mean and Harmonic Mean with example.

Ans.

Relationship among A.M., G.M. and H.M.

$$\text{G.M.} = \sqrt{\text{A.M.} \times \text{H.M.}} \text{ for two positive numbers.}$$

To prove the above relations let take '*a*' and '*b*' two positive numbers.

So, $$\text{A.M.} = \frac{a+b}{2}, \text{ G.M.} = \sqrt{ab}, \text{ H.M.} = \frac{2ab}{a+b}$$

Now, $$\text{A.M.} \times \text{H.M.} = \frac{a+b}{2} \times \frac{2ab}{a+b} = ab = (\text{G.M.})^2$$

$\therefore$ $$\text{G.M.} = \sqrt{\text{A.M.} \times \text{H.M.}}$$

Example : If the A.M. and G.M. are two numbers are 20 and 16 respectively then find the harmonic mean of two numbers and also find two numbers.

Sol. Given : A.M. = 20, G.M. = 16

The relation between A.M., G.M. and H.M. is given as :

$$\sqrt{(\text{A.M.})(\text{H.M.})} = \text{G.M.} \Rightarrow \sqrt{(20)(\text{H.M.})} = (16)$$

Square of the both means, we get

$$\text{H.M.} = \frac{256}{20} = 12.8$$

Let the two numbers be X_1 and X_2. We are given that

$$\text{A.M.} = \frac{X_1 + X_2}{2} = 20 \text{ and G.M.} = \sqrt{X_1 \cdot X_2} = 16$$

$$\Rightarrow \quad X_1 + X_2 = 40 \quad \text{...(1)}$$

$$\Rightarrow \quad X_1 \cdot X_2 = 256 \quad \text{...(2)}$$

We can write,

$$(X_1 - X_2)^2 = (X_1 + X_2)^2 - 4X_1 \cdot X_2 = (40)^2 - 4 \times 256$$

$$= 600 - 1{,}024 = 576$$

$$X_1 - X_2 = 24 \quad \text{...(3)}$$

Adding equation (1) and (3), we get

$$2X_1 = 6.4$$

$$\therefore \quad X_1 = 32 \quad \text{Also } X_2 = 8$$

1. When all the values of the series differ in size, A.M. is greater than G.M. and G.M. is greater than harmonic mean, *i.e.*,

A.M. > G.M. > H.M.

Example : Using the values 4, 8 and 16, verify that A.M. > G.M. > H.M.

Sol.

$$\text{A.M.} = \frac{4+8+16}{3} = 9.33 \text{ approx}$$

$$\text{G.M.} = 3\sqrt{4 \times 8 \times 16} = (512)^{1/3} = 8$$

$$\text{H.M.} = \frac{3}{\frac{1}{4} + \frac{1}{8} + \frac{1}{12}} = \frac{3}{\frac{11}{24}} = \frac{3 \times 24}{11} = \frac{72}{11} + 6.54$$

Thus, A.M. > G.M. > H.M.

2. If all the values of the series are equal, then A.M. is equal to G.M. and G.M. is equal to H.M. *i.e.*,

A.M. = G.M. = H.M.

Example : Using the values 12 and 12, verify that A.M. = G.M. = H.M.

Sol.

$$\text{Mean} = \frac{12+12}{2} = 12$$

$$\text{Geometric Mean} = \sqrt{12 \times 12} = 12$$

$$\text{Harmonic Mean} = \frac{2}{1/12 + 1/12} = \frac{2}{2/12} = \frac{2 \times 12}{2} = 12$$

Thus, $\text{A.M.} = \text{G.M.} = \text{H.M.}$

Q.8. What do you mean by measure of dispersion. Discuss its types, objects and importance. Also write the methods of measuring dispersion?

Ans.

Measure of Dispersion

According to *Kafka*, "The measurement of a scatteredness of the mass of figures in a series about an average is called measure of dispersion or measure of variation."

Measures of Dispersion are called the *Averages of Second Order* when they are based on the average of the deviations of the different values from their mean.

Types of Measure of Dispersion

There are two types of measure of dispersion :

I. Absolute Measure of Dispersion

1. The measure of dispersion which is expressed in terms of the units of the observation (*e.g.*, Rupees, Metre, Years, etc.) is called absolute *measure*.
2. Absolute measure is not suitable to compare dispersion in two series due to two reasons :
   (i) The two series may be in different units, and
   (ii) The measure may depend on a measure of central tendency or any other fixed point.

For example, there is no comparison possible between Rupees and Kilogram.

II. Relative Measure of Dispersion

As per reasons given in absolute measure, a measure of dispersion which is independent of unit and/or may involve the point about which the deviations are taken is suggested, is known as **relative measure of dispersion** or **coefficient of Dispersion**. There is one relative measure corresponding to an absolute measure of dispersion.

Objects/Purposes of Measuring Dispersion

The objects or purposes of a measure of dispersion are as follows :

1. To find the average distance of the items from an average.
2. To know the structure of the series.
3. To gauge the reliability of an average. When the dispersion is small, the average is reliable.
4. To know the limits of the items.
5. To serve as a basis for control of the variability itself.
6. To compare two or more series with regard to their variability.

Importance of Measuring Dispersion

On the importance of measuring dispersion *Darrell Huff* opines, "Place little faith in average.....where those important figures are missing. Otherwise you are as blind as a man choosing a camp site from a report of mean temperature along....... You can freeze or roast if you ignore the range."

The importance of measuring dispersion is considered in the light of its need, objects, purpose, uses, etc. To summarize the characteristics of a set of data, both average and measure of dispersion must be presented. Dispersion is important not as merely supplementary to the average, but because the scatter in the distribution may itself be significant. According to *Spurr and Bonini*, "In matters of health, variations in body temperature, pulse beat, and blood pressure are basic guides to diagnosis. In industrial production, efficient operation requires control of quality variation, the causes of which are sought through inspection and quality control programmes."

A study of dispersion is useful to find out degree of uniformity or consistency in the two or more sets of data.

Methods of Measuring Dispersion

Since there are two meanings of dispersion, so we have two mathematical methods of measuring dispersion : (i) The method of limits, and (ii) The method of Averaging Deviation. Dispersion can also be measured graphically :

1. Method of Limits :
   (a) Range, (b) Inter-Quartile Range,
   (c) Percentile Range.
2. Method of Averaging Deviations :
   (a) Quartile Deviation or Semi-inter-quartile Range,
   (b) Mean Deviation, (c) Standard Deviation.
3. Graphic Method : Lorenze Curve

Q.9. What do you mean by Quartiles? Define the methods of calculation of Quartiles with examples.

Ans.

Quartile

The measure of central tendency which divides a group of data into four subgroups or parts then it is called **quartiles**. Firstly, date arranging into ascending or descending order for calculating quantites after that it's divided into four equal parts. The three quartiles are denoted as Q_1, Q_2 and Q_3. The first quartile, Q_1, divides a frequency distribution in such a way that one-fourth (25%) of the distribution has a value less than Q_1 and three-fourth (75%) have a value more than Q_1

The second quartile, Q_2 divides a frequency distribution in such a way that it has equal number of observations above and below it. Hence, it is equal to the median of the data. The third quartile, Q_3 divides a frequency distribution in such a way that three-fourth (75%) of the observations have a value less than Q_3 and one-fourth (25%) have a value more than Q_3.

Methods of Calculating Quartiles

I. By Individual Series

Let the series be in ascending or descending order. Let N be the number of observations. Then,

$$\text{First quartile } (Q_1) = \left(\frac{N+1}{4}\right)^{\text{th}} \text{ item;}$$

$$\text{Second quartile } (Q_2) = \left[\frac{2(N+1)}{4}\right]^{\text{th}} \text{ item or } \left(\frac{N+1}{2}\right)^{\text{th}} \text{ item}$$

$$\text{Third quartile } (Q_3) = \left[\frac{3(N+1)}{4}\right]^{\text{th}} \text{ item}$$

Example : Calculate quartiles from the following data :

Marks obtained : 06, 30, 37, 18, 14, 42, 34, 11, 09, 26, 22, 03, 28, 52, 48

Sol. Arranging the series in ascending order, we get,

Serial No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Marks Obtained	03	06	09	11	14	18	22	26	28	30	34	37	42	48	52

Number of observations, $N = 15$

$$Q_1 = \left(\frac{N+1}{4}\right)\text{th item} = \left(\frac{15+1}{4}\right)\text{th item} = 4 \text{ th item} = 11;$$

$$Q_2 = \left[\frac{2(N+1)}{4}\right]\text{th item} = \left[\frac{2(15+1)}{4}\right]\text{th item} = 8\text{th item} = 26$$

$$Q_3 = \left[\frac{3(N+1)}{4}\right]\text{th item} = \left[\frac{3(15+1)}{4}\right]\text{th item} = 12\text{th item} = 37$$

II. By Discrete Series

The computation of quartiles from discrete series involves the following steps :

1. Arrange the data in ascending or descending order of magnitude (if not arranged).
2. Find the less than type cumulative frequencies.
3. Calculate quartiles using the formulae :

$$Q_1 = \frac{N+1}{4}\text{th item};$$

$$Q_2 = \left[\frac{2(N+1)}{4}\right]\text{th item};$$

$$Q_3 = \left[\frac{3(N+1)}{4}\right]\text{th item}.$$

Example : Calculate quartile from the following data :

X	3	5	8	12	18	21	26
f	6	11	24	21	16	13	9

Sol. Quartiles can be calculated as follows :

Variable (X)	Frequency (f)	Cumulative frequency (c.f.)
3	6	6
5	11	17
8	24	41

12	21	62
18	16	78
21	13	91
26	9	100
Total	**$N=100$**	**—**

$$Q_1 = \left(\frac{N+1}{4}\right)\text{th item} = \left(\frac{100+1}{4}\right)\text{th item} = 25.25\text{ th item}$$

$$= 25\text{th item} + 0.25\ (26\text{th item} - 25\text{th item})$$

$$= 8 + 0.25\ (8-8) = 8$$

$$Q_2 = \left[\frac{2(N+1)}{4}\right]\text{th item} = 50.50\text{th item} = 50\text{th item} + 0.5\ (51\text{st item} - 50\text{th item})$$

$$= 12 + 0.5\ (12-12) = 12$$

$$Q_3 = \left[\frac{3(N+1)}{4}\right]\text{th item} = 75.75\text{th item} = 75\text{th item} + 0.75\ (76\text{th item} - 75\text{th item})$$

$$= 18 + 0.75\ (18-18) = 18$$

III. By Grouped Series

The process of computing quartiles in case of a grouped frequency distribution involves the following steps :

1. Arrange the data in ascending order.
2. Obtain less than type cumulative frequencies.
3. Convert the classes into exclusive form if given otherwise.
4. Use the following formula to calculate quartiles:

Formula :

$$Q_k = L_1 + \frac{\frac{kN}{4} - C}{f}(L_2 - L_1)$$

Where, $k = 1, 2, 3; i = (L_2 - L_1) = k$th quartile class,
N = total frequency; f = frequency of kth quartile class,
c = cumulative frequency of the class procedding the kth quartile class.

Obviously :

$$Q_1 = L_1 + \frac{\frac{N}{4} - c}{f} \times i$$

$$Q_2 = L_1 + \frac{\frac{2N}{4} - c}{f} \times i$$

$$Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i$$

The values of L_1, L_2, c, i, f are taken according to respective quartiles

Example : Calculate quartiles from the following data :

Age (Years)	0-10	10-20	20-30	30-40	40-50
No. of Persons	3	8	20	12	7

Sol. Quartiles can be calculated as follows :

Age	Number of Persons (f)	Cumulative frequency (c.f.)
0-10	3	3
10-20	8	11
20-30	20	31
30-40	12	43
40-50	7	50
Total	$N = 50$	—

$$Q_1 = \left(\frac{N}{4}\right)\text{th} = \left(\frac{50}{4}\right)\text{th}$$

$= 12.5$ th item which lies in the class-interval (20-30)

Hence, $L_1 = 20, L_2 = 30, f = 20, c = 11, i = (L_2 - L_1) = (30 - 20) = 10$

$$\Rightarrow \quad Q_1 = L_1 = \frac{\frac{N}{4} - c}{f} \times i = 20 + \frac{12.5 - 11}{20} \times 10 = 20.75$$

$$Q_2 = \left(\frac{2N}{4}\right)\text{th item} = \left(\frac{2 \times 5}{4}\right)\text{th item}$$

$= 25$th item which lies in the clas-interval (20-30)

Hence, $L_1 = 20, L_2 = 30, f = 20, c = 11, i = (L_2 - L_1) = (30 - 20) = 10$

$$\Rightarrow \quad Q_2 = L_1 = \frac{\frac{2N}{4} - c}{f} \times i = 20 + \frac{25 - 11}{20} \times 10 = 27$$

$$Q_3 = \left(\frac{3N}{4}\right)\text{th item} = \left(\frac{3 \times 50}{4}\right)\text{th item}$$

$= 37.5$th item which lies in the class-interval (30-40).

Hence, $L_1 = 30, L_2 = 40, f = 12, c = 31, i = (L_2 - L_1) = (40 - 30) = 10$

$$\Rightarrow \quad Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i = 30 + \frac{37.5 - 31}{12} \times 10 = 35.42$$

Q.10. Explain the methods of calculation of mean deviation with examples.

Ans.

Method of Calculation of Mean Deviation

Mean deviation can be calculated by following three methods :

I. Computation of Mean Deviation by Individual Series

Following steps are involved in the calculation of the mean deviation :

1. First, measure the average mean, median or mode of the series.
2. Find the deviation of the items from the average, while ignoring the positive (+) and negative (–) signs. The resultant deviation is dentoed by $|dx|$.
3. Next, measure the total sum of these deviations. It is represented by $\Sigma|dx|$.
4. In the last step, divide the sum obtained by the number of items.

Symbolically : Mean Deviation $= \frac{\Sigma|d_x|}{N}$

Where, $|d_x|$ = deviation from mean (or median) ignoring ± signs.

N = Number of items.

Example : The following are the monthly expenditure of six families. Calculate mean deviation from mean and mean deviation from median.

Expenditure (₹)	4,260	4,980	8,460	5,240	4,780	6,480

Sol. First we arrange the given data in ascending order :

Expenditure (₹)	4,260	4,780	4,980	5,240	6,480	8,460

Here, $N = 6$ which is even, so

$$\text{Median} = \frac{\left(\frac{N}{2}\right)\text{th value} + \left(\frac{N}{2}+1\right)\text{th value}}{2} = \frac{\left(\frac{6}{2}\right)\text{th value} + \left(\frac{6}{2}+1\right)\text{th value}}{2}$$

$$= \frac{\text{3rd value} + \text{4th value}}{2} = \frac{4980+5240}{2} = 5110$$

$$\bar{X} = \frac{4,260+4,780+4,980+5,240+6,480+8,460}{6} \Rightarrow \frac{34,200}{6} = 5,700$$

About Mean

S.No.	***X*** **(₹)**	**Deviation from A.M. ignoring ± sign** $|d_x|$
1	4260	1440
2	4780	920
3	4980	720
4	5240	460
5	6480	780
6	8460	2760
Total	$\Sigma X = 34200$	$\Sigma|d_x| = 7080$

About Median

S.No.	***X*** **(₹)**	**Deviation from Median ignoring ± sign** $|d_x|$
1	4260	850
2	4780	330
3	4980	130
4	5240	130
5	6480	1370
6	8460	3350
Total	$\Sigma X = 34200$	$\Sigma|d_x| = 6160$

Mean deviation (about mean) $= \frac{\Sigma|d_x|}{N} = \frac{7080}{6} = 1180$

Mean deviation (about mean) $= \frac{\Sigma|d_x|}{N} = \frac{6160}{6} = 1026.67$

II. Computation of Mean Deviation by Discrete Series

In case of discrete series following steps are involved :

1. First, calculate the averages such as mean, mode or median.
2. From the central tendency calculate the deviation of the size, ignore the positive and negative signs. The result is represented by $|d_x|$.
3. In this step, deviation of every size ($|d_x|$) is multiplied by their respective frequency (f) and summated ($\Sigma f|d_x|$).
4. In the last step, total sum is divided by the total frequency. Find result known as mean deviation.

Symbolically : $\text{M.D.} = \frac{\Sigma f|d_x|}{\Sigma f}$

Where, $|d_x|$ = deviation from mean (or median) ignoring ± signs, N = total frequency.

Example : Calculate the Mean-deviation from the following data :

Quantity Demanded (Units)	10	20	30	40	50	60	70	80	90	100
Frequency	7	13	16	6	14	19	28	17	21	9

Sol.

Calculation of Mean Deviation

Quantity Demanded (X)	Frequency (f)	fX	$\lvert d_x\rvert = (X - \bar{X})$ ignoring ± sign	$f\lvert d_x\rvert$
10	7	70	50	350
20	13	260	40	520
30	16	480	30	480
40	6	240	20	120
50	14	700	10	140
60	19	1140	0	0
70	28	1960	10	280
80	17	1360	20	340
90	21	1890	30	630
100	9	900	40	360
Total	$\Sigma f = 150$	$\Sigma fX = 9000$		$\Sigma f\lvert d_x\rvert = 3220$

Arithmetic Mean $(\bar{X}) = \frac{\Sigma fX}{\Sigma f} = \frac{9000}{150} = 60$

Mean Deviation $= \frac{\Sigma f\,|d_x|}{\Sigma f} = \frac{3220}{150} = 21.5$

III. Computation of Mean Deviation by Continuous Series

In case of continuous series, following steps are involved :

1. First, calculate the mid-value of the class intervals.
2. Measure the median or arithmetic mean.
3. Find $|d_x|$ and $f|d_x|$.
4. Divide the total sum $\Sigma f|d_x|$ by the total frequency Σf.

Example : Calculate Mean Deviation from the Mean for the following data :

Sales (₹ in thousand)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of days	5	6	8	15	17	7	9	3

Sol. **Calculation of Mean Deviation**

Sales	Mid-Value (X)	No. of days (f)	fX	$\lvert d_X\rvert = \lvert X - \bar{X}\rvert$	$f\lvert d_X\rvert$
0-10	5	5	25	35	175
10-20	15	6	90	25	150
20-30	25	8	200	15	120
30-40	35	15	525	5	75
40-50	45	17	765	5	85
50-60	55	7	385	15	105
60-70	65	9	585	25	225
70-80	75	3	225	35	105
		$\Sigma f = 70$	$\Sigma fX = 2800$		$\Sigma f\lvert d_x\rvert = 1040$

$$\text{Mean } (\bar{X}) = \frac{\Sigma fX}{\Sigma f} = \frac{2800}{70} = 40;$$

$$\text{Mean Deviation (M.D.)} = \frac{\Sigma f\,|d_x|}{\Sigma f} = \frac{1040}{70} = 14.86$$

Q.11. Discuss the methods of calculation of Standard Deviation with examples.

Ans.

Methods of Calculation of Standard Deviation

Standard deviation can be calculated by the following methods :

I. Calculation of Standard Deviation by Individual Series

There are two methods of calculating standard deviation in an individual observation or series :

1. **Deviation Taken from Actual Mean :** This method is adopted when the mean is a whole number.

   The following are the steps :

   (i) Find out the actual mean of the series.

   (ii) Find out the deviation of each value from the mean (d_x or $x = X - \bar{X}$).

   (iii) Square the deviations and take the total of squared deviations Σx^2.

(iv) Divide the total (Σx^2) by the number of observations. The square root of the quotient is standard deviation. Thus, apply the following formula :

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} \text{ or } \sqrt{\frac{\Sigma (X-\bar{X})^2}{N}}$$

2. **Deviation Taken from Assumed Mean (Short-cut Method) :** This method is adopted when the arithmetic average is a fractional value. Taking deviations from fractional value would be a very difficult and tedious task. To save time and labour, we apply short-cut method; deviations are taken from an assumed mean.

The following are the steps :

(i) Assume any one of the item in the series as an average (A).

(ii) Find out the deviations from the assumed mean; *i.e.,* ($X - A$) denoted by d_x.

(iii) Find out the total of the deviations; *i.e.,* Σd_x.

(iv) Square the deviations; *i.e.,* d_x^2 and add up the squares of deviations, *i.e.,* Σd_x^2.

(v) Then substitute the values in the following formula :

$$\sigma = \sqrt{\frac{\Sigma d_x^2}{N} - \left(\frac{\Sigma d_x}{N}\right)^2}$$

Where, σ = standard deviation; d_x stands for the deviation from assumed mean.

d_x^2 = Deviation taken from Assumed Mean Square; N = No. of Items

$\bar{X}$ = Mean, A = Assumed Mean.

Example : Find S.D. of (₹) 8, 10, 15, 24, 28.

Sol. Let the assumed mean (A) = 16.

(a) Calculation from Arithmetic Mean

X	$x = (X - \bar{X})$	x^2
8	–9	81
10	–7	49
15	–2	4
24	7	49
28	11	121
$\Sigma X = 85$		$\Sigma x^2 = 304$

(b) Calculation from Assumed Mean

X	$d_x = (X - A)$	d_x^2
8	–8	64
10	–6	36
15	–1	1
24	8	64
28	12	144
	5	$\Sigma d_x^2 = 309$

(a) **From Arithmetic Mean :** $\bar{X} = \frac{\Sigma X}{N} = \frac{85}{5} = 17$; $\sigma = \sqrt{\frac{\Sigma x^2}{N}}$ or $\frac{\Sigma (X-\bar{X})^2}{N}$

$$= \sqrt{\frac{1}{5} \times 304} = \sqrt{60.8} = 7.8$$

(b) **From Assumed Mean :** Let A (assumed mean) = 16

$$\sigma = \sqrt{\frac{\Sigma d_x^2}{N} - \left(\frac{\Sigma d_x}{N}\right)^2} = \sqrt{\frac{309}{5} - \left(\frac{5}{5}\right)^2} = \sqrt{61.8 - (1)^2} = \sqrt{60.8} = 7.8$$

Note : If the actual mean is in fraction, then it is better to take deviations from an assumed mean for avoiding too much calculation.

II. Calculation of Standard Deviation by Discrete Series (Ungrouped Data)

There are three methods for calculating standard deviation in discrete series :

1. **Actual mean Method :** It includes the following steps :
   (i) Calculate the mean of the series.
   (ii) Find deviations for various items from the mean *i.e.*, $(X - \bar{X}) = d_x$.
   (iii) Square the deviations d_x^2 and multiply by the respective frequencies (f). We get fd_x^2.
   (iv) Total the product (Σfd_x^2). Then apply the formula

$$\sigma = \sqrt{\frac{\Sigma fd_x^2}{\Sigma f}}$$

Example : Calculate standard deviation from the following series :

Age (in years)	15	25	35	45	55	65
No. of Persons	7	25	20	16	11	6

Sol.

Calculation of Standard Deviation

Age (X)	No. of Persons (f)	fX	$d_x = (X - \bar{X})$	fd_x^2
15	7	105	–22	3388
25	25	625	–12	3600
35	20	700	–2	80
45	16	720	8	1024
55	11	605	18	3564
65	6	390	28	4704
	$\Sigma f = 85$	$\Sigma fX = 3145$		$\Sigma fd_x^2 = 16360$

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{3145}{85} = 37$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\Sigma f(d_x)^2}{\Sigma f}} = \sqrt{\frac{16360}{85}} = \sqrt{192.5} = 13.87$$

2. **Assumed Mean Method :** We applied short-cut method in individual observation. Similarly we can apply it in discrete series also. Here deviations are taken, not from the actual mean, but from an assumed mean. It involves the following steps :
   (i) Assume any one of the items in the series as an average and this is called assumed average and denoted by A.
   (ii) Find out the deviations from assumed mean, *i.e.*, $(X - A)$, and denote it by d_x.

(iii) Multiply these deviations by the respective frequencies and get the Σfd_x.

(iv) Square the deviations (d_x^2).

(v) Multiply the squared deviations (d_x^2) by the respective frequencies (f) and get Σfd_x^2.

(vi) Substitute the values in the following formula :

$$\sigma = \sqrt{\frac{\Sigma fd_x^2}{\Sigma f} - \left(\frac{\Sigma fd_x}{\Sigma f}\right)^2} \quad \text{Where } d_x = X - A$$

Example : Find the Standard Deviation of the following series :

X	10	11	12	13	14	**Total**
f	4	16	22	14	6	**62**

Sol. Let the assumed mean $(A) = 12$.

Calculation of Standard Deviation

X	f	$d_x = (X - A)$	d_x^2	fd_x	fd_x^2
10	4	–2	4	–8	16
11	16	–1	1	–16	16
12	22	0	0	0	0
13	14	1	1	14	14
14	6	2	4	12	24
Total	$\Sigma f = 62$			$\Sigma fd_x = 2$	$\Sigma fd_x^2 = 70$

$$\sigma = \sqrt{\frac{\Sigma f\,(d_x)^2}{\Sigma f} - \left(\frac{\Sigma fd_x}{\Sigma f}\right)^2} = \sqrt{\frac{70}{62} - \left(\frac{2}{62}\right)^2} = \sqrt{1.13 - 0.001} = 1.06$$

3. **Step Deviation Method :** Here we take a common factor for all the items of the series. In this method the calculation becomes easy and simple. The formula for this is :

$$\sigma = \sqrt{\frac{\Sigma fd'^2_x}{\Sigma f} - \left(\frac{\Sigma fd'_x}{\Sigma f}\right)^2} \times i$$

Where, $d'_x = \frac{X - A}{i}$, i = Common Factor

Example : Find the Standard Deviation for the following distribution :

X	4.5	14.5	24.5	34.5	44.5	54.5	64.5
f	5	3	7	18	14	9	4

Sol. Let the assumed Mean $(A) = 34.5$

Calculation of Standard Deviation

X	f	$d_x = (X - A)$	$d_x = \left(\frac{d_x}{10}\right)$	fd'_x	$f(d'_x)^2$
4.5	5	– 30	– 3	– 15	45
14.5	3	– 20	– 2	– 6	12
24.5	7	– 10	– 1	– 7	7
34.5	18	0	0	0	0
44.5	14	10	1	14	14
54.5	9	20	2	18	36
64.5	4	30	3	12	36
	$\Sigma f = 60$			$\Sigma fd_x = 16$	$\Sigma f(d'_x)^2 = 150$

$$\sigma = \sqrt{\left\{\frac{\Sigma f(d'_x)^2}{\Sigma f} - \left(\frac{\Sigma fd'_x}{\Sigma f}\right)^2\right\}} \times i = \sqrt{\frac{150}{60} - \left(\frac{16}{60}\right)^2} \times 10 = 1.56 \times 10 = 15.6$$

III. Calculation of Standard Deviation by Continuous Series (Grouped Data)

In the continuous series, the method of calculating standard deviation is almost the same as in a discrete frequency distribution. But in a continuous series, mid-values of the class-intervals are to be found out.

The step deviation method is widely used. It involves the following steps :

1. Find out the mid-value of each group or class.
2. Assume one of the mid-values as an average and denote it by A.
3. Find out deviation of each mid-value from the assumed average A and denote these deviations by d_x.
4. If the class-intervals are equal, then take a common factor. Divide each deviation by the common factor and denote this column by d'_x.
5. Multiply these deviations d'_x by the respective frequencies and get $\Sigma fd'_x$.
6. Square the deviations and get d_x^2.
7. Multiply the squared deviation (d_x^2) by the respective frequencies (f). Then obtain the total : Σfd_x^2.
8. Substitute the values in the following formula to get the standard deviation.

$$\sigma = \sqrt{\frac{\Sigma fd'^2_x}{\Sigma f} - \left(\frac{\Sigma fd'_x}{\Sigma f}\right)^2} \times i\,;$$

Where, $d'_x = \frac{X - A}{i}$, i = Common factor

Example : Find the S.D. from the following figures :

Height (Inches)	44-46	46-48	48-50	50-52	52-54	Total
No. of Children	5	25	28	22	5	**85**

Sol. Let A (assumed mean) $= 49$.

Height (Inches)	Mid-point (X)	No. of Children (f)	$d_x = X - 49$	$d'_x = d_x / 2$	fd'_x	fd'^2_x
44-46	45	5	– 4	– 2	– 10	20
46-48	47	25	– 2	– 1	– 25	25
48-50	49	28	0	0	0	0
50-52	51	22	2	1	22	22
52-54	53	5	4	2	10	20
Total		$\Sigma f = 85$			$\Sigma fd'_x = -3$	$\Sigma fd'^2_x = 87$

$$\sigma = \sqrt{\left\{\frac{\Sigma f(d_x)^2}{\Sigma f} - \left(\frac{\Sigma d_x}{\Sigma f}\right)^2\right\}} \times i = \sqrt{\frac{87}{85} - \left(\frac{-3}{85}\right)^2} \times 2 = 1.01 \times 2 = 2.02$$

❑

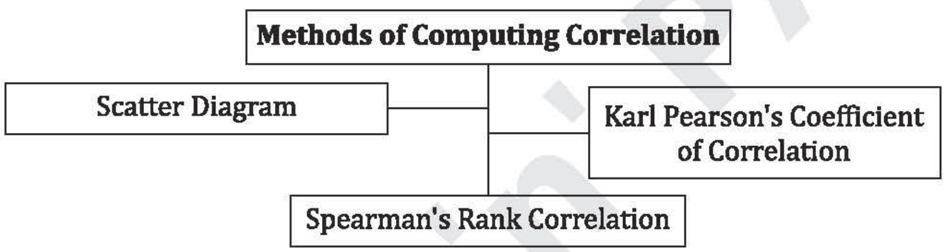
UNIT-III

Correlation

SECTION-A VERY SHORT ANSWER TYPE QUESTIONS

Q.1. Write the methods of computing correlation.

Ans. The various method of computing linear correlation can be shown diagrammatically as follows :



Q.2. What do you mean by scatter diagram.

Ans. Graphical presentation of the relationship between two variables calculated on the same set of individual is known as scatter diagram.

Q.3. The coefficient of correlation between two variables X, Y is 0.6. Their covariance is 18. The variance of X is 25. Find variance of Y series.

Sol. Given, $r = 0.6, Cov.(X, Y) = 18$ and $Var.(X) = 25$

$$r = \frac{Cov.(X,Y)}{\sqrt{Var.(X)}\sqrt{Var.(Y)}}$$

$$\Rightarrow \quad 0.6 = \frac{18}{\sqrt{25}\sqrt{Var.(Y)}} \quad \Rightarrow 0.6 = \frac{18}{5\sqrt{Var.(Y)}}$$

$$\Rightarrow \quad 3 = \frac{18}{\sqrt{Var.(Y)}} \quad \Rightarrow \quad \sqrt{Var.(Y)} = \frac{18}{3} = 6 \quad \Rightarrow \quad Var.(Y) = 36$$

Q.4. Karl Pearson's coefficient of correlation between two variables X and Y is 0.75. Their covariance is + 21. If the variance of X is 16. Find standard deviation of Y.

Sol. $r = \frac{Cov\,(x.Y)}{\sqrt{Var\,(X)\,Var\,(Y)}}$

Here, $Cov\,(X,Y) = +21,\ r = 0.75;\ Var = (X) = 16$

Hence, $$0.75 = \frac{21}{\sqrt{16}\sqrt{Var\ (Y)}}$$

$\Rightarrow$ $$\sqrt{Var\ (Y)} = \frac{21}{4 \times 0.75} = \frac{21}{3} = 7$$

$\Rightarrow$ $$\text{S.D.}\ (Y) = \sqrt{Var\ (Y)} = 7$$

Q.5. Write the methods of determining correlation.

Ans. There are various methods of determining correlation in two variables :

1. The Graphic Method, 2. Scatter Diagram or Dot Diagram or Dotogram, 3. Correlation Table 4. Karl Pearson's Coefficient of Correlation, 5. Coefficient of Concurrent Deviation, 6. Spearman's Rank Correlation Method, 7. Method of Least Squares.

Q.6. Discuss the advantages and disadvantages of Scatter Diagram.

Ans. **Advantages of Scatter Diagram :** 1. It is very simple and non-mathematical technique.

2. It is not influenced by the size of extreme item.
3. It is the very basic step to find out the relationship between two variables.

Disadvantages of Scatter Diagram : The main disadvantage of this technique is that it cannot find out an exact degree of correlation between two variables. We can only view the visual form of correlation and direction (positive or negative) on the chart.

Q.7. Calculate coefficient of correlation when Covariance of *X* and *Y* is 488 and Variance of *X* is 824 and Variance of *Y* is 325.

Ans. Coefficient of Correlation $(r) = \frac{Cov.\ (X, Y)}{\sqrt{Var.\ (X)}\sqrt{Var.\ (Y)}} = \frac{488}{\sqrt{824 \times 325}} = 0.943.$

Q.8. Define the advantages bond disadvantages of Spearman's rank method.

Ans. **Advantages of Spearman's Rank Method :** 1. Easy to understand and simple to calculate.

2. When data are qualitative (*e.g.*, intelligence, efficiency) in nature then it is very useful.
3. It also applies when actual data are given.

Disadvantages of Spearman's Rank Method : 1. Not useful in frequency distribution.

2. If we take a large sample $(n > 30)$ then calculation becomes more difficult and takes too much time.

Q.9. Write the merits of Karl Pearson's coefficient of correlation.

Ans. Merits of Karl Pearson's coefficient of correlation are :

1. Satisfactory from algebraic point of view : This measure is quite satisfactory on algebraic grounds, because its calculation is based on all values of both the series.
2. Ideal statistical measure : It is ideal from statistical point of view because it is based on arithmetic mean and standard deviation.
3. Measurement of both the degree and direction : An important feature of the measure is that it gives idea about the direction as well as degree of correlation. The positive or negative sign of coefficient tells about the direction, while the quantitative measurement explains the degree or density of the relationship.

Q.10. Explain the limitations of Karl Pearson's coefficient of correlation.

Ans. Limitations of Karl Pearson's coefficient of correlation are :

1. **Complicated calculations :** The process of calculation is comparatively complicated.
2. **Assumption of linear relationship :** An assumption of this coefficient is that there is linear relationship between two series but in many cases it may not be so.
3. **Need of careful interpretation :** The coefficient of correlation needs careful interpretation, so that it can be easily understood.
4. **Effect of extreme values :** The value of Karl Pearson's coefficient of correlation is unduly affected by extreme items.

SECTION-B SHORT ANSWER TYPE QUESTIONS

Q.1. Define the importance and applications of correlation analysis.

Ans. Importance and Applicatons of Correlation Analysis

Correlation analysis is a very important technique in Statistics. It is useful in physical and social sciences and business and economics.

The correlation analysis is useful in the following cases :

1. To have more reliable forecasting.
2. To study economic activities.
3. To estimate the variable values on the basis of an other variable values.
4. To make analysis, drawing conclusions etc. in the research or statistical investigations.

In economics, we study the relationship between price and demand, price and supply, income and expenditure, etc. According to *Neiswanger*, "Correlation analysis contributes to the understanding of economic behaviour, aids in locating the critically important variables on which others depend, may reveal to the economist the connections by which disturbances spread and suggests to him the paths through which stabilizing forces may become effective."

To a businessman correlation analysis helps to estimate costs, sales, prices and other related variables.

Correlation analysis is the basis of the concept of regression and ratio of variation.

According to *Tippet*, "The effects of the correlation is to reduce the range of uncertainty of our prediction."

Q.2. From the data given below, calculate Karl Pearson's coefficient of correlation and interpret it.

Price	11	12	13	14	15	16	17	18	19	20
Supply	30	29	29	25	24	24	24	21	18	25

Sol. Calculation of Karl Pearson's Coefficient of Correlation

Price (X)	Supply (Y)	$x = X - \bar{X}$	x^2	$y = (Y - \bar{Y})$	y^2	xy
11	30	– 4.5	20.25	5.1	26.01	– 22.95
12	29	– 3.5	12.25	4.1	16.81	– 14.35

13	29	– 2.5	6.25	4.1	16.81	– 10.25
14	25	– 1.5	2.25	0.1	0.01	– 0.15
15	24	– 0.5	0.25	– 0.9	0.81	0.45
16	24	0.5	0.25	– 0.9	0.81	– 0.45
17	24	1.5	2.25	– 0.9	0.81	– 1.35
18	21	2.5	6.25	– 3.9	15.21	– 9.75
19	18	3.5	12.25	– 6.9	47.61	– 24.15
20	25	4.5	20.25	0.1	0.01	0.45
$\Sigma X = 155$	$\Sigma Y = 249$		$\Sigma x^2 = 82.5$		$\Sigma y^2 = 124.9$	$\Sigma xy = -82.5$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{155}{10} = 15.5; \bar{Y} = \frac{\Sigma Y}{N} = \frac{249}{10} = 24.9$$

$$\text{Coefficient of correlation } (r) = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{-82.5}{\sqrt{82.5 \times 124.9}}$$

$$= \frac{-82.5}{\sqrt{10304.25}} = \frac{-82.5}{\sqrt{101.51}} = -0.81$$

Q.3. Calculate the coefficient of correlation for the following data :

Price	50	55	48	54	60	56	58	59
Supply	90	110	75	100	120	110	115	120

Sol.

Calculation of Coefficient of Correlation

Price (X)	**Supply (Y)**	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
50	90	– 5	– 15	25	225	75
55	110	0	5	0	25	0
48	75	– 7	– 30	49	900	210
54	100	– 1	– 5	1	25	5
60	120	5	15	25	225	75
56	110	1	5	1	25	5
58	115	3	10	9	100	30
59	120	4	15	16	225	60
$\Sigma X = 440$	$\Sigma Y = 840$			$\Sigma x^2 = 126$	$\Sigma y^2 = 1750$	$\Sigma xy = 460$

$$\bar{X} \frac{\Sigma x}{N} = \frac{440}{8} = 55, \bar{Y} = \frac{\Sigma Y}{N} = \frac{840}{8} = 105$$

$$\text{Coefficient of correlation } (r) = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}} = \frac{460}{\sqrt{126 \times 1750}} = \frac{460}{469.57} = 0.9796$$

Q.4. Calculate Karl Pearson's Coefficient of Correlation from the following data. Assume 125 and 190 as working mean for X and Y.

X-Series	112	114	108	124	145	150	190	125	147	150
Y-Series	200	190	214	187	170	170	210	190	180	180

Sol. Given Assumed Means :

$$A_x = 125; A_y = 190; d_x = X - A_x; d_y = X - A_y; N = 10$$

X	Y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
112	200	−13	10	169	100	−130
114	190	−11	0	121	0	0
108	214	−17	24	289	576	−408
124	187	−1	−3	1	9	3
145	170	20	−20	400	400	−400
150	170	25	−20	625	400	−500
190	210	65	20	4225	400	1300
125	190	0	0	0	0	0
147	180	22	−10	484	100	−220
150	180	25	−10	625	100	−250
		$\Sigma d_x = 115$	$\Sigma d_y = -9$	$\Sigma d_x^2 = 6939$	$\Sigma d_y^2 = 2085$	$\Sigma d_x d_y = -605$

$$r = \frac{N\Sigma d_x d_y - \Sigma d_x \Sigma d_y}{\sqrt{\{N\Sigma d_x^2 - (\Sigma d_x)^2\}\{N\Sigma d_y^2 - (\Sigma d_y)^2\}}}$$

$$= \frac{10(-605) - (115)(-9)}{\sqrt{\{10(6939) - (115)^2\}\{10(2085 - (-9)^2\}}}$$

$$= \frac{-5015}{34153.93} = -0.1468.$$

Hence, $r = -0.1468$

Q.5. Calculate Karl Pearson's coefficient of correlation for the following series :

Sale (in lakh ₹)	16	20	26	30	28	18	22	32
Expenditure (in lakh ₹)	36	42	48	52	50	38	44	50

Sol. Let the assumed mean for A_x is 30, Assumed mean for A_y is 40.

X	$d_x = (X - A_x)$	$d'_x = d_x/2$	$(d'_x)^2$	Y	$d_y = (Y - A_y)$	$d'_y = d_y/2$	$(d'_y)^2$	$d'_x\ d'_y$
16	−14	−7	49	36	−4	−2	4	14
20	10	−5	25	42	2	1	1	−5
26	−4	−2	4	48	8	4	16	−8
30	0	0	0	52	12	6	36	0

28	–2	–1	1	50	10	5	25	–5
18	–12	–6	36	38	–2	–1	1	6
22	–8	–4	16	44	4	2	4	–8
32	2	1	1	50	10	5	25	5
		$\Sigma d'_x = -24$	$\Sigma d'^2_x = 132$			$\Sigma d'_y = 20$	$\Sigma d'^2_y = 112$	$\Sigma d'_x d'_y = -1$

$$r = \frac{\Sigma d'_x d'_y - \frac{\Sigma d'_x d'_y}{N}}{\sqrt{\Sigma (d'_x)^2 - \frac{(\Sigma d'_x)^2}{N}}\sqrt{\Sigma (d'_y)^2 - \frac{(\Sigma d'_y)^2}{N}}} = \frac{-1-\left(-\frac{1}{8}\right)}{\sqrt{132 - \frac{(-24)^2}{8}}\sqrt{112 - \frac{(20)^2}{8}}}$$

$$= \frac{\left(\frac{-8+1}{8}\right)}{\sqrt{132-72} \times \sqrt{112-50}} = \frac{\left(\frac{-7}{8}\right)}{\sqrt{60} \times \sqrt{62}} = \frac{-7}{8 \times 61} = -0.014 \times 2 = -0.028$$

Q.6. A researcher wished to determine if a person's age is related to the number of hours he or she exercises per week. The data obtained from a sample is given. State your opinion based on Karl Pearson's coefficient of correlation for the data.

Age	16	24	38	46	54	62
Hours	11	7	5	4	2	1

Sol.

Calculation of Coefficient of Correlation

Age (X)	Hours (Y)	X^2	Y^2	XY
16	11	256	121	176
24	7	576	49	168
38	5	1444	25	190
46	4	2116	16	184
54	2	2916	4	108
62	1	3844	1	62
$\Sigma X = 240$	$\Sigma Y = 30$	$\Sigma X^2 = 11152$	$\Sigma Y^2 = 216$	$\Sigma XY = 888$

Coefficient of correlation $(r) = \dfrac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$

$$= \frac{6 \times 888 - (240)(30)}{\sqrt{6 \times 11152 - (240)^2}\sqrt{6 \times 216 - (30)^2}}$$

$$= \frac{-1872}{\sqrt{9312} \times \sqrt{396}} = \frac{-1872}{1920.3} = -0.975$$

So, there is a negative correlation between age and number of hours he/she exercises per week.

Q.7. From the following data, calculate coefficient of correlation between the age of students and their playing habits.

Age (Years)	15	16	17	18	19	20
No. of Students	250	200	150	120	100	80
Regular Players	200	150	90	48	30	12

Sol. Since it is asked to find the correlation between age and playing habits, it is required to find the percentage of regular players which is obtained as follows :

No. of Students	Regular Players	% of Regular Players
250	200	$\frac{200}{250}\times 100 = 80$
200	150	$\frac{150}{200}\times 100 = 75$
150	90	$\frac{90}{150}\times 100 = 60$
120	48	$\frac{48}{120}\times 100 = 40$
100	30	$\frac{30}{100}\times 100 = 30$
80	12	$\frac{12}{80}\times 100 = 15$

Let the Assumed mean for X is 17, Assumed mean for Y is 40.

X (Age)	Y (% of Regular Player)	$d_x = (X - 17)$	$d_y = (Y - 40)$	d_x^2	d_y^2	$d_x d_y$
15	80	– 2	40	4	1600	– 80
16	75	– 1	35	1	1225	– 35
17	60	0	20	0	400	0
18	40	1	0	1	0	0
19	30	2	– 10	4	100	– 20
20	15	3	– 25	9	625	– 75
Total		$d_x = 3$	$d_y = 60$	$\Sigma d_x^2 = 19$	$\Sigma d_y^2 = 3950$	$\Sigma d_x d_y = -210$

$$r = \frac{N\Sigma d_x d_y - \Sigma d_x \Sigma d_y}{\sqrt{\{N\Sigma d_x^2 - (\Sigma d_x)^2\} \times \{N\Sigma d_y^2 - (\Sigma d_y)^2\}}} = \frac{6(-210) - 3\times 60}{\sqrt{6\times 19 - (3)^2} \times \sqrt{6\times 3950 - (60)^2}} = -0.99$$

The variable X and Y are perfectly negatively correlated.

Q.8. A survey regarding the income and savings of 100 school teachers in a certain city is provided in the following data :

Income (₹)	Savings (₹)				Total
	50	100	150	200	
400	8	4	–	–	12
600	–	12	24	6	42
800	–	9	7	2	18
1000	–	–	10	5	15
1200	–	–	9	4	13
Total	**8**	**25**	**50**	**17**	**100**

Calculate coefficient of correlation (r).

Sol. Let the income be denoted by X and the savings by Y.

Calculation of Coefficient of Correlation

		Savings							
		50	100	150	200				
		dy –1	0	1	2	f	fd_x	fd_x^2	fd_xd_y
400	d_x –2	16(fd_xd_y) 8	0 4			12	–24	48	16
600	–1		0 12	–24 24	–12 6	42	–42	42	–36
800	0		0 9	0 7	0 2	18	0	0	0
1000	1			10 10	10 5	15	15	15	20
1200	2			18 9	16 4	13	26	52	34
f		8	25	50	17	$N=100$	$\Sigma fd_x=-25$	$\Sigma fd_x^2=157$	$\Sigma fd_xd_y=34$
fd_y		–8	0	50	34	$\Sigma fd_y=76$			
fd_y^2		8	0	50	68	$\Sigma fd_y^2=126$			
fd_xd_y		16	0	4	14	$\Sigma fd_xd_y=34$			

$$r=\frac{\Sigma f_xd_y-\frac{(\Sigma fd_x)(\Sigma fd_y)}{N}}{\sqrt{\Sigma fd_x^2-\frac{(fd_x)^2}{N}}\sqrt{\frac{(\Sigma fd_y)^2}{N^2}}}$$

$\Sigma fd_x d_y = 34, \Sigma fd_x = -25, \Sigma fd_y = 76; \Sigma fd_x^2 = 157, \Sigma fd_y^2 = 126, N = 100$

$$r = \frac{34 - \frac{(-25)\times 76}{100}}{\sqrt{\left\{157 - \frac{(-25)^2}{100}\right\} \times \left\{126 - \frac{(76)^2}{100}\right\}}} = 0.523$$

Q.9. Given below the following data, compute Karl Pearson's coefficient of correlation :

X/Y	0-20	20-40	40-60
10-25	10	5	3
25-40	4	40	8
40-55	6	9	15

Sol. $d_x = \frac{m.p_x - 30}{20}, d_y = \frac{m.p_y - 32.5}{15}$

X \ *Y*	$m.p_x$		0-20	20-40	40-60				
			10	30	50				
	$m.p_y$	d_x \ d_y	−1	0	1	f	fd_y	fdy^2	$fd_x d_y$
10-25	17.5	−1	10 ($fd_x d_y$) 10	0 5	−3 3	18	−18	18	7
25-40	32.5	0	0 4	0 40	0 8	52	0	0	0
20-55	47.5	1	−6 6	0 9	15 15	30	30	30	9
		f	20	54	26	$\Sigma f = 100$	$\Sigma fd_y = 12$	$\Sigma fd_y^2 = 48$	$\Sigma fd_x d_y = 16$
		fd_x	−20	0	26	$\Sigma fd_x = 6$			
		fd_x^2	20	9	26	$\Sigma fd_x^2 = 46$			
		$fd_x d_y$	4	0	12	$\Sigma fd_x d_y = 16$			

$$b_{xy} = \frac{\Sigma fd_x d_y - \frac{(\Sigma fd_x)(\Sigma fd_y)}{N}}{\Sigma fd_y^2 - \frac{(\Sigma fd_y)^2}{N}} \times \frac{i_x}{i_y} = \frac{16 - \frac{6\times 12}{100}}{48 - \frac{(12)^2}{100}} \times \frac{20}{15}$$

$$= \frac{(16 - 0.72)}{(48 - 1.44)} \times \frac{4}{3} = \frac{61.12}{139.68} = 0.438$$

$$b_{yx} = \frac{\Sigma f d_x d_y - \frac{(\Sigma f d_x)(\Sigma f d_y)}{N}}{\Sigma f d_x^2 - \frac{(\Sigma f d_x)^2}{N}} \times \frac{i_y}{i_x} = \frac{16 - \frac{6 \times 12}{100}}{46 - \frac{(6)^2}{100}} \times \frac{15}{20}$$

$$= \frac{(16 - 0.72)}{(46 - 0.36)} \times \frac{3}{4} = \frac{45.84}{182.56} = 0.251$$

$$r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{0.438 \times 0.251} = \sqrt{0.1099} = 0.331 \text{ (Approx)}$$

Q.10. Ten girls in a beauty contest are ranked by three Judges in the following order :

Judge *A*	3	8	2	4	7	9	5	10	6	1
Judge *B*	6	10	1	4	8	7	3	9	5	2
Judge *C*	5	7	2	3	9	10	6	1	8	4

Using Spearman's coefficient of correlation, find which pair of Judges have the nearest approach to common taste in beauty?

Sol.

Calculation of Rank Correlation

R_1	R_2	R_3	$D_1 = (R_1 - R_2)^2$	$D_2 = (R_2 - R_3)^2$	$D_3 = (R_1 - R_3)^2$
3	6	5	9	1	4
8	10	7	4	9	1
2	1	2	1	1	0
4	4	3	0	1	1
7	8	9	1	1	4
9	7	10	4	9	1
5	3	6	4	9	1
10	9	1	1	64	81
6	5	8	1	9	4
1	2	4	1	4	9
$N = 10$			**26**	**108**	**106**

Rank correlation is given by $R = 1 - \frac{6\Sigma D^2}{N^3 - N}$

Rank correlation between Judge *A* and *B*

$$= 1 - \frac{6\Sigma D_1}{N^3 - N} = 1 - \frac{6 \times 26}{10^3 - 10} = 1 - \frac{156}{990} = \frac{990 - 156}{990} = 0.842$$

Rank correlation between Judge *B* and *C*

$$= 1 - \frac{6\Sigma D_2}{N^3 - N} = 1 - \frac{6 \times 108}{10^3 - 10} = 1 - \frac{648}{990} = \frac{990 - 648}{990} = 0.345$$

Rank correlation between Judge A and C

$$=1-\frac{6\Sigma D_3}{N^3-N}=1-\frac{6\times 106}{10^3-10}=1-\frac{636}{990}=\frac{990-636}{990}=0.357$$

Since coefficient is maximum in the judgements of the Judge A and Judge B, so they have the nearest approach to common taste in beauty.

Q.11. The marks obtained by 9 students in two subjects *A* and *B* are given below. Compute the Rank Correlation Coefficient of Spearman.

Marks in *A*	35	23	47	17	10	43	9	6	28
Marks in *B*	30	33	45	23	8	49	12	4	31

Sol. **Rank Correlation Coefficient of Spearman**

x	y	Ranks in x R_1	Ranks in y R_2	$D=R_1-R_2$	D^2
35	30	3	5	–2	4
23	33	5	3	2	4
47	45	1	2	–1	1
17	23	6	6	0	0
10	8	7	8	–1	1
43	49	2	1	1	1
9	12	8	7	1	1
6	4	9	9	0	0
28	31	4	4	0	0
$N=9$					$\Sigma D^2=12$

$$r=1-\frac{6\Sigma D^2}{N(N^2-1)}=1-\frac{6\times 12}{9(81-1)}=1-\frac{72}{720}=1-\frac{1}{10}=\frac{9}{10}=0.9$$

The high value of r indicates a very high relationship. This means that the students who are good in A are good in B also and *vice versa*.

Q.12. From the following data, calculate rank correlation coefficient :

X	115	109	112	87	98	98	120	100	98	118
Y	75	73	85	70	76	65	82	73	68	80

Sol. **Calculation of Rank Correlation Coefficient**

X	Rank (R_1)	Y	Rank (R_2)	$D=(R_1-R_2)$	D^2
115	3	75	5	–2	4
109	5	73	6.5	–1.5	2.25
112	4	85	1	3	9
87	20	70	8	2	4
98	8	76	4	4	16

98	8	65	10	−2	4
120	1	82	2	−1	1
100	6	73	6.5	−0.5	0.25
98	8	68	9	−1	1
118	2	80	3	−1	1
					$\Sigma D^2 = 42.5$

Coefficient of Rank Correlation $R = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2)\ldots\right)}{N^3 - N}$

$$= R = 1 - \frac{6\left(\Sigma 42.5 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2)\ldots.\right)}{10^3 - 10} = 0.724 \text{ approx}$$

Q.13. Calculate Coefficient of Rank Correlation for the following data :

Capital (in lakh ₹)	66	55	46	33	22	18	11	8	7	11
Profit (in lakh ₹)	58	43	36	27	15	9	12	15	6	14

Sol. **Calculation of Coefficient of Rank Correlation**

Capital (X)	Profit (Y)	Rank (R_1)	Rank (R_2)	$D = R_1 - R_2$	D^2
66	58	10	10	0	0
55	43	9	9	0	0
46	36	8	8	0	0
33	27	7	7	0	0
22	15	6	5.5	0.5	0.25
18	9	5	2	3	9
11	12	3.5	3	0.5	0.25
8	15	2	5.5	−3.5	12.25
7	6	1	1	0	0
11	14	3.5	4	−0.5	0.25
$N = 10$					$\Sigma D^2 = 22$

Here, number 11 is repeated twice in series X and number 15 is repeated twice in series Y. Therefore, in $X, m_1 = 2$ and in $Y, m_2 + 2$.

$$R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2)\right]}{N^3 - N}$$

$$= 1 - \frac{6\left[22 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{10^3 - 10}$$

$$=1-\frac{6\left[22+\frac{1}{12}(6)+\frac{1}{12}(6)\right]}{1000-10}$$

$$=1-\frac{6\left[22+\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right)\right]}{990}=1-\frac{6(23)}{990}$$

$$=1-\frac{138}{990}=\frac{990-138}{990}=\frac{852}{990}=0.86$$

Q.14. The value of ordinary correlation (r) for the following data is 0.636.

X	0.05	0.14	0.24	0.30	0.47	0.52	0.57	0.61	0.67	0.72
Y	1.08	1.15	1.27	1.33	1.41	1.46	1.54	2.72	4.01	9.63

Calculate Spearman's rank correlation (P) for this data.

Sol. **Calculation of Rank Correlation Coefficient**

X	Rank (R_1)	Y	Rank (R_2)	$D=(R_1-R_2)$	D^2
0.5	1	1.08	1	0	0
0.14	2	1.15	2	0	0
0.24	3	1.27	3	0	0
0.30	4	1.33	4	0	0
0.47	5	1.41	5	0	0
0.52	6	1.46	6	0	0
0.57	7	1.54	7	0	0
0.61	8	2.72	8	0	0
0.67	9	4.01	9	0	0
0.72	10	9.63	10	0	0
					$\Sigma D^2=0$

Coefficient of Rank Correlation $(R)=1-\frac{6\Sigma D^2}{N(N^2-1)}=1-\frac{6\times 0}{10(100-1)}=1$

Q.15. Calculate r_k of the following data :

X	75	73	72	72	63	62	55	50
Y	62	58	67	45	81	60	67	48

Sol. **Calculation of Rank Correlation**

X	Rank (R_1)	Y	Rank (R_2)	$D=(R_1-R_2)$	D^2
75	1	62	4	– 3	9
73	2	58	6	– 4	16
72	3.5	67	2.5	1	1

72	3.5	45	8	– 4.5	20.25
63	5	81	1	4	16
62	6	60	5	1	1
55	7	67	2.5	4.5	20.25
50	8	48	7	1	1
					$\Sigma D^2 = 84.5$

In series X the value 72 occurs thrice $(m_1 = 2)$, *i.e.*, at 3rd and 4th rank. Hence all two values are given the average rank, *i.e.*, $\left(\frac{3+4}{2}\right)$th $= 3.5$th rank. While in series Y, the value 67 occurs both at the 2nd and 3rd rank, $(m_2 = 2)$. Hence, both are given the average rank, this is,

$$= \left(\frac{2+3}{2}\right)\text{th} = 2.5\text{th rank}$$

Coefficient of Rank Correlation

$$(r_k) = 1 - \frac{6\left[\Sigma D^2 + \frac{m_1^3 - m_1}{12} + \frac{m_2^3 - m_2}{12}\right]}{N(N^2 - 1)}$$

$$= 1 - \frac{6\left[84.5 + \frac{2^3 - 2}{12} + \frac{2^3 - 2}{12}\right]}{8(8^2 - 1)}$$

$$= 1 - \frac{6[84.5 + 0.5 + 0.5]}{8(8^2 - 1)}$$

$$= 1 - \frac{6[85.5]}{504} = 1 - \frac{513}{504} = 1 - 1.0178 = -0.0178$$

SECTION-C LONG ANSWER TYPE QUESTIONS

Q.1. Explain the meaning of correlation. Mention also its various types.

Ans.

Meaning and Definition of Correlation

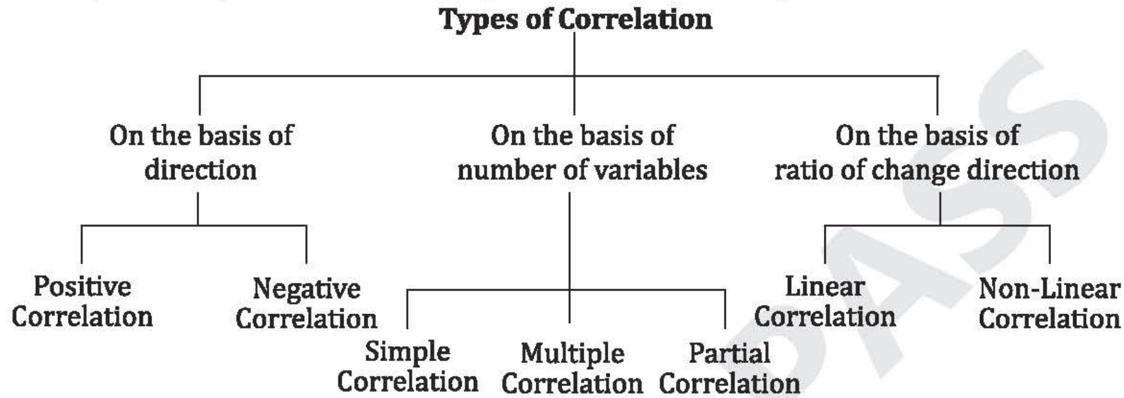
Study of the linear relationship between two variables is known as correlation. If there are two variables and changes in the value of one variable will affect the value of the other variable then both the variables are correlated. **Correlation** is a statistical tool used to measure the relationship between two sets of variables and express each in a precise manner.

According to *Croxton and Cowden*, "The appropriate statistical tool for discovering and measuring the relationship of quantitative nature and expressing it in brief formula is known as correlation".

According to *A.M. Tuttle*, "An analysis of the covariance of two or more variables is usually called correlation".

Types of Correlation

Categorisation of correlation is done on various bases such as on the direction of change of the variable, constancy of the ratio of change between the variable, etc.



I. Positive Correlation and Negative Correlation

Depending on the direction of change of the variables, correlation may be positive or negative.

1. **Positive Correlation :** Correlation is positive when two variables vary in the same direction.

   For example, consider the correlation between sales and expenses of a firm (table below).

Firms	1	2	3	4	5	6
Sales (in ₹)	50	60	70	80	90	100
Expenses (in ₹)	11	15	19	23	27	31

2. **Negative correlation :** Two variables are said to be negatively correlated when both the variables vary in the opposite direction. When one variable increases then other variable decreases and *vice versa* is also a negative correlation. *For example,* consider the correlation between production and price of crop.

Production in (kg)	100	200	300	400	500	600	700	800
Price (₹ per kg)	10	9	8	7	6	5	4	3

II. Linear Correlation and Non-linear (Curvilinear) Correlation

On the basis of constancy of the ratio of change between the variable, correlation may be linear or non-linear :

1. **Linear Correlation :** In a linear correlation, change in the values of one variable has a fixed ratio to the variation in the values of other variable. When these variables are plotted on a graph, then plotted points would fall on a straight line. *For example,* consider the following relationship shown in the table.

Mustard (kg)	10	20	30	40	50	60	70	80
Oil (kg)	3	6	9	12	15	18	21	24

2. **Non-Linear (Curvilinear) Correlation :** In a non-linear correlation, change in values of one variable does not have a fixed ratio to the variation in the value of other variable. When these variables are plotted on a graph, then plotted points would fall on a curve. *For example,* consider the following relationship shown in the table.

Use of Fertilizer (in kg)	1	2	3	4	5
Production of Rice (in kg)	4	6	9	17	28

III. Simple, Partial and Multiple Correlation

Depending on the number of variables studied, it can be categorised into three types :

1. **Simple Correlation :** When we measure the linear relationship between two variables then this interpretation is known as simple correlation, e.g. relationship between sales and expenses, income and consumption, etc.
2. **Partial Correlation :** If we have various related variables and try to find out the relationship between two variables then it is known as partial correlation. *For example,* consider the two variables height and weight, which are partially correlated because of the effect of third variable 'age' on height and weight. In this condition, if we neglect the effect of age, and study the relationship between height and weight, then this correlation is known as partial correlation.
3. **Multiple Correlations :** It is defined as the measurement of the effect of multiple variables on one variable. *For example,* if we try out to find the relationship of rainfall and temperature on the yield of wheat, then this is known as multiple correlations.

Q.2. What do you mean by Degree of correlation and correlation coefficient? Write the applications of correlation also.

Ans.

Degree of Correlation

On the basis of coefficient of correlation, the degree of correlation is determined as follows :

1. **Perfect Correlation :** If two variables change in the same direction and in the same proportion, the correlation between the two variables is called perfect positive correlation. According to **Karl Pearson**, the coefficient of correlation in this case is +1. If two variables change in the opposite direction and in the same proportion, the correlation between the two variables is called perfect negative correlation. In this case the coefficient of correlation is – 1.
2. **Absence of Correlation :** If two series of two variables show no relation between them or change in one variable does not lead to a change in the other variable, then it means that there is no relationship between variables. In this case, the coefficient of correlation is zero.
3. **Limited Degree of Correlation :** If two variables are not perfectly correlated or there is an absence of perfect correlation, then it is referred as Limited correlation. It may be positive, negative or zero but lies with the limits ± 1. High degree, moderate degree and low degree are the three categories of this kind of correlation.

   If the value of coefficient of correlation (represented by '*r*') lies between ± 0.75 to ± 1, it is known as **High degree** of correlation and when the value of r lies between ± 0.25 to ± 0.75 and it is known as Moderate degree of correlation. When the correlation lies between 0 and ± 0.25, low degree of correlation is said to exist look at the table below :

Table : Degree of Correlation

	Degree	Positive Correlation Co-efficient	Negative Correlation Co-efficient
1.	Perfect	+1	− 1
2.	Limited		
	(i) High	Between + 0.75 to + 1	Between − 0.75 to − 1
	(ii) Moderate	Between + 0.25 to + 0.75	Between − 0.25 to − 0.75
	(iii) Low	Between 0 to 0.25	Between 0 to − 0.25
3.	Absence	Zero	Zero

Correlation Coefficient

The extent of measurement as to how much one number can expect to be influenced by changes in another number is known as correlation coefficient. *For example*, the measuring value that how much the share price of a company is expected to get influenced, when there are fluctuations in the index. If the two numbers are perfectly correlated then it means that value of correlation coefficient is ± 1.

If one grows, the other also grows, and the change in one is a multiple of the change in the other. If the numbers are perfectly inversely correlated then it means that value of correlation coefficient is − 1. If one grows, the other falls. The growth in one variable is a negative multiple of the growth in the other variable. If the two variables are not related then the correlation coefficient is zero.

When numbers are related and the coefficient is either 1 or − 1 and other influences and relationship between two numbers is not fixed then this is known as non-zero correlation coefficient. If one number is known, the other number can be calculated, but with uncertainty. Uncertainty increases when correlation coefficient is more towards to zero. But this relationship is not useful, if the correlation coefficient is very low.

Applications of Correlation

1. The nature, direction and degree of relationship between two or more variables are determined by the use of correlation analysis.
2. It is used for estimating the change in value of one variable occurs due to the changes in the value of other variable.
3. Under some specific conditions, it helps the researchers to understand the behaviour of certain events. *For example*, one can find the various factors which are related to rainfall, in a specific area and determine how these factors affect the production of rice.
4. Correlation analysis provides the platform of decision-making to business word and reduces the factor of uncertainty from the decision-making.
5. Correlation analysis is helpful in making predictions.

Q.3. Define Karl Pearson's coefficient of correlation, its properties, advantages and disadvantages. Also discuss the methods of calculation of Karl Pearson's coefficient of correlation with examples.

Ans. Karl Pearson's Coefficient of Correlation

To calculate the magnitude of linear relationship between two variables Karl Pearson (biometrician and statistician) gave a quantitative technique. This technique is known as Pearsonian Coefficient of Correlation (r) and is extensively used in practice.

Properties of Karl Pearson's Coefficient of Correlation

1. Ideal measure of correlation and is independent of the units of the variables.
2. Free from change of origin and scale.
3. Based on all observations.
4. Lies between – 1 and + 1 and when :
   (i) $r = -1$, shows a perfect negative correlation between variables.
   (ii) $r = 0$, then there is no correlation between variables.
   (iii) $r = +1$, shows a perfect positive correlation between variables.

Advantage of Karl Pearson's Coefficient of Correlation

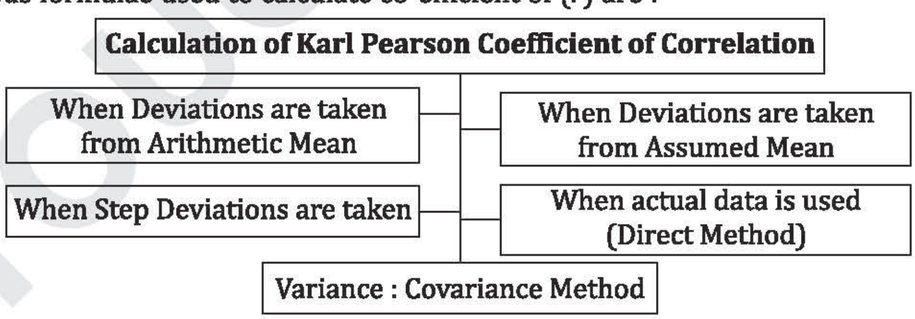
The main advantage of the Karl Pearson's coefficient of correlation is that it gives the result in one value and also summarizes the degree of correlation and direction.

Disadvantages of Karl Pearson's Coefficient of Correlation

1. Every time assumes only a linear relationship between variables.
2. Interpreting the value of correlation coefficient (r) is difficult.
3. Significance of Correlation coefficient (r) is affected by the extreme values.
4. It is a time consuming method.
5. Does not convey the cause and affect relationship.

Calculation of Karl Pearson Coefficient of Correlation

The various formulae used to calculate co-efficient of (r) are :



I. When Deviations are Taken from Arithmetic Mean

When deviation is taken from arithmetic mean, then r is calculated as :

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} \text{ or } \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} \text{ or } \frac{\Sigma xy}{N\sigma_x \cdot \sigma_y} \text{ or } r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\,\Sigma(Y-\bar{Y})^2}}$$

Where, $x = X - \bar{X}$, (deviation from Arithmetic Mean (A.M.) of X series)
$y = Y - \bar{Y}$, (deviation from Arithmetic Mean (A.M.) of Y series)

σ_x = Standard Deviation (S.D.) of X series

σ_y = Standard Deviation (S.D.) of Y series

And N = Number of observation.

Example : Calculate Karl Pearson's Coefficient of Correlation. From the following data :

X	42	52	55	60	66	68	65	60	58	34
Y	11	13	18	22	26	40	31	27	24	18

Sol. **Calculation of Karl Pearson's Coefficient of Correlation**

X	Y	$x=(X-\bar{X})$	x^2	$y=(Y-\bar{Y})$	y^2	xy
42	11	– 14	196	– 12	144	168
52	13	– 4	16	– 10	100	40
55	18	– 1	1	– 5	25	5
60	22	4	16	– 1	1	– 4
66	26	10	100	3	9	30
68	40	12	144	17	289	204
65	31	9	81	8	64	72
60	27	4	16	4	16	16
58	24	2	4	1	1	2
34	18	– 22	484	– 5	25	110
$\Sigma X = 560$	$\Sigma Y = 230$		$\Sigma x^2 = 1058$		$\Sigma y^2 = 674$	$\Sigma xy = 643$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{560}{10} = 56;\ \bar{Y} = \frac{\Sigma Y}{N} = \frac{230}{10} = 23$$

$$\text{Coefficient of correlation } (r) = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{643}{\sqrt{1058 \times 674}} = 0.76$$

II. When Deviations are Taken from Assumed Mean

When deviations are taken from assumed mean, then the following formula is used to calculate the value of r.

$$r = \frac{N \times \Sigma d_x d_y - \Sigma d_x \cdot \Sigma d_y}{\sqrt{N \cdot \Sigma d_x^2 = -(\Sigma d_x)^2}\ \sqrt{N \cdot \Sigma d_y^2 - (\Sigma d_y)^2}}$$

Where $d_x = X - A_x$ and $d_y = Y - A_y$, A = Assumed Mean

Example : Find the coefficient of correlation from the following data :

X	1	2	3	4	5	6	7	8	9	10
Y	62	56	48	41	36	28	21	16	12	8

Sol. Let the assumed mean for X is 5, Assumed mean for Y is 40.

X	Y	$d_x=(X-5)$	$d_y=(Y-40)$	d_x^2	d_y^2	$d_x d_y$
1	62	– 4	22	16	484	– 88
2	56	– 3	16	9	256	– 48
3	48	– 2	8	4	64	– 16
4	41	– 1	1	1	1	– 1
5	36	0	– 4	0	16	0
6	28	1	– 12	1	144	– 12
7	21	2	– 19	4	361	– 38
8	16	3	– 24	9	576	– 72
9	12	4	– 28	16	784	– 112
10	8	5	– 32	25	1024	– 160
Total		$d_x=5$	$d_y=-72$	$\Sigma d_x^2=85$	$\Sigma d_y^2=3710$	$\Sigma d_x d_y=-547$

$$r=\frac{N\Sigma d_x d_y-\Sigma d_x \Sigma d_y}{\sqrt{\{N\Sigma d_x^2-(\Sigma d_x)^2\}\times\{N\Sigma d_y^2-(\Sigma d_y)^2\}}}$$

$$=\frac{10(-547)-5(-72)}{\sqrt{10\times 85-(5)^2}\sqrt{10\times 3710-(-72)^2}}=-0.996$$

The variable X and Y are perfectly negatively correlated.

III. When Step Deviations are Taken

To calculate the r following formula is used when step deviation is taken.

$$r=\frac{N\times\Sigma d'_x d'_y-\Sigma d'_x\cdot\Sigma d'_y}{\sqrt{N\cdot\Sigma d'^2_x-(\Sigma d'_x)^2}\sqrt{N\cdot\Sigma d^2_y-(\Sigma d'_y)^2}};$$

Where, $d'_x=\frac{d_x}{i_x}, d'_y=\frac{d_y}{i_y}$

Example : Calculate Coefficient of Correlation between the expenses and saving of any family.

Expenses (₹)	0	15	25	35	75
Saving (₹)	70	55	35	25	15

Sol. Let the assumed mean for X is 25, assumed mean for Y is 45.

X	$d_x=(X-25)$	$d'_x=\frac{d_x}{5}$	$(d'_x)^2$	Y	$d_y=(Y-45)$	$d'_y=\frac{d_y}{5}$	$(d'_y)^2$	$d'_x d'_y$
0	– 25	– 5	25	70	25	5	25	– 25
15	– 10	– 2	4	55	10	2	4	– 4

25	0	0	0	35	– 10	– 2	4	0
35	10	2	4	25	– 20	– 4	16	– 8
75	50	10	100	15	– 30	– 6	36	– 60
$N=5$		$\Sigma d'_x=5$	$\Sigma d'^2_x=133$	$N=5$		$\Sigma d'_y=-5$	$\Sigma d'^2_y=85$	$\Sigma d'_x d'_y=-97$

$$r=\frac{\Sigma d'_x d'_y-\frac{\Sigma d'_x d'_y}{N}}{\sqrt{\Sigma d'^2_x-\frac{(\Sigma d'_x)^2}{N}}\sqrt{\Sigma(d'_y)^2-\frac{(\Sigma d'_y)^2}{N}}}$$

$$=\frac{-97-\left(-\frac{97}{5}\right)}{\sqrt{133-\frac{25}{5}}\sqrt{85-\frac{(-5)^2}{5}}}=\frac{-77.6}{\sqrt{128}\times\sqrt{80}}$$

$$=\frac{-77.6}{\sqrt{10240}}=\frac{-77.6}{101.2}=-0.767$$

IV. When Actual Data is used (Direct Method)

In case of actual data following formula is used to calculate the coefficient r :

$$r=\frac{N\times\Sigma XY-\Sigma X.\Sigma Y}{\sqrt{N.\Sigma X^2-(\Sigma X)^2}\sqrt{N.\Sigma Y^2-(\Sigma Y)^2}}$$

Example : Calculate the Karl Pearson coefficient of correlation for the following data :

X	4	7	11	14	19	15
Y	18	16	17	19	19	21

Sol. **Calculation of Karl Pearson's Coefficient of Correlation**

X	*Y*	X^2	Y^2	*XY*
4	18	16	324	72
7	16	49	256	112
11	17	121	289	187
14	19	196	361	266
19	19	361	361	361
15	21	225	441	315
$\Sigma X=70$	$\Sigma Y=110$	$\Sigma X^2=968$	$\Sigma Y^2=2032$	$\Sigma XY=1313$

Coefficient of Correlation, $r=\frac{N\Sigma XY-(\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2-(\Sigma X)^2}\sqrt{N\Sigma Y^2-(\Sigma Y)^2}}$

$$= \frac{6\times1313-70\times110}{\sqrt{6\times968-(70)^2}\sqrt{6\times2032-(110)^2}}$$

$$= \frac{178}{\sqrt{83536}} = \frac{178}{289} = 0.616$$

V. Variance : Covariance Method

Following formula is used to calculate the value of r when variance and covariance are given :

$$r = \frac{Cov.(X,Y)}{\sqrt{Var.(X)}\sqrt{Var.(Y)}}$$

Where, $$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\,\Sigma(Y-\bar{Y})^2}}$$

Divide Numerator and Denominator by N we have,

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})/N}{\sqrt{\frac{\Sigma(X-\bar{X})^2\,\Sigma(Y-\bar{Y})^2}{N}}}$$

$$= \frac{\Sigma(X-\bar{X})(Y-\bar{Y})/N}{\sqrt{\frac{\Sigma(X-\bar{X})^2}{N}\times\frac{\Sigma(Y-\bar{Y})^2}{N}}}$$

$$= \frac{Cov.(X,Y)}{\sqrt{Var.(X)}\sqrt{Var.(Y)}}$$

Example : If covariance between X and Y variables is 12.5 and the variance of x and y are respectively 16.4 and 13.8, find the coefficient of correlation between them.

Sol. Covariance of $XY = 12.5$, Variance of $X = 16.4$, Variance of $Y = 13.8$

$$r = \frac{Cov.(X,Y)}{\sqrt{Var.(X)}\sqrt{Var.(Y)}} = \frac{12.5}{\sqrt{16.4}\sqrt{13.8}}$$

$$= \frac{12.5}{\sqrt{226.32}} = \frac{12.5}{15.04} = 0.83 = 0.83$$

Q.4. Calculate Karl Pearson's correlation coefficient between X and Y :

X	58	43	41	39	43	46	43	45	41	47	45	44
Y	11	27	31	42	30	28	28	20	19	20	32	30

Sol. **Calculation of Karl Pearson's Correlation Coefficient**

X	$d_x = X-45$	d_x^2	Y	$d_y = Y-27$	d_y^2	$d_x d_y$
58	+ 13	169	11	– 16	256	– 208
43	– 2	4	27	0	0	0
41	– 4	16	31	+ 4	16	– 16

39	– 6	36	42	+ 15	225	– 90
43	– 2	4	30	+ 3	9	– 6
46	+ 1	1	28	+ 1	1	+ 1
43	– 2	4	28	+ 1	1	– 2
45	0	0	20	– 7	49	0
41	– 4	16	19	– 8	64	+ 32
47	+ 2	4	20	– 7	49	– 14
45	0	0	32	+ 5	25	0
44	– 1	1	30	+ 3	9	– 3
—	$\Sigma d_x = -5$	$\Sigma d_x^2 = 255$	—	$\Sigma d_y = -6$	$\Sigma d_y^2 = 704$	$\Sigma d_x d_y = -306$

Formula

$$r = \frac{\Sigma d_x d_y - \frac{(\Sigma d_x)(\Sigma d_y)}{N}}{\sqrt{\left\{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}\right\}}\sqrt{\left\{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}\right\}}}$$

$$= \frac{-306 - \frac{(-5)(-6)}{12}}{\sqrt{\left[255 - \frac{(-5)^2}{12}\right]}\sqrt{\left[704 - \frac{(-6)^2}{12}\right]}}$$

$$= \frac{-306 - \frac{30}{12}}{\sqrt{\left\{255 - \frac{25}{12}\right\}}\sqrt{\left\{704 - \frac{36}{12}\right\}}}$$

$$= \frac{-306 - 2.5}{\sqrt{(255 - 2.1)(704 - 3)}}$$

$$= \frac{-308.5}{\sqrt{252.9 \times 701}} = \frac{-308.5}{\sqrt{1,77,282.9}} = \frac{-308.5}{421.05} = -0.733$$

Q.5. Define coefficient of correlation in continuous series with example.

Ans.

Correlation in Continuous Series

When we have to find coefficient of correlation from a bivariate grouped data the following formula is used.

$$r = \frac{N \times \Sigma f d_x d_y - \Sigma f d_x . \Sigma f d_y}{\sqrt{N.\Sigma f d_x^2 - (\Sigma f d_x)^2}\sqrt{N.\Sigma f d_y^2 - (\Sigma f d_y)^2}}$$

Where, $d_x = X - A_x$ and $d_y = Y - A_y$; A_x = assumed mean of x and A_y = Assumed mean of y.

Example : From the table given below, calculate the Coefficient of Correlation between the ages of husbands and wives :

Age of Wives (*Y*-Series)	Age of Husbands (*X*-Series)					
	20-30	**30-40**	**40-50**	**50-60**	**60-70**	**Total**
15-25	5	9	3	–	–	17
25-35	–	10	25	2	–	37
35-45	–	1	12	2	–	15
45-55	–	–	4	16	5	25
55-65	–	–	–	4	2	6
Total	**5**	**20**	**44**	**24**	**7**	**100**

Sol. Calculation of Coefficient of Correlation between the Age of Husbands and Wives is given as follows :

Age of Wives (*Y*)				**Age of Husbands (*X*)**								
				Age group 20-30	**30-40**	**40-50**	**50-60**	**60-70**				
				m.p. 25	35	45	55	65				
			d_x	−20	−10	0	+10	+20	**Total**			
	d_y			−2	−1	0	+1	+2	***f***	fd_y	fd_y^2	$fd_x d_y$
Age group 15-25	m.p. 20	−20	−2	20 ($fd_x d_y$) 5	18 9	0 3			17	−34	68	38
25-35	30	−10	−1		10 10	0 25	−2 2		37	−37	37	8
35-45	40	0	0		0 1	0 12	0 2		15	0	0	0
45-55	50	+1	+1			0 4	16 16	10 5	25	+25	25	26
55-65	60	+20	+2				8 4	8 2	6	+12	24	16
			Total *f*	5	20	44	24	7	100	Σfd_y =−34	Σfd_y^2 =154	$\Sigma fd_x d_y$ =88
			fd_x^2	−10	−20	0	+24	+14	Σfd_x =8			
			fd_x	20	20	0	24	28	Σfd_x^2 =92			
			$fd_x d_y$	20	28	0	22	18				

$$d_x = \frac{X - A_x}{h}, d_y = \frac{Y - A_y}{h}$$

Substituting the above values in Karl Perarson's formula :

$$r = \frac{\Sigma fd_x d_y - n\left(\frac{\Sigma fd_x}{n}\right)\left(\frac{fd_y}{n}\right)}{\sqrt{\frac{\Sigma fd_x^2}{n} - \left(\frac{\Sigma fd_x}{n}\right)^2}\sqrt{\frac{\Sigma fd_y^2}{n} - \left(\frac{\Sigma fd_y}{n}\right)^2}}$$

We get,

$$r = \frac{88 - 100\left(\frac{8}{100}\right)\left(\frac{-34}{100}\right)}{100\sqrt{\frac{92}{100} - \left(\frac{8}{100}\right)^2}\sqrt{\frac{154}{100} - \left(\frac{-34}{100}\right)^2}}$$

$$= \frac{9072}{\sqrt{9136 \times 14244}} = \frac{9072}{11407.59} = 0.74$$

Q.6. Define Spearman's rank correlation. What are the various conditions for calculating the rank correlation? Explain with examples.

Ans.

Spearman's Rank Correlation

Spearman's Rank Correlation coefficient named after **Charles Edward Spearman** and denoted by *R*, is a technique to find the correlation between the ranks of two series. This technique is used when value of variable cannot be calculated quantitatively. To apply the Spearman's rank correlation technique we need to first arrange the value of variable in serial order. We apply this method when we deal with qualitative characteristics.

For example, when we test the *IQ* of a person with the number of hours spent in front of TV per week, etc. Spearman's rank correlation coefficient (*R*) is defined as,

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

Where, R = Rank Coefficient of Correlation,

D = Difference between Two Ranks $(R_1 - R_2)$,

N = Number of Pair of Observations.

Rank Correlation Coefficient (*R*) always exists between –1 and +1 and also used when the measurements are given for both the series.

Calculation of Rank Correlation

Various conditions for calculating the Rank Correlation are as follows :

1. **When Ranks are given :** When the actual ranks (say R_1 and R_2) are given, then following steps are used to calculate the rank.
   (i) Calculate the difference of two ranks $(R_1 - R_2)$ and assign it to D, *i.e.*, $D = (R_1 - R_2)$.
   (ii) Then calculate the square of D and summate the D as ΣD^2.
   (iii) In the last, put all the calculated values in the formula.

Example : Following are the rank obtained by 10 students in two subjects, English and History. To what extent the knowledge of the students in the two subjects is related?

English	1	2	3	4	5	6	7	8	9	10
History	2	4	1	5	3	10	9	6	7	8

Sol. **Calculation of Rank Correlation Coefficient**

Rank of English (R_1)	Rank of History (R_2)	$D=(R_1-R_2)$	D^2
1	2	–1	1
2	4	–2	4
3	1	+2	4
4	5	–1	1
5	3	+2	4
6	10	–4	16
7	9	–2	4
8	6	+2	4
9	7	+2	4
10	8	+2	4
			$\Sigma D^2=46$

$$\text{Rank correlation coefficient } (R)=1-\frac{6\Sigma D^2}{N(N^2-1)}=1-\frac{6\times 46}{10(10^2-1}$$

$$=1-\frac{276}{10(100-1)}=1-\frac{276}{990}=1-0.28=+0.72$$

2. **When Ranks are not given :** If no rank is given, then we first calculate the rank of the given data (actual data) :
   (i) To assign the rank we take the lowest number of rank as 1, the next lowest number a rank is 2, and so on. As well as assign the rank lighest number.
   (ii) Calculate the difference of two ranks (R_1-R_2) and assign it to D, *i.e.*, $D=(R_1-R_2)$.
   (iii) Then calculate the square of D and summate the D as ΣD^2.
   (iv) In the last, put all the calculated values in the formula.

Example : Calculate coefficient of correlation by rank method :

X	54	58	85	75	65	90	80	50
Y	120	134	150	115	110	140	142	100

Sol. **Calculation of Rank Correlation Coefficient**

X	Rank (R_1)	Y	Rank (R_2)	$D=(R_1-R_2)$	D^2
54	2	120	4	– 2	4
58	3	134	5	– 2	4
85	7	150	8	– 1	1

75	5	115	3	+2	4
65	4	110	2	+2	4
90	8	140	6	+2	4
80	6	142	7	– 1	1
50	1	100	1	0	0
					$\Sigma D^2 = 22$

Coefficient of Rank Correlation

$$R = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)} = 1 - \frac{6 \times 22}{8(64 - 1)} = 1 - \frac{132}{504} = \frac{504 - 132}{504} = \frac{372}{504} = 0.738$$

3. **Tied Ranks/Equal or Repeated Ranks :** When values are repeated, then ranks assigned to these values are similar. In this case, calculate the average of rank and assign it to the repeated values. *For example*, if two persons are positioned on the fifth place, they each of them are given the rank $\frac{5+6}{2} = 5.5.$, which is common rank to be assigned; and the next rank will be 7. In case if three are ranked equal at the fifth place, they all are given the rank $\frac{5+6+7}{3} = 6$, which is the common rank to be assigned to each; and the next rank will be 8, in this case. In this (Tied Ranks/Equal or Repeated Ranks) case, to perform calculation following formula is used :

$$R = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2)\ldots\right)}{N^3 - N}$$

Where, m = Number of items having common ranks.

Example : Obtain the rank correlation coefficient for the following data :

X	68	64	74	50	64	80	74	40	55	64
Y	62	58	67	45	81	60	67	48	50	70

Sol.

Calculation of Rank Correlation

X	Rank (R_1)	Y	Rank (R_2)	$D = (R_1 - R_2)$	D^2
68	7	62	6	1	1
64	5	58	4	1	1
74	8.5	67	7.5	1	1
50	2	45	1	1	1
64	5	81	10	– 5	25
80	10	60	5	5	25
74	8.5	67	7.5	1	1
40	1	48	2	– 1	1
55	3	50	3	0	0
64	5	70	9	– 4	16
					$\Sigma D^2 = 72$

In series X the value 64 occurs thrice $(m_1 = 3)$, *i.e.*, at 4th, 5th and 6th rank. Hence, all three values are given the average rank, *i.e.*, $\frac{4+5+6}{3}$th = 5th rank. Also in series X the value 74 occurs twice $(m_2 = 2)$, *i.e.*, at 8th and 9th rank. Hence, values are given the average rank, *i.e.*, $\frac{8+9}{2}$th = 8.5th rank.

While, in series Y, the value 67 occurs both at the 7th and 8th rank, $(m_3 = 2)$. Hence, both are given the average rank, this is, $\frac{7+8}{2}$th = 7.5th rank.

Coefficient of Rank Correlation

$$(R) = 1 - \frac{6\left[\Sigma D^2 + \frac{m_1^3 - m_1}{12} + \frac{m_2^3 - m_2}{12} + \frac{m_3^3 - m_3}{12}\right]}{N(N^2 - 1)}$$

$$= 1 - \frac{6\left[72 + \frac{3^3 - 3}{12} + \frac{2^3 - 2}{12} + \frac{2^3 - 2}{12}\right]}{10(10^2 - 1)} = 1 - \frac{6[72 + 2 + 0.5 + 0.5]}{10(10^2 - 1)} = 1 - 0.4545$$

$$R = 0.5455$$

Q.7. Find the Spearman's rank coefficient of correlation between sales and profits of the following 10 firms :

Firms	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Sales	50	50	55	60	65	65	65	60	60	50
Profit	11	13	14	16	16	15	15	14	13	13

Sol. Calculation of Spearman's Rank Coefficient of Correlation

Firms	**Sales, X**	**Profit, Y**	**Ranks**		**Difference of Rank, D**	**D^2**
			X	**Y**		
A	50	11	9	10	– 1	1
B	50	13	9	8	1	1
C	55	14	7	5.5	1.5	2.25
D	60	16	5	1.5	3.5	12.25
E	65	16	2	1.5	0.5	0.25
F	65	15	2	3.5	– 1.5	2.25
G	65	15	2	3.5	– 1.5	2.25
H	60	14	5	5.5	– 0.5	0.25
I	60	13	5	8	– 3	9
J	50	13	9	8	1	1
Total						**31.50**

Here in series X, 50, 60 and 65 are repeated 3 times each and in series Y, 14, 15 and 16 are repeated 3 times each and 13 is repeated 3 times. Hence,

Let
$$T = \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3) + \frac{1}{12}(m_4^3 - m_4) + \frac{1}{12}(m_5^3 - m_5) + \frac{1}{12}(m_6^3 - m_6) + \frac{1}{12}(m_7^3 - m_7)$$

$$T = \frac{1}{12}(3^3 - 3) + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)$$

$$= \frac{4}{12}(27 - 3) + \frac{3}{12}(8 - 2)$$

$$= \frac{24}{3} + \frac{6}{4} = \frac{96 + 18}{12} = \frac{114}{12} = 9.5$$

Thus,
$$r_s = 1 - \frac{6(\Sigma D^2 - T)}{N(N^2 - 1)}$$

$\therefore$
$$r_s = 1 - \frac{6(31.5 + 9.5)}{10(10^2 - 1)}$$

$$= 1 - \frac{6 \times 41}{10 \times 99} = 1 - \frac{246}{990} = 1 - 0.25 = +0.75$$

Q.8. Find the Karl Pearson's coefficient of correlation on the basis of data ahead related to age of husbands and their wives taking deviations from actual means 30 and 25 respectively :

Couple	Age of Husband (in years)	Age of Wife (in years)
1	23	18
2	?	20
3	28	22
4	28	27
5	29	?
6	30	29
7	31	27
8	?	29
9	35	28
10	36	29

Assume that age of eight husbands is 6 years more than second husband.

Sol. Suppose the age of husband of second couple $= a$ years

Hence, the age of husband of 8th couple $= (a + 6)$ years

$$\bar{X} = \frac{\Sigma X}{n}$$

$\Rightarrow \quad 30 = \dfrac{23 + a + 28 + 28 + 29 + 30 + 31 + (a + 6) + 35 + 36}{10}$

$\Rightarrow \quad 30 \times 10 = 2a + 246$

$\Rightarrow \quad 2a = 300 - 246 = 54 \qquad \therefore a = \dfrac{54}{2} = 27$

Then $\quad \bar{Y} = \dfrac{\Sigma Y}{n}$

$\Rightarrow \quad 25 = \dfrac{18 + 20 + 22 + 27 + b + 29 + 27 + 29 + 28 + 29}{10}$

$\Rightarrow \quad 25 \times 10 = b + 229$

$\Rightarrow \quad b = 250 - 229 = 21$

Thus, substituting the values of unknowns, the table for computation of Karl Pearson coefficient of correlation is as follows :

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
23	18	– 7	– 7	49	49	49
27	20	– 3	– 5	9	25	15
28	22	– 2	– 3	4	9	6
28	27	– 2	2	4	4	– 4
29	21	– 1	– 4	1	16	4
30	29	0	4	0	16	0
31	27	1	2	1	4	2
33	29	3	4	9	16	12
35	28	5	3	25	9	15
36	29	6	4	36	16	24
300	**250**	**0**	**0**	**138**	**164**	**123**

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$= \frac{123}{\sqrt{138 \times 164}} + \frac{123}{\sqrt{22632}} = \frac{123}{150.44} = 0.8176 \simeq 0.82$$

Q.9. Discuss about the measurement of correlation by Scatter diagram or dotogram.

Ans.

Measurement of Correlation by Scatter Diagram or Dotogram

Let the two variables be X and Y for which a set of n pairs of values is known. To get some idea whether there is any relationship present in two variables we plot these values on a chart known as *Scatter Diagram*.

This diagram allows a visual examination of the extent to which the variables are related. By means of scatter diagram one can quickly judge the type of correlation between the two variables.

Construction

The diagram is prepared by measuring *X* on the horizontal axis and *Y* on the vertical axis and plotting a point for each pair of observations of *X* and *Y*. The diagram is in the shape of dots or points.

Conclusion

Case 1. If the points take the shape of a line, then their a linear correlation.

(i) If the line rises from left bottom going up toward the right top, then there is a perfect positive correlation between the two variables.

(ii) If the line moves in the reverse way, then there is perfect negative correlation between the two variables. This is shown in Fig. 1 and Fig. 2.

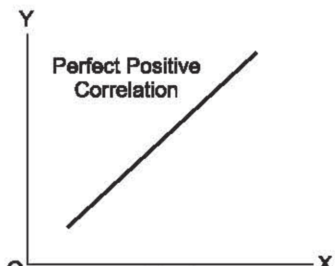


Fig 1

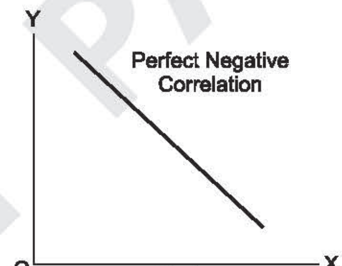


Fig 2

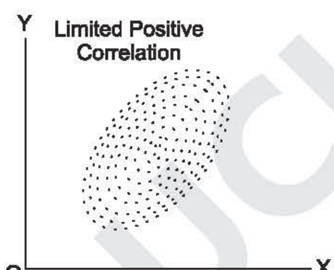


Fig 3

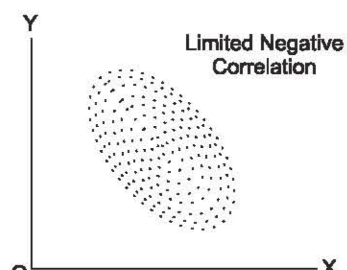


Fig 4

Case 2. If the points show some trend either upward (Fig. 3) or downward (Fig. 4), the two variables are correlated.

(i) If the trend of the points is upward rising from the left bottom and going up towards right top (Fig. 3), correlation is positive.

(ii) On the other hand, if the tendency is reverse so that the points show a downward trend from the left top to the right bottom (Fig. 4), correlation is negative.

Case 3. If the plotted points do not show any trend the two variables have no correlation (Fig. 5).

Case 4. If the tendency of the dots is to concentrate round a circle (Fig. 6) or any known form other than a straight line, the correlation is non-linear.

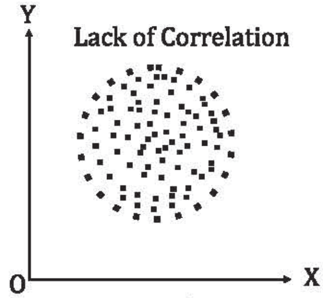


Fig. 5

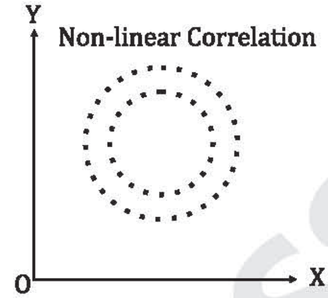


Fig. 6

Example : Draw a scatter diagram and study correlation from the given data of a industrial city :

Sale ('000 in ₹)	125	170	175	180	190	210	250	300	320	400
Profit ('000 in ₹)	20	29	32	35	34	41	55	60	64	70

Sol.

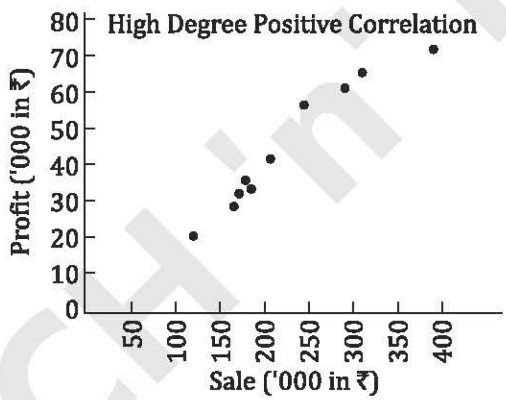


Fig. 7

❑

UNIT-IV

Index Number

SECTION-A VERY SHORT ANSWER TYPE QUESTIONS

Q.1. Write the importance of Index Number.

Ans. Importance of Index Numbers :

1. Trend in production can be studied by using indices and it facilitates the government to make policies for the industries.
2. Rise or fall in foreign trade can be known by the use of such indices.
3. Indices like cost of living index number can be used to interpret the changes in cost of living in a country.
4. Interpretation of changes in internal price level can be done using index numbers.

Q.2. Write any two definitions of index number.

Ans. According to *Maslow*, "It is a numerical value characterizing the change in complex economic phenomena over a period of time or space."

According to *John Riffin*, "Index numbers are used to measure changes over time in magnitudes which are not capable of direct measurement."

Q.3. What are the limitations of index numbers.

Ans. Limitations of Index Numbers are below :

1. Manipulation in index numbers can be done to serve particular purposes. High profit year can be made base period thereby showing the profit in current year as less.
2. Tastes and preferences of customers change with the changing time so quality of item can change. Using new quality and obsolescence of other can make the comparison wrong over long period.
3. Different index numbers are constructed by different methods. Thus, it can give different results in different cases.
4. Samples of items and quantities are used to calculate index numbers. Sampling is generally biased thus introducing errors in the calculation. So efforts have to be made to minimise such errors.

Q.4. Define the construction of Price Index Number.

Ans. Price index number is used to compare a group of products at a given time and place. Base period price or place is used to compare the given price. To calculate the price index number following formula is used :

$$P_{01} = \frac{p_1}{p_0} \times 100$$

Where, P_{01} = Price index number of the current year on the basis of the base year's price.

Let's take an example where price of rice in India in 2014 was ₹850 per quintal and in 2013 it was ₹750 per quintal. Then the price relative for 2014 is :

$$\frac{85}{75}\times 100 = 133.33\%$$

Note : 2013 taken as base year.

Thus, Price relative is the percentage of price of the commodity in a specific year and the price of the same commodity in a given year.

Q.5. Distinguish between fixed base and chain base year.

Ans. Major differences between fixed base and chain base year are discussed in the following table :

S.No.	Fixed Base Year	Chain Base Year
1.	The fixed base index becomes out moded on account of the changes in habits and tastes of persons, etc.	This is not the case with chain base index.
2.	Fixed base method is good for studying long time changes.	Chain base method is not good for studying the long time changes.
3.	The inclusion and exclusion of commodities is not easy.	New commodities can be included and old one can be debted easily.
4.	Selection of a normal year is a problem.	There is no problem of selection.
5.	It remains same for all the years.	It changes with every year.
6.	It is easy to calculate.	It involves lengthy calculations.

Q.6. Explain the process of construction of quality index number?

Ans. To find out the variation in level of quantities of items (produced, consumed or distributed) in a year with reference to base year quantity, this index number is used. Following is the simplest formula used to calculate the quantity index number :

$$Q_{01} = \frac{q_1}{q_0}\times 100$$

Where, Q_{01} = Quantity index number of the current year on the basis of the base year's quantity.

Q.7. Why Fisher's Index Number is ideal?

Ans. Fisher's Index No. is considered an 'ideal index' on account of its following characteristics :

1. **Use of Variable Weights :** In Fisher's Index different weights are used in base year and current year on the basis of quantities in these years. It is correct also because quantity of consumption may not remain the same every year.
2. **Based on Geometric Mean :** Fisher's index is based on geometric mean, which is supposed to be the best average for the construction of index numbers.
3. **Satisfaction of Reversal Tests :** It satisfies both the tests *i.e.,* time reversal as well as the factor reversal tests.

Q.8. Write the methods of measurement of trend.

Ans. The following are the four important methods which are used in estimating the trend :

(I) Free-hand Curve Method, (II) Semi-average Method, (III) Moving-average Method, (IV) Method of Least Squares.

Q.9. Write the advantages of Fisher's 'Ideal' Index.

Ans. Following are the main advantages of ideal method :

1. It includes the effects of both current and base years.
2. All the data such as p_1, p_0, q_1, and q_0 are used to this method.
3. It does not show an upward or downward bias because it takes both current and base years quantities as weight.
4. It fulfils the requirements of the following tests :
   (i) Unit Test
   (ii) Time Reversal Test
   (iii) Factor Reversal Test.
5. This method uses the most suitable average (geometric mean) to calculate the index number.

Q.10. What are the limitations of Fisher's formula?

Ans. Though Fisher's formula is an ideal formula of index number, yet it has certain limitations that the data of quantity of each year also should be available, otherwise this index number cannot be calculated. Moreover, this index number does not indicate the change in price only but it gives information about the mixed change in price and quantity. *Boddington* has remarked that "Unfortunately, while this formula apparently meets most of the mathematical requirements of a perfect index formula, it is objected to, on the score that it is not clear what it measures, i.e., the result combines both price and volume changes, when usually we want the one to be separated from the other."

Q.11. Define Wholesale Price Index (WPI).

Ans. Wholesale Price Index (WPIs) is taken in order to indicate the average price variation in the goods soid in huge quantities and these are the collection of various indicates that can be used to determine the growth and inflation in the country. This is the index that is commonly used to determine and track the variation in the prices of the goods before the retail level. This includes the wholesale prices where the base year quantities and current year quantities are not the same.

SECTION-B SHORT ANSWER TYPE QUESTIONS

Q.1. What is type of index number?

Ans. Fundamentally index numbers can be categorised into following types :

1. **Quantity Index Number :** In a given time period, quantity index number helps in calculating and comparing the physical quantity of goods (sold, purchased and produced by an organisation).
2. **Price Index Number :** These are the simplest and most commonly used index numbers. They are used to measure the variation in the price of commodities consumed in a given period of time. To measure the variation or change in price of commodities it uses the base period as a reference.

Following are the types of price index numbers :

(i) **Wholesale Price Index Number :** It is used to find out the variation in the general price level of a product or service.

(ii) **Retail Price Index Number :** It is used to find the variation in retail price of a product or service (bought and sold in the retail market).

3. **Value Index Number :** These are used to find out of the variation in the value of commodities and collection of commodities consumed or purchased in a time period with reference to the base period.
4. **Simple Index Number :** Index numbers for individual commodities are known as simple index numbers.
5. **Cost of Living Index Number :** It is also known as 'consumer price index number'. It is used to compare the average variation in expenses and consumption of the commodity from one time period to another. It is also used to measure the variation of an individual class of customers.
6. **Aggregate (or Composite) Index Number :** Creation of index number for a group of commodities is termed as composite (aggregate) index numbers.

Q.2. Write the merits and limitations of Least Squares Method.

Ans.

Merits of Least Squares Method

1. **Line of best fit :** The trend line obtained by this method is the line of best fit, because the sum of positive and negative deviations of original data from this line is zero and sum of squares of deviations is minimum.
2. **Forecasting :** The equation of a straight line based on least squares method establishes a functional relationship in between x and y series and through this relationship forecasting can easily be made for future values.
3. **Trend for the entire period :** In moving average method trend values cannot be found out for the entire period while the method of least squares gives the trend values for the entire time period.
4. **Completely objective :** This method is completely objective, in which trend values are calculated on the basis of well-defined mathematical principles and formulae. There is no possibility of personal bias in this method.
5. **Calculation of change rate :** If data are on yearly basis then annual growth rate or decline rate can be obtained by this method.

Limitations of Least Squares Method

1. **Limitations of predictions :** Prediction in this method is based on long-term trend and the impact of seasonal, cyclical or irregular variations is ignored.
2. **Lack of flexibility :** If even a single value is added or deleted to the series, it becomes necessary to do all the computations again and generally the equation of trend also does change.
3. **Wrong selection of equation :** If the selection of trend equation (linear, parabolic or some other type) is not proper, it may lead to fallacious results.

4. **Tedious and complicated :** This method is tedious and complicated from the point of mathematical calculations.

Q.3. Construct an index for the year 2014 taking 2013 as base year from the data given below :

Commodity	Price in 2013 (₹)	Price in 2014 (₹)
P	70	80
Q	80	60
R	40	35
S	50	50

Sol.

Construction of Price Index

Commodity	Price 2013 (P_0)	Price 2014 (P_1)
P	70	80
Q	80	60
R	40	35
S	50	50
	$\Sigma P_0 = 240$	$\Sigma P_1 = 225$

By simple aggregative method,

$$P_{01} = \frac{\Sigma P_1}{\Sigma P_0} \times 100 = \frac{240}{225} \times 100 = 106.66$$

It means there is a net increase of 6.66% (=106.66 − 100) in prices in 2014 as compared to prices in 2013.

Q.4. Using Fisher's formula, find the price index number from the following data :

Commodity	Unit	Base Year		Current Year	
		Price (₹)	Value (₹)	Quantity	Value (₹)
P	kg	12.5	125	12	156
Q	kg	14	112	9	135
R	Metre	11	88	9	108
S	kg	13	78	6	90

Sol.

Calculation of Price Index Number

Commodity	p_0	q_0	p_1	q_1	p_0q_0	p_1q_0	p_1q_1	p_0q_1
P	12.5	10	13	12	125	130	156	150
Q	14	8	15	9	112	120	135	126
R	11	8	12	9	88	96	108	99
S	13	6	15	6	78	90	90	78
					403	**436**	**489**	**453**

Fisher's Index Number :

$$\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{436}{403} \times \frac{480}{453}} \times 100 = \sqrt{1.082 \times 1.079} \times 100 = 108$$

Q.5. Calculate Fisher's Index Number from the following information :

Items	Price	Base Year	Current Year	
		Expenditure (p_0q_0)	Price	Expenditure
P	5	120	8	96
Q	6	360	12	600
R	9	450	15	585
S	14	406	7	140

Sol. In this problem we are given the expenditure (e) and the prices (p) per unit for different commodities.

We have, Expenditure = Price × Quantity

$$\Rightarrow \quad \text{Quantity} = \frac{\text{Expenditure}}{\text{Price}} \quad \Rightarrow \quad q = \frac{e}{p} \qquad ...(1)$$

From equation (1) we first obtain the quantities consumed for the base year and the current year as given in the following table :

Computation of Fisher's Ideal Index Number

Commodities	p_0	q_0	p_1	q_1	p_0q_0	p_1q_1	p_1q_0	p_0q_1
P	5	24	8	12	120	96	192	60
Q	6	60	12	50	360	600	720	300
R	9	50	15	39	450	585	750	351
S	14	29	7	20	406	140	203	280
					Σp_0q_0 =1336	**Σp_1q_1 =1421**	**Σp_1q_0 =1865**	**Σp_0q_1 =991**

Hence, Fisher's Ideal Price Index is given by

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{1865 \times 1421}{1336 \times 991}} \times 100 \sqrt{\frac{2650165}{1323976}} \times 100 = \sqrt{2} \times 100 = 141.4$$

Q.6. Construct Fisher's ideal Index from the following data and show that it satisfies time reversal and factor reversal tests :

Commodity	2013		2014	
	Price	Value	Price	Value
P	18	140	22	170
Q	22	90	28	120
R	12	150	15	180
S	26	90	32	110
T	38	440	60	500

Sol. **Construction of Fisher's Ideal Index**

Commodities	Base Year		Current Year		p_0q_0	p_1q_0	p_1q_1	p_0q_1
	p_0	q_0	p_1	q_1				
P	18	140	22	170	2520	3080	3740	3060
Q	22	90	28	120	1980	2520	3360	2640
R	12	150	15	180	1800	2250	2700	2160
S	26	90	32	110	2340	2880	3520	2860
T	38	440	60	500	16720	26400	30000	19000
Total					$\Sigma p_0q_0 = 25360$	$\Sigma p_1q_0 = 37130$	$\Sigma p_1q_1 = 43320$	$\Sigma p_0q_1 = 29720$

Fisher's Ideal Index
$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{37130}{25360} \times \frac{43320}{29720}} \times 100$$
$$= \sqrt{\frac{1608471600}{753699200}} \times 100 = 1.461 \times 100 = 146.1$$

Time-reversal test is satisfied if
$$P_{01} \times P_{10} = 1 = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times \frac{\Sigma p_0q_1}{\Sigma p_1q_1} \times \frac{\Sigma p_0q_0}{\Sigma p_1q_0}}$$

Substituting the values from the table
$$P_{01} \times P_{10} = \sqrt{\frac{37130}{25360} \times \frac{43320}{29720} \times \frac{29720}{43320} \times \frac{25360}{37130}} = \sqrt{1} = 1$$

∴ Time-reversal test is satisfied by Fisher's Ideal formula.

Factor-reversal test is satisfied if
$$P_{01} \times Q_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times \sqrt{\frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times \frac{\Sigma q_1p_1}{\Sigma q_0p_1}}$$

Substituting the values
$$P_{01} \times Q_{01} = \sqrt{\frac{37130}{25360} \times \frac{43320}{29720} \times \frac{29720}{25360} \times \frac{43320}{37130}} = \frac{43320}{25360}$$

Now $\frac{\Sigma p_1q_1}{\Sigma p_0q_0}$ is also equal to $\frac{43320}{25360}$.

∴ Fisher's Ideal Index satisfies factor-reversal test also.

Q.7. Compute Fisher's Ideal Index and show that it satisfies the reversibility tests :

Items	Base Year		Current Year	
	Price	Quantity	Price	Quantity
P	10	12	12	15
Q	7	14	5	20

R	5	24	9	30
S	16	5	14	10

Sol. **Construction of Fisher's Ideal Index**

Commodities	Base Year		Current Year		p_0q_0	p_1q_0	p_1q_1	p_0q_1
	p_0	q_0	p_1	q_1				
P	10	12	12	15	120	144	180	150
Q	7	14	5	20	98	70	100	140
R	5	24	9	30	120	216	270	150
S	16	5	14	10	80	70	140	160
Total					$\Sigma p_0q_0 = 418$	$\Sigma p_1q_0 = 500$	$\Sigma p_1q_1 = 690$	$\Sigma p_0q_1 = 600$

Fisher's Ideal Index
$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{500}{418} \times \frac{690}{600}} \times 100$$
$$= \sqrt{\frac{3450}{2508}} \times 100 = 117.3$$

Time-reversal test is satisfied if
$$p_{01} \times P_{10} = 1 = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times \frac{\Sigma p_0q_1}{\Sigma p_1q_1} \times \frac{\Sigma p_0q_0}{\Sigma p_1q_0}}$$

Substituting the values from the table
$$P_{01} \times P_{10} = \sqrt{\frac{500}{418} \times \frac{690}{600} \times \frac{600}{690} \times \frac{418}{500}} = \sqrt{1} = 1$$

∴ Time-reversal test is satisfied by Fisher's Ideal formula

Factor-reversal test is satisfied if
$$P_{01} \times Q_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$
$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times \frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times \frac{\Sigma q_1p_1}{\Sigma q_0p_1}}$$

Substituting the values
$$P_{01} \times Q_{01} = \sqrt{\frac{500}{418} \times \frac{690}{600} \times \frac{600}{418} \times \frac{690}{500}} = \frac{690}{418}$$

Now $\frac{\Sigma p_1q_1}{\Sigma p_0q_0}$ is also equal to $\frac{690}{418}$

∴ Fisher's Ideal Index satisfies factor-reversal test also.

Q.8. From the following average prices (₹ per unit) of three commodities, find the index numbers chained to 1988 using chain base method.

Commodities	1988	1989	1990	1991	1992
A	8	10	12	15	12
B	10	12	15	18	20
C	6	9	12	15	18

Sol. **Construction of Chain Base Index Number**

Commodities	1988		1989		1990		1991		1992	
	p_0	*LR*	p_1	*LR*	p_2	*LR*	p_3	*LR*	p_4	*LR*
X	8	100	10	$\frac{10}{8}\times100$ =125	12	$\frac{12}{10}\times100$ =120	15	$\frac{15}{12}\times100$ =125	12	$\frac{12}{15}\times100=80$
Y	10	100	12	$\frac{12}{10}\times100$ =120	15	$\frac{15}{12}\times100$ =125	18	$\frac{18}{15}\times100$ =120	20	$\frac{20}{18}\times100$ =111.1
Z	6	100	9	$\frac{9}{6}\times100$ =150	12	$\frac{12}{9}\times100$ =133.3	15	$\frac{15}{12}\times100$ =125	18	$\frac{18}{15}\times100=120$
Total of *LR*	–	300	–	395	–	378.3	–	370	–	311.1
ALR		100		131.7		126.1		123.3		103.7
Chain Indices based to 1988		100		$\frac{100\times131.7}{100}$ =131.7		$\frac{131.7\times126.1}{100}$ =166.1		$\frac{166.1\times123.2}{100}$ =204.8		$\frac{204.8\times103.7}{100}=212.4$

Q.9. From the following data, calculate 4 year moving average :

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Production	30	32	22	24	12	26	20	14	16	18

Sol. **Calculation of 4 Yearly Moving Averages of the Data**

(1) Year	(2) Production	(3) 4 Year Moving Total	(4) 4 Year moving Average (3) ÷ 4	(5) 2 items moving total of column (4)	(6) 4 year centered moving average (5) ÷ 2
2001	30				
2002	32				
		108	27		
2003	22			49.5	24.75
		90	22.5		
2005	24			43.5	21.75
		84	21		

2004	12			41.5	20.75
		82	20.5		
2006	26			38.5	19.25
		72	18		
2007	20			37	18.5
		76	19		
2008	14			36	18
		68	17		
2009	16	–	–		
2010	18	–	–		

Working Notes : 1st 4 year moving total $= 30+32+22+24 = 108$,

2nd 4 year moving total $= 32+22+24+12 = 90$

1st 4 year moving average $= \frac{108}{4} = 27$ and so on, 1st 2 item moving total

$= 27 + 22.5 = 49.5$ and so on

1st 4 year centered moving average $= 49.5/2 = 24.75$

SECTION-C LONG ANSWER TYPE QUESTIONS

Q.1. Define Index Numbers and mention its importance.

Ans.

Meaning and Definition of Index Numbers

Index number indicates the level of certain phenomenon at a given time (or place) in comparison with the level of the same phenomenon at some standard time (or place) or other characteristics.

In its simplest form an index number is a percentge relative that compares economic measures in a given period to those same measures at a fixed time period. *For example*, changes in income, wages, prices, production, expenditure, exports, imports over a period of time.

Some definitions given by various persons are as follows :

(i) "An index number is a statistical measure designed to show changes in a variable or group of related variables with respect to time, geographical location or other characteristics." *—Murray Spiegal*

(ii) "Index numbers are a series of numbers by which changes in the magnitude of a phenomenon are measured from time to time or from place to place." *—Horace Secrist*

(iii) "An index number is special type of average that provides a measurement of relative change from time to time or from place to place." *—Wessell, Willet and Simone*

(iv) "Index numbers are devices for measuring differences in the magnitude of a group of related variables." *—Croxton and Cowden*

Importance of Index Numbers

The primary purpose of index numbers is to measure relative temporal or cross-sectional changes in a variable or a group of related variables which are not capable of being directly measured. The greatest purpose of index numbers has been to measure and compare changes in prices and purchasing power of money which have received great attention of economists for many years.

The mechanism of Index Number, though originated in 1764 by Italian Statistician *Carli* who used it for measuring price changes, has assumed importance. Now-a-days it is not only used for measuring price changes alone. The factors like, wages, employment, production, trade, demand, supply, business condition, industrial activity, economic condition, financial problems, etc. are also studied through this statistical device.

According to *M.M. Blair*, "Index numbers are the signs and guide posts along the business high way that indicate to the businessman how he should derive or manage his affairs." As a Barometer measure the pressure of atmosphere or gases so the index numbers measure the pressure of economic behaviour, and thus the index numbers are called *Economic Barometers*.

In brief the following are the main uses of index numbers by which the importance (significance) of index numbers can be best appreciated :

1. **Simplify data :** Index numbers provide a means of simplifying data that is of reducing complex forms of measurement to simple numbers which reflect the changes in the variable.
2. **Measure purchasing power of money :** Index numbers (wholesale price index number) measure the purchasing power of money. The purchasing power of money indicates the economic set-up of a country.
3. **Comparative study :** With the help of the index numbers measurement taken at various points of time or space are readily compared in respect to relative change. They combine units of dissimilar nature into one meaningful value.
4. **Useful in deflating :** Index numbers are very useful in deflating *i.e.,* Index numbers help in finding out the real income of the people.
5. **Reveal trends and tendencies :** Index numbers reveal trends and tendencies and are helpful in forecasting the future economic trends. In the words of *G. Simpson* and *F. Kafka*, "Index numbers are today one of the most widely used statistical devices......, they are used to take the pulse of the economy and they have come to be used as indicator of inflationary or deflationary tendencies."
6. **National income :** The index numbers of business conditions measure the changes in the general economic activity of a country and give an idea about the fluctuation in the real national income.
7. **Provide guidelines to Economic policy and informulating decisions :** Index numbers are used to determine dearness allowance, bonus, etc. to be given to employees to bear the increased cost of living. The index numbers of production and sales, etc. of various commodities are used to study the progress of general industrialization of a country. Investment index numbers are obtained to have some idea about the stock market.
8. **Essential control by government :** On the basis of the index numbers, Government can control the prices, etc. The decisions regarding investment, income, employment, trade, consumption, etc. are taken with the help of index numbers.
9. **Universal utility :** Index numbers are useful in various fields *e.g.,* Sociology, Economics, Commerce, Psychology, Medical Science, Educational Research Organizations, Life Insurance Corporation, etc.

10. **Comparative study is made possible :** Index numbers are used to compare the total variations in the prices of different commodities in which the units of measurements are different.
11. **Change in cost of living :** Index numbers (consumer price index number) measure the changes in the cost of living of different groups of people.

Q.2. Discuss the main characteristics and limitations of Index Number.

Ans.

Main Characteristics of Index Numbers

On the basis of study and analysis of the various definitions of index number, the following characteristics of index numbers are apparent. These points are noteworthy for a proper understanding of the term "index number" :

1. **Relative measure :** Index numbers measure changes which are not capable of direct measurement.
2. **Specialised averages :** Index number represents a special case of an average, in general, a weighted average. It is a special type of average, because whereas in a simple average, the data are homogeneous having the same unit of measurement, they average variables having different units of measurement.
3. **Expressed in percentage :** Index numbers are expressed in terms of percentages so as to show the extent of relative change where the value of base is assumed to be 100 but the sign of percentage (%) is not used.
4. **Basis of comparison :** Index numbers by their very nature are comparative. They compare changes overtime or between places or like categories.
5. **Expressed in numbers :** Index number represents the relative changes such as production is increased, prices are down, etc. in numbers.

Limitations or Defects of Index Numbers

Index numbers being even the most widely used statistical devices have their own limitations as given below :

1. **International comparison is not possible :** The base period may be different for different countries so their comparison is not possible.
2. **Fallacious conclusions :** Since index numbers are based on the selection of representative commodities, they may give fallacious conclusions.
3. **Time factor :** The tastes, habits and customs of people change in course of time so comparison of new index numbers is not suitable for old index numbers.
4. **Different purposes :** Index numbers which are useful for one purpose, may not be useful for another.
5. **No attention towards qualitative improvement of the items :** Index numbers do not reveal the changes in the qualitative characteristics of the items.
6. **Limitations of averages :** Since index numbers are specialised averages limitations of these averages are inherent.
7. **Approximate indicator :** Index numbers are only approximate indicators of the relative changes. They are true on the average. They give the direction of change.
8. **Other limitations :** Their may be error in each stage of the construction of index numbers, *e.g.,* selection of items, selection of base period, selection of weights, selection of averages, selection of formula, retail prices, etc.

Q.3. Explain the various problems in the construction of Index number.

Ans.

Problems in the Construction of Index Numbers

To follow the steps (given below) involved in the construction of index numbers many problems are encountered which are to be considered carefully :

1. **Purpose of index number :** The steps which are taken in the construction of index numbers generally depends on the purpose of the index number; so the purpose of an index number must be defined clearly and precisely. According to *A. L. Toole*, "Without a precise formulation of the purpose it may be impossible to know how to carry out properly the steps involved in construction of the index." For instance,

   The purpose of the general index number of wholesale price index number is to know the general price level. The purpose of the consumer price index number is to give an idea of the effect of change in ratail prices on the cost of living of classes of people.

2. **Selection of base period :** The base period of an index number is the period of time against which the comparisons are made. There are three types of based period :

   (i) Fixed Base (a single period)

   (ii) Fixed Base (an average of selected periods)

   (iii) Chain Base.

   While selecting the base a decision has to be made as to whether we have a fixed base or chain base. In a fixed base (a single period) :

   (a) The base period must be a *normal period.* By normal period we mean that period which is free from all sorts of abnormalities or random causes such as strikes of labourers, wars, financial crisis, floods, famines, earthquakes, etc.

   (b) The base period should not be too distant in the past.

   (c) The base period should be a period for which reliable figures are available.

   When it is difficult to choose just one single period as the normal, an average of several periods is usually a better base.

   If comparisons are desired from year, to year a system of chain base is used. In chain base method, there is no fixed base for comparing the values of subsequent years; but the value of each year is compared with the value of the preceding year.

3. **Selection of commodities :**

   (a) First problem in the selection of commodities because it is not feasible to include all commodities is their number. The purpose of the index number shall help in deciding the number of commodities. In a general index number, we may consider a large number of commodities.

   (b) Which commodities are to be included? A careful selection of the commodities must be made in such a way that :

   (i) it represents the real habits, tastes and the customs of the people;

   (ii) it must be easily recognizable and describable;

   (iii) it should be of a standard quality and there must be no significant variation in the quality;

   (iv) it should not be a non-tangible commodity (such as personal services, etc.).

(c) **Classification of commodities :** Sometimes (as in case of cost of living index numbers) we classify the selected commodities into different groups such as :

(i) Food,

(ii) Clothing and Footwear,

(iii) Fuel and Lighting,

(iv) House Rent and

(v) Miscellaneous.

4. **Selection of the representative prices :** In the collection of price quotations we have to consider the following points :

   (i) The method of quoting prices of the commodities—money price or quantity price.

   (ii) The type of quotations—whether wholesale prices or retail prices.

   (iii) The place from where the quotations are to be obtained—The wholesale prices are obtained from wholesale dealers or stores or *mandis* or standard trade journals or agencies or government departments.

5. **System of weighting :** The term 'weight' refers to the relative importance of the different commodities included in the construction of index numbers. There are two methods of assigning weights :

   (i) Implicit Method and (ii) Explicit Method.

   In implicit method, several varieties of a certain type of commodity under study are used. Such weights are called Implicit weights.

   In explicit method, the weights are laid down on the basis of some outward evidence of importance of commodities. One of the problem in the selection of appropriate weight is to decide this evidence. Another problem with regard to the system of weighting is whether weights should be fixed or fluctuating.

6. **Selection of the average :** To find composite index number we can use any average such as arithmetic mean, geometric mean, harmonic mean, median and mode. The use of an average depends on the relative merits and demerits of the various averages. The average may be weighted or unweighted.

   Comparative study of averages : Arithmetic mean, geometric mean and median are the commonly used in practice.

   Arithmetic mean, although simple in its application, gives greater importance to the larger values and less to smaller values. Median is the quickest average to be applied to give the required information, it does not take account of the magnitudes of the extreme values.

   Geometric mean, though a bit difficult in computations, gives the best results. It does not give undue importance to any observation but gives more importance to small values and less importance to bigger values, thus it helps in the evaluation of the real position. The two properties of a good index number "time reversability and factor reversability" are fulfilled, in certain cases when geometric mean is used. Geometric mean makes it easy to change the base year of the index number.

7. **Selection of suitable formula :** There are various formulae for computing index numbers so the selection of a suitable formula also poses some problem. A particular formula is suitable in a particular situation.

Q.4. Describe fixed base method of construction of an index number with examples.

Ans.

Fixed Base Method

In this method, Price Index numbers are calculated on the basis of base year's price, Generally there are two method of base year (a) Fixed single year base method and (b) Multi-year average base year method.

I. Single Year Fixed Base Method

In this method, any normal year is selected as base year. The price of base year is denoted as P_0 and the price of current year as P_1, after this the following formula is used for calculating Index number. Under this method Index number is also known as price relatives :

Formula : Index No. for Price Relative (P.R.) $= \dfrac{P_1 \times 100}{P_0}$

Explanation : P_1 = Price of current year for which Index number is to be calculated.

P_0 = Price of base year

Example 1. Calculate Price Relative (Simple Index Number) based on 1988 for all the years from the following data :

Year	1988	1989	1990	1991	1992	1993
Price (in ₹)	120	140	150	165	175	240

Sol. Thus, Price of the year 1988 $= p_0 = 120$ and so on.

Price Relative (index number) based on 1988 for the current year

$$= \frac{\text{Price of the current year}}{\text{Price of the base year}} \times 100$$

Year	Prices	Calculation	Index Number (Base 1988 = 100)
1988	120	—	100
1989	140	$= \frac{140}{120} \times 100$	116.7
1990	150	$= \frac{150}{120} \times 100$	125.0
1991	165	$= \frac{165}{120} \times 100$	137.5
1992	175	$= \frac{175}{120} \times 100$	145.8
1993	240	$= \frac{240}{120} \times 100$	200

II. Multi-year Average Base Method

In such condition where normal year is not selected easily, the average price of few year is taken as base and this average price is denoted as P_a. The formula is used for the calculation of Index number is as follows :

$$\text{Index No. or P.R.} = \frac{P_1 \times 100}{P_a};$$

Explanation : P_1 = Price of current year for which Index No. is calculated.
P_a = Average Price of few years

Example 2. From the following data calculate index number for all the years by taking 1992 to 1995 as base period :

Year	1990	1991	1992	1993	1994	1995	1996	1997
Price	84	105	123	147	168	210	189	231

Sol. P_a = Average price (for the years 1992 to 1995)

$$= \frac{123+147+168+210}{4} = \frac{648}{4} = 162$$

Formula : Index Number of current year

$$= \frac{\text{Price of current year}}{\text{Average price for the years 1992 to 1995}} \times 100$$

Year	Price	Calculation	Index No. *PR*
1990	84	$\frac{84}{162} \times 100$	51.85
1991	105	$\frac{105}{162} \times 100$	64.81
1992	123	$\frac{123}{162} \times 100$	75.93
1993	147	$\frac{147}{162} \times 100$	90.74
1994	168	$\frac{168}{162} \times 100$	103.70
1995	210	$\frac{210}{162} \times 100$	129.63
1996	189	$\frac{189}{162} \times 100$	116.67
1997	231	$\frac{231}{162} \times 100$	142.59

Q.5. Describe chain base index number with example and write also its advantages and disadvantages.

Ans.

Chain Base Index Numbers

In this method, price relative is also known as Link Relative. Link Relative is calculated on the basis of price of immediately preceeding year. It means base year changes every year for calculating Index number on chain base method. The formula to calculate link relative is as follows :

Formula No. 1

$$\text{Chain Base Index No. or Link Relative (L.R.)} = \frac{\text{Price of Current Year}}{\text{Price of Previous Year}} \times 100$$

Formula No. 2 : *Average link relative :* Under this method, Average Link Relative is calculated as follows :

$$= \frac{\text{Total of Link Relatives}}{\text{No. of Commodities or Items}}$$

Formula No. 3 : Chain Indices chained to a common Base :

Formula : Chained Index No. for Current Year

$$= \frac{\text{Chain Index No. of Previous Year} \times \text{Average link Relative of Current Year}}{100}$$

Example : The prices of wheat for 6 years are given below. Calculate Price Link Relatives (chain base index) upto the nearest integer :

Year	1989	1990	1991	1992	1993	1994
Price (in ₹ per quintal)	300	345	355	400	415	425

Sol. Price Link Relative of the current year $= \frac{\text{Price of the Current Year}}{\text{Price of the Previous Year}} \times 100$

Year	**Price**	**Calculation**	**Price Link Relative (Chain base Index)**
1989	300	—	100
1990	345	$\frac{345}{300} \times 100$	115
1991	355	$\frac{355}{345} \times 100$	103
1992	400	$\frac{400}{355} \times 100$	113
1993	415	$\frac{415}{400} \times 100$	104
1994	425	$\frac{425}{415} \times 100$	102

Advantages of Chain Base Index Number

Following are the advantages of chain base Index Number :

1. Under this method, it is easier to introduce new items or commodities and delete old ones.
2. These Index numbers make easy and reliable comparison of the current situation with past situation.
3. These Index numbers are relatively not affected with seasonal and cyclical variation.

Disadvantages or Limitations of Chain Base Index Number

1. The construction of such Index number is comparatively difficult and calculative.
2. These Index numbers are not useful for long term comparisons.
3. Under this method one error at any place has cumulative effects on all succeeding calculations.

Q.6. Explain Base conversion and base shifting with examples.

Ans.

Base Conversion

Sometimes it is required to change the fixed base index into chain base index and *vice versa*. Following are the methods used for the conversion :

I. Conversion from Chain Base Index (CBI) to Fixed Base Index (FBI)

To convert chain base to fixed base index following method is used :

1. FBI has same base as CBI, for the first year. FBI is taken as 100, if first year is working as base in a problem.
2. Following formula is used to calculate the fixed base indices for the succeeding years :

$$\text{Current Year's FBI} = \frac{\text{Current Year's CBI} \times \text{Previous Year's FBI}}{100}$$

Example : From the chain base index number given below, prepare fixed base index numbers :

Year	2015	2016	2017	2018	2019
CBI	136	162	148	154	185

Sol.

Conversion of CBI into FBI

Year	CBI	Conversion	FBI
2015	136	—	136.00
2016	162	$\frac{162 \times 136}{100}$	220.32
2017	148	$\frac{148 \times 220.32}{100}$	326.1
2018	154	$\frac{154 \times 326.1}{100}$	502.2
2019	185	$\frac{185 \times 502.2}{100}$	929.07

II. Conversion from Fixed Base Index (FBI) to Chain Base Index (CBI)

Following process is used to convert the fixed base to chain base index :

1. CBI has same base as FBI, for the first year.
2. Following formula is used to calculate the chain based index from the fixed base index for the succeeding year :

$$\text{Current Year's CBI} = \frac{\text{Current Year's FBI}}{\text{Previous Year's FBI}} \times 100$$

Example : From the following fixed base index numbers, prepare chain base index numbers :

Year	2015	2016	2017	2018	2019	2020
FBI	120	150	190	260	320	450

Sol.

Conversion of FBI into CBI

Year	FBI	Conversion	CBI
2015	120	—	120

2016	150	$\frac{150}{120}\times 100$	125.00
2017	190	$\frac{190}{150}\times 100$	126.67
2018	260	$\frac{260}{190}\times 100$	136.84
2019	320	$\frac{320}{260}\times 100$	123.08
2020	450	$\frac{450}{320}\times 100$	140.63

Base Shifting

The base shifting refers to the shifting of the base of a series of index numbers from one period to another when the given base year becomes outdated. There is a better possibility of comparison as well as new construction of the whole series on the shifting of the base. Thus, during the shifting of the base the new index number is determined by dividing the old index number of the current year by the old index number of the base year. The result is multiplied by 100.

Following is the formula of base shifting : $P_{01} = \frac{p_0}{p_1} \times 100$

Where, p_0 = Index number of the base year,

p_i = Index number of the new (or current) year taken as the base.

Example : Shift the base from 2010 to 2014 and construct new series for the following data :

Year	Index Numbers (2010 = 100)	Year	Index Numbers (2010 = 100)
2010	100	2010	370
2011	130	2011	410
2012	180	2012	440
2013	220	2013	455
2014	300	2014	448

Sol. Computation of New Index Numbers

Year	Old Index No. (2010 = 100)	New Index No. (2014 = 300)	Year	Old Index No. (2010 = 100)	New Index No. (2014 = 300)
2010	100	$\frac{100}{300}\times 100 = 33.33$	2010	370	$\frac{370}{300}\times 100 = 123.33$
2011	130	$\frac{130}{300}\times 100 = 43.33$	2011	410	$\frac{410}{300}\times 100 = 136.67$
2012	180	$\frac{180}{300}\times 100 = 60$	2012	440	$\frac{440}{300}\times 100 = 146.67$
2013	220	$\frac{220}{300}\times 100 = 73.33$	2013	455	$\frac{455}{300}\times 100 = 151.67$
2014	300	100.00	2014	448	$\frac{448}{300}\times 100 = 149.33$

Q.7. Explain deflating and splicing of index numbers with examples.

Ans.

Deflating Index Numbers

The technique of adjusting the values of a series after allocating changes in price level is known as 'deflation'. If the prices increase then one can purchase fewer amounts from the same amount of money in earlier years. It is said that the 'purchasing power of money' decreases. As a result, it cannot be concluded that resulting from any increase in the money income, there will be the same percentage increase in the 'standard of living'. It means that 'money income' is more than the 'real income'. It is required to adjust the money income for the changes in the cost of living index numbers for proper comparison. Hence, after the deflation of the money income by the cost of living index, the real income is derived.

$$\text{Real Income} = \frac{\text{Money Income}}{\text{Cost of Living Index}} \times 100$$

To find 'real wages' as well as to deflate value series, rupee sales etc., technique of deflation is extensively used alongwith the equivalent price index numbers. By this technique a series of values which is calculated at current prices is converted into constant prices for a given year and used to measure the purchasing power of money.

$$\text{Real Wage} = \frac{\text{Money or Nominal Wage}}{\text{Price Index}} \times 100$$

$$\text{Real Value of Deflated Value} = \frac{\text{Current Value}}{\text{Price Index of Current Year}} \times 100$$

Example : Deflate the per capita income shown in the following table or the basis of the rise in the cost of living index :

Year	2001	2002	2003	2004	2005	2006	2007	2008
Cost of Living Index	100	115	130	140	155	220	245	340
Per Capital Income (₹)	75	90	100	110	130	145	150	170

Sol.

Deflating Per Capita Income

Year	Cost of Living Index (2006 =100)	Actual Per Capita Income (₹)	"Real Income" (₹)
2001	100	75	75.00
2002	115	90	$\frac{90}{115} \times 100 = 78.26$
2003	130	100	$\frac{100}{130} \times 100 = 76.92$
2004	140	110	$\frac{110}{140} \times 100 = 78.57$
2005	155	130	$\frac{130}{155} \times 100 = 83.87$
2006	220	145	$\frac{145}{220} \times 100 = 65.90$

2007	245	150	$\frac{150}{245}\times 100 = 61.22$
2008	340	170	$\frac{170}{340}\times 100 = 50.00$

Splicing

When the base year of an index number series becomes very old, it is usually discontinued and a new series with a recent past year as base is started. Sometimes, it may become necessary to combine these two series. The process of combining of two or more overlapping index number series is called splicing. If there are two index number series, say *A* or old series and *B* or new series, then *B* can be spliced to *A* or *vice-versa*. When *B* is spliced to *A* the base of the spliced series would be same as that of series *A*.

Similarly, when *A* is spliced to *B*, the base of the spliced series would be same as that of series *B*. For splicing, there must be at least one year having index numbers of both the series. Normally, this year is the base year of the new series. Given this, we can find a correction factor for each of the two situations.

1. When *B* is spliced to *A*; Correction factor

$$= \frac{\text{Index of Series } A \text{ in the year corresponding to Base Year of series } B}{100}$$

   All the index numbers of series B are multiplied by this correction factor to get the spliced series.

2. When *A* is spliced to *B*. In this case all the index numbers of series *A* are divided by the above correction factor to get the spliced series.

Example : Given below are the two index number series, one with 1981 as base and the other with 1989 as base.

Series *A*	**Year**	1981	1982	1983	1984	1985	1986	1987	1988	1989
	Index No.	100	110	120	130	170	200	240	300	350
Series *B*	**Year**	1989	1990	1991	1992					
	Index No.	100	125	160	190					

1. Splice Series *B* to series *A* (or series *A* forward)
2. Splice series *A* to series *B* (or series *B* forward)

Sol. The Correction factor $= \frac{350}{100} = 3.5$

Year	**Series *A***	**Series *B***	**Series *B* Spliced to Series *A***	**Series *A* Spliced to Series *B***
1981	100		100	$\frac{100\times 100}{350} = 28.6$
1982	110		110	$\frac{110\times 100}{350} = 31.4$
1983	120		120	$\frac{120\times 100}{350} = 34.3$

1984	130		130	$\frac{130\times100}{350}=37.1$
1985	170		170	$\frac{170\times100}{350}=48.6$
1986	200		200	$\frac{200\times100}{350}=57.1$
1987	240		240	$\frac{240\times100}{350}=68.6$
1988	300		300	$\frac{300\times100}{350}=85.7$
1989	350	100	350	100
1990		125	$\frac{125\times350}{100}=437.5$	125
1991		160	$\frac{160\times350}{100}=560$	160
1992		190	$\frac{190\times350}{100}=665$	190

Q.8. Define Consumer Price Index. Write also its objectives, uses and construction.

Ans. Consumer Price Index (CPI)/Cost of Living Index Number

Consumer price index is generally called cost of living index. It gives the price change paid by actual user of product. Price change affects standard of living of different people in different ways. This cannot be measured using general index number. Due to this reason consumer price index is constructed.

People use various types of goods and the habit of people varies from person to person, place to place and class to class. The scope of the CPI is to define or specify the group of population covered, *e.g.*, poor class, middle class, etc. Urban, rural, etc., are the covered geographical areas.

Objectives of Consumer Price Index

1. The main objective of the CPI is to determine the effect of 'price change' on the purchasing power of the after tax money income earners.
2. It tracks the inflation.
3. In the condition when decisions are taken with the help of monetary policy, it focuses on controlling the inflation.

Uses of Consumer Price Index

Some important uses of the consumer price index :

1. **In the Regulation of Dearness Allowances and Bonus Policy :** To decide the dearness allowances and bonus, various employers and government use the index number as a tool.
2. **In the Determination of Purchasing Power of Money and Value of Real Wages :** It works as a tool which measures the purchasing power of money and value of real wages, as under :

(i) Purchasing Power of Money $= \frac{1}{\text{Price Index}} \times 100$

(ii) Real Wage $= \frac{\text{Money wage}}{\text{Price Index}} \times 100$

Real Income Index $= \frac{\text{Real Income of the current year}}{\text{Real Income of the fixed base year}} \times 100$

Where, Price Index = Cost of living index number.

3. **In the Deflation of Income and Values :** It is used in deflating the income of a national account.
4. **In the Determination of Certain Government Policies :** Government uses CPI to regulate the various policies such as economic policy, income policy, etc.
5. **In the Analysis of Price Situations :** It is used to analyse the price situation of a specific community.

Construction of Consumer Price Index

Following are the two methods used in the construction of consumer price index :

1. **Aggregate Expenditure Method :** It is based on the Laspeyre's method. It is widely used to measure the quantities of goods consumed by a specific group in the base year :

$$\text{Consumer Price Index number} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Example : Construct the consumer price index number for 2014 on the basis of 2010 from the following data using Aggregate expenditure method.

Commodities	Quantity Consumed	Price in	
		2010	2014
P	150	10	14
Q	30	5	9
R	22	7	16
S	28	22	26

Sol. For Aggregate Expenditure Method we have to find the value of $\Sigma p_0 q_0$, $\Sigma p_1 q_0$ which are calculated in the following table :

Commodities	q_0	p_0	p_1	$p_0 q_0$	$p_1 q_0$
P	150	10	14	1500	2100
Q	30	5	9	150	270
R	22	7	16	154	352
S	28	22	26	616	728
				$\Sigma p_0 q_0 = 2420$	$\Sigma p_1 q_0 = 3450$

Consumer price index by Aggregate Expenditure Method

$$= \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{3450}{2420} \times 100 = 142.56$$

2. **Family Budget Method or Method of Weighted Relatives :** It is used to determine the aggregate expenditure of an average family on different goods. Following formula is used to calculate the CPI,

$$\text{Consumer Price Index Number} = \frac{\Sigma PW}{\Sigma W}$$

Where, $P = \frac{p_1}{p_0} \times 100$ for each item; $W =$ value weight $(i.e.,)$ $p_0 q_0$.

Note : "Weighted average price relative method" and "Family Budget method" are the same for finding out consumer price index.

Example : Construct the consumer price index number for 2013 on the basis of 2009 from the following data using method of weighted relatives.

Commodities	Quantity Consumed	Price in	
		2009	2013
P	110	6	9
Q	40	8	14
R	15	4	19
S	70	12	21

Sol.

Construction of Consumer Price Index Number

Commodities	q_0	p_0	p_1	Price Relative $(P) = \frac{p_1}{p_0} \times 100$	Weight (W) $p_0 q_0$	$P \times W$
P	110	6	9	150	660	99000
Q	40	8	14	175	320	56000
R	15	4	19	475	60	28500
S	70	12	21	175	840	147000
Total				$\Sigma P = 975$	$\Sigma W = 1880$	$\Sigma PW = 330500$

$$\text{Consumer Price Index Number} = \frac{\Sigma PW}{\Sigma W} = \frac{330500}{1880} = 175.80$$

Q.9. Explain Fisher's formula for price index with example. Write its merits and demerits.

Ans.

Fisher's Formula

Prof. Irwing Fisher has suggested a compromise between the Laspeyre's and Paasche's formulae by taking geometric mean of these formulae. Thus Fisher's formula for price index is given by

$$^F P_{01} = \sqrt{^L P_{01} \times {^P P_{01}}}$$

where, $^L P_{01} =$ Laspeyre's Price Index, $^P P_{01} =$ Paasche's Price Index

$$^F P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

Merits

1. It satisfies both the 'time reversal test' as well as the 'factor reversal test'. This is why it is called an 'Ideal Formula'.
2. It is free from bias, upward as well as downward.
3. This formula takes into account both current years as well as base year prices and quantities.

Demerits

1. It is not, however, a practical index to compute because it is excessive ly laborious.
2. It requires the prices and quantities for base year and current year.
3. This formula is difficult to interpret.

Since Fisher's Formula satisfies Time Reversal Test and Factor Reversal Test, it is called Fisher's Ideal Index Formula and the index number so obtained is known as Fisher's Ideal Index Number.

Fisher's Formula does not satisfy circular test. Although Fisher's Formula satisfies the tests of a good index number but nobody can say as to what exactly Fisher's Index Number is supposed to measure.

According to **Boddington**, "Unfortunately, while this formula apparently meets most of the mathematical requirements of a perfect index formula, it is objected to. On the score that it is not clear what it measures *i.e.* The result combines both price and value changes, when usually we want the one to be separated from the other."

Example : From the following data, find out price index number by Fisher Ideal Formula for 1995 based on 1985 :

Commodity	1985		1995	
	Price	Quantity	Price	Quantity
P	12	120	20	120
Q	4	200	4	240
R	8	120	12	150
S	20	60	24	50

Sol.

Calculation of Prime Index Number

Commodity	1985		1995		p_0q_0	p_0q_1	p_1q_0	p_1q_1
	p_0	q_0	p_1	q_1				
P	12	100	20	120	1,200	1,440	2,000	2,400
Q	4	200	4	240	800	960	800	960
R	8	120	12	150	960	1,200	1,440	1,800
S	20	60	24	50	1,200	1,000	1,440	1,200
Total					**4,160**	**4,600**	**5,680**	**6,360**

$$\text{Laspeyre's Formula} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{5{,}680}{4{,}160} \times 100 = 136.5$$

$$\text{Paasche's Formula} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100 = \frac{6,360}{4,600} \times 100 = 138.3$$

∴ Fisher's Ideal Price Index Number

$$= \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100 = \sqrt{\frac{5680}{4160} \times \frac{6360}{4600}} \times 100$$

$$= \sqrt{1.365 \times 1.383} \times 100 = \sqrt{1.888} \times 100$$

$$= 1.374 \times 100 = 137.4$$

Alternative Method : Fisher's Ideal Price Index Number

$$= \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100 = \sqrt{\frac{5,680}{4,160} \times \frac{6,360}{4,600}} \times 100$$

$$= \sqrt{1.365 \times 1.383} \times 100 = \sqrt{1.888} \times 100$$

$$= 1.374 \times 100 = 137.4$$

Q.10. What do you understand by Time Reversal and Factor Reversal Tests?

Ans.

Time Reversal Test

The time reversal test was developed by *Irwing Fisher* from the view point that an index number should work both ways (backward and forward) with respect to time. According to him, "If a formula for constructing an index number is such that it will give the same ratio between one point of comparison and the other, no matter which of the two is taken as base, then formula is said to have obeyed the time reversal test."

In other words, "When the data for any two years are treated by the same formula, but with the bases reversal, the two index numbers obtained should be reciprocals of each other so that their product is unity (neglecting the factor 100)."

Let P_{01} = Index Number for year 1, taking base as 0

P_{10} = Index Number for year 0, taking base as 1

If $P_{01} \times P_{10} = 1$, neglecting the factor 100; otherwise $P_{01} \times P_{10} = 100^2$

Then the formula for the constructions of index number is said to satisfy the time reversal test.

Time reversal test is satisfied by :

(i) Fisher's Formula :

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}, P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}, P_{01} \times P_{10} = 1$$

(ii) Simple geometric mean of relatives.

(iii) Weighted aggregative with fixed weights.

(iv) Weighted geometric mean of relatives with fixed weights.

(v) Marshall-Edgeworth's Formula :

$$P_{01} = \frac{\Sigma p_1 (q_0 + q_1)}{\Sigma p_0 (q_0 + q_1)}, P_{10} = \frac{\Sigma p_0 (q_1 + q_0)}{\Sigma p_1 (q_1 + q_0)}, P_{01} \times P_{10} = 1$$

Time reversal test is not satisfied by :

(i) Laspeyre's Formula :

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}, \quad P_{10} = \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1}, \quad P_{01} \times P_{10} \neq 1$$

(ii) Paasche's Formula :

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}, \quad P_{10} = \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}, \quad P_{01} \times P_{10} \neq 1$$

Factor Reversal Test

The factor reversal test was developed by Irwing Fisher to extend the logical property

Price × Quantity = Value to a group of commodities.

In the words of *Fisher*, "Just as each formula should permit the interchange of the two times without giving inconsistent results, so it ought to permit inter-changing the prices and quantities without giving in consistent result *i.e.,* the two results multiplied together should give the true ratio." In other words, "When prices and quantities are inter-changed in the price index formula we get quantity index formula which when multiplied by the price index it should give the Value Index, by omitting the factor 100."

Let P_{01} = price index for year '1' with year '0' as base

Q_{01} = quantity index for year '1' with year '0' as base

V_{01} = value index for year '1' with year '0' as base

$$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

If $P_{01} \times Q_{01} = V_{01}$, neglecting the factor 100

Otherwise $P_{01} \times Q_{01} = V_{01} \times 100$

then the formula for the construction of index number is said to satisfy the 'factor reversal test'.

The factor reversal test is satisfied only by Fisher's Index Formula :

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

Changing p to q and q to p, we have

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$\therefore \quad P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \sqrt{\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0}}$$

$$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = V_{01} \qquad \{\because \Sigma p_1 q_1 = \Sigma q_1 p_1 ; \Sigma p_0 q_0 = \Sigma q_0 p_0$$

Marshall-Edgeworth's formula does not satisfy Factor Reversal Test :

$$P_{01} = \frac{\Sigma p_1 (q_0 + q_1)}{\Sigma p_0 (q_0 + q_1)} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1}$$

$$Q_{01} = \frac{\Sigma q_1(p_0 + p_1)}{\Sigma q_0(p_0 + q_1)} = \frac{\Sigma q_1 p_0 + \Sigma q_1 p_1}{\Sigma q_0 p_0 + \Sigma q_0 p_1}$$

$$V_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$P_{01} \times Q_{01} \neq V_{01}$$

Q.11. Construct index number from the following data by using Fisher's Formula and prove Time Reversal Test :

Commodities	Base Year		Current Year	
	Price (₹)	Quantity (kg)	Price (₹)	Quantity (kg)
P	6	50	10	56
Q	2	100	2	120
R	4	60	6	60
S	10	30	12	24
T	8	40	12	36

Sol. **Construction of Index Number**

Commodities	Base Year		Current Year		p_0q_0	p_1q_0	p_0q_1	p_1q_1
	p_0	q_0	p_1	q_1				
P	6	50	10	56	300	500	336	560
Q	2	100	2	120	200	200	240	240
R	4	60	6	60	240	360	240	360
S	10	30	12	24	300	360	240	288
T	8	40	12	36	320	480	288	432
	—	—	—	—	Σp_0q_0 =1,360	Σp_1q_0 =1,900	Σp_0q_1 =1,344	Σp_1q_1 =1,880

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100 = \sqrt{\frac{1,900}{1,360} \times \frac{1,880}{1,344}} \times 100$$

$$= \sqrt{1.397 \times 1.399} \times 100 = \sqrt{1.4 \times 1.4} \times 100$$

$$= 1.4 \times 100 = 140$$

With the help of log tables :

$$\sqrt{\frac{1,900}{1,360} \times \frac{1,880}{1,344}} = \text{Antilog}\left[\frac{1}{2}(\log 1,900 + \log 1,880 - \log 1,360 - \log 1,344)\right]$$

$$= \text{Antilog}\left[\frac{1}{2}(3.2788 + 3.2742 - 3.1335 - 3.1284)\right]$$

$$= \text{Antilog}\left[\frac{1}{2}(6.5530 - 6.2619)\right]$$

$$= \text{Antilog}\left[\frac{1}{2}(0.2911)\right]$$

$$= \text{Antilog}(0.1456) = 1.398 \text{ or } 1.40$$

Now,
$$P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} \times 100$$

$$= \sqrt{\frac{1{,}344}{1{,}880} \times \frac{1{,}360}{1{,}900}} \times 100$$

$$= \sqrt{0.71 \times 0.72} \times 100 = 0.715 \times 100 = 71.5$$

Further
$$P_{01} \times P_{10} = 1.4 \times 0.715 \text{ (neglecting the factor 100)}$$

$$= 0.994 \text{ or } 1$$

Fisher's Formula satisfies Time Reversal Test.

Q.12. Construct index number by using Fisher's Ideal Formula and show how it satisfies Factor Reversal Test from the data given below :

Commodities	Value in ₹		Base Year	Current Year
	Base Year	Current Year	Quantity (kg)	Price (₹)
P	300	560	50	10
Q	200	240	100	2
R	240	360	60	6
S	300	288	30	12
T	320	432	40	12

Sol. Since we are given Value in Rupees (price × quantity) for the base and current years, the price for the base year and quantity consumed in the current years are to be obtained :

Commodity	Price for Base Year	Quantity for Current
P	$\frac{300}{50} = 6$	$\frac{560}{10} = 56$
Q	$\frac{200}{100} = 2$	$\frac{240}{2} = 120$
R	$\frac{240}{60} = 4$	$\frac{360}{6} = 60$
S	$\frac{300}{30} = 10$	$\frac{288}{12} = 24$
T	$\frac{320}{40} = 8$	$\frac{432}{12} = 36$

Computation of Fisher's Ideal Formula

Items	Base Year		Current Year		p_0q_0	p_1q_1	p_0q_1	p_1q_0
	p_0	q_0	p_1	q_1				
P	6	50	10	56	300	560	336	500
Q	2	100	2	120	200	240	240	200
R	4	60	6	60	240	360	240	360
S	10	30	12	24	300	288	240	360
T	8	40	12	36	320	432	288	480
					Σp_0q_0 =1,360	Σp_1q_1 =1,880	Σp_0q_1 =1,344	Σp_1q_0 =1,900

Fisher's Price Index Number

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

$$\therefore \quad P_{01} = \sqrt{\frac{1{,}900}{1{,}360} \times \frac{1{,}880}{1{,}344}} \times 100 = 139.79$$

Fisher's Quantity Index Number,

$$Q_{01} = \sqrt{\frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times \frac{\Sigma q_1p_1}{\Sigma q_1p_0}} \times 100$$

Factor Reversal Test is satisfied if

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma q_1p_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times \frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times \frac{\Sigma q_1p_1}{\Sigma q_0p_1}} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

Hence, we conclude that Fisher's Formula satisfies Factor Reversal Test.

Q.13. What is a time series? Explain its importance.

Ans.

Meaning of Time Series

Time series refers to such a series in which statistical data are presented on the basis of time of occurrence or in a chronological order. The measurement of time may be either year, month, week, day, hour or even minutes or seconds. Technically, in a time series the time factor is an independent variable and the dependent variable in the form of statistical data represents the movements related to time factor. The following two series are the examples of time series :

Population in India		Wholesale Price Index No. (2004-05 =100)	
Year	Population (Crores)	Month (Year 2012)	Index No.
1971	54.0	April	197
1981	68.4	May	195
1991	84.4	June	196
2001	102.7	July	196
2011	120.2		

It is worth mentioning here, that the time series is also termed as *'Historical Series'* or *'Chronological Series'*. A few definitions of time series are as follows :

1. "A set of data depending on the time is called a time series." —*Kenny & Kieping*
2. "A time series is a sequence of values of the same variate corresponding to successive points in time". —*Werner Z. Hirsch*
3. "A time series may be defined as a collection of reading belonging to different time periods, of some economic variable or composite of variables". —*Ya-Lun-Chou*

Conclusively, it can be said that time series is an arrangement of statistical data in a chronological order. It reflects the dynamic pace of movements of a phenomenon over a period of time. For example, annual production of sugar during ten years, population census after every ten years, the variations in the temperature of a patient during a certain period, monthly price index numbers during a year, etc.

Importance or Utility of Analysis of Time Series

The analysis of time series is extremely useful in business decision-making as it is clear from the following points :

1. **Estimation of Trade Cycles :** Trade cycles can be estimated on the basis of cyclical fluctuations and it helps the businessman plan and regulate his activities.
2. **Evaluation of Performance :** It helps in the evaluation of current achievements. For example, the progress of long-term plan can be evaluated on the basis of comparison of yearly targets of growth rate and actual annual achievements.
3. **Comparative Study :** Analysis of time series facilitates the comparative study of data for two or more related periods. For example, to compare the monthly data of production in a factory during a period of ten years from 2000 to 2010.
4. **Forecasting :** The knowledge of trend and fluctuations on the basis of analysis of time series makes it possible forecast the future behaviour of the variable under study. On this basis businessmen, administrators and planners can lay down future policies and targets and can prepare future plan. *Prof. Werner Z. Hirsch* writes, "The main objective in analysing time series is to understand, interpret and evaluate changes in economic phenomena in the hope of more correctly anticipating the course of the events."
5. **Analysis of Past Behaviour :** Analysis of time series helps in the analysis of past behaviour of a variable which discloses the circumstances and causes influencing the movement of data. This information can be helpful in the arrangement for controlling the behaviour of the variable.

Q.14. Explain the main components of a Time series.

Ans.

Components of a Time Series

Time series is influenced collectively by a large variety of factors and forces. The effects of these forces can be classified in some definite categories. These categories are called the components of time series. The main components of time series may be classified as shown below :

Original Data (O or Y)

- Long-term Movement or Trend or Secular Trend (T)
- Short-time Fluctuations (O–T)
  - Regular Short-time Fluctuations (S + C)
    - Seasonal Variations (S)
    - Cyclical Fluctuations (C)
  - Irregular or Random Fluctuations (I)

These various components of time series have been examined in detail in the following paragraphs :

1. Secular Trend or Long-term Movement or Trend

Trend refers to that tendency which indicates the general direction of fluctuation in a long period. In simple words, it can be stated that despite various fluctuations from time to time, there will be an underlying tendency of movement in a particular direction and this tendency is called as long-term trend. *For example*, that despite the fluctuations in prices in our country, the long-term trend is of increasing. There are certain such facts also, in which tendency moves to one direction only such as continuous increase in population, continuous decline in death rate, etc.

Some important characteristics of secular trend are as follows :

(a) **Three Aspects :** There may be three aspects of long-term trend : Upward trend, Downward trend, and Stable trend. An upward trend is usually observed in time series relating to prices, population, etc., while a downward trend is noticed in data of illiteracy, death rate, etc. The stable trend is found in pure and natural sciences such as tendency of temperature of an individual on 98.6°F.

(b) **Different Trends during Different Periods :** It is not necessary that the trend should be in the same direction throughout the given period. It may be possible that different tendencies of increase, decrease or stability are observed in different periods of time. However, the overall tendency may be upward, downward or stable.

(c) **Relative Concept :** The term 'long period of time' is a relative concept, which is influenced by the characteristics of the series. *For example*, during the days of curfew, the period of one month may be a long period in the context of data relating to number of deaths, while in price statistics a period of several years will be considered as a long period of time.

The symbol of T is used for denoting long-term trend in the formulae relating to analysis of time series.

2. Regular Short-time Oscillations

Most of the time series are influenced by such factors or forces which repeat themselves periodically. The variations arising out on account of such regular or periodical repetitions are called regular short-time oscillations, which may be classified into the following two categories :

(a) **Seasonal Variations :** Seasonal variations refer to such movements which are regular and repetitive and which operate in a regular and periodic manner over a span of less than a year. This span may be a day, week, month, quarter, half-year, etc.

Characteristics : The main characteristics of seasonal variations are as follows :

(i) **Regular movement :** Seasonal variations occur regularly almost at the same time and about the same proportion within a period of less than one year.

(ii) **Swings in both directions :** These variations may swing to any direction-upward or downward.

(iii) **Easy forecast :** Seasonal variations can easily be forecasted. Various economic and business activities are operated on the basis of these forecasting. Consumers, producers and sellers give due consideration to these variations while taking decisions about their operations. These variations are denoted by letter S in the analysis of time series.

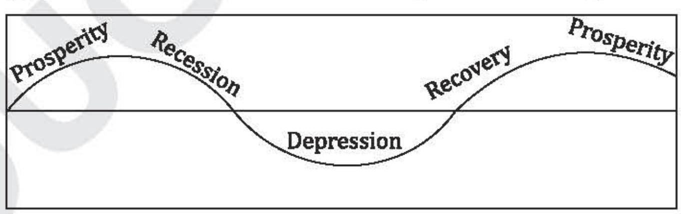
(b) **Cyclical Fluctuations :** Cyclical fluctuations also occur periodically like seasonal variations, but the period of their reoccurrence is more than a year. These fluctuations are called cyclical because they occur in cyclical nature and in this cycle there are four stages :

(i) Prosperity,

(ii) Recession,

(iii) Depression,

(iv) Recovery, which can be demonstrated by the following chart :



It is worth-mentioning that there is no definite period of cyclical fluctuation. Generally, this period varies from 3 to 10 years. The concept of cyclical fluctuation can be understood more precisely by the examples that after every three years there is tendency of bumper crops of mangoes or after every four years production of sugarcane reaches to the peak, etc.

In formulae, the cyclical fluctuation is denoted by the symbol of C.

3. Irregular or Random Fluctuations

Irregular or random fluctuations occur accidently in time series. For instance, decline in profits due to break of fire in the factory in a particular year, decrease in production due to sudden strike or scarcity of petroleum products due to war, etc. The following facts are important about the characteristics of these fluctuations :

(i) **No Forecasting :** These fluctuations are the results of such forces working randomly, which cannot be predicted or forecasted.

(ii) **No Definite Pattern :** Irregular fluctuations do not exhibit any definite pattern and there is no regular period or time of their occurence. That is why, they are called irregular.

(iii) **Short-term :** Generally, they occur as short-term variations but sometimes their effect may be so intense that they give rise to new cyclical or other movements.

(iv) **Coverage :** Irregular fluctuations cover all such variations in a time series, which are not covered within the gamut of trend, seasonal and cyclical movements.

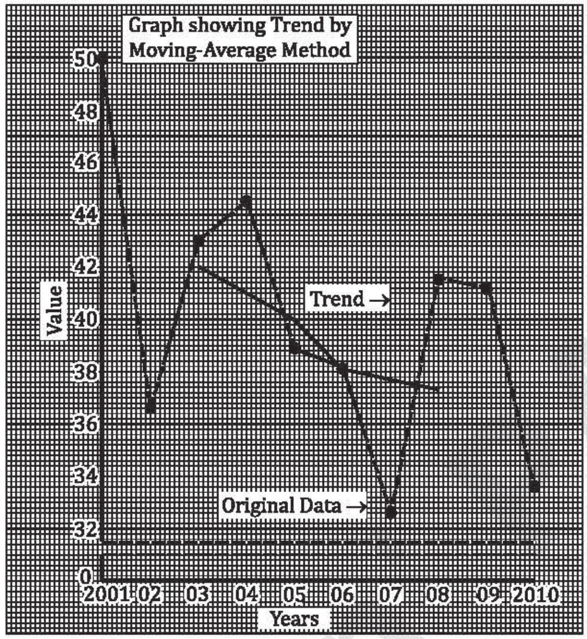
In the process of analysis, irregular fluctuations are denoted by the symbol of *T*.

Q.15. From the following data, calculate the 4-yearly moving averages and determine the trend values. Plot the original data and the trend on a graph :

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Value	50.0	36.5	43.0	44.5	38.9	38.1	32.6	41.7	41.1	33.8

Sol. Calculation of Trend Values by 4-yearly Moving Averages

Year	Value	4-yearly Moving Total	2 Periods Moving Totals Centred	Moving Averages (Trend Values)
2001	50.0			
2002	36.5			
		174.0		
2003	43.0		336.9	42.11
		162.9		
2004	44.5		327.4	40.93
		164.5		
2005	38.9		318.6	39.83
		154.1		
2006	38.1		305.4	38.18
		151.3		
2007	32.6		304.8	38.10
		153.5		
2008	41.7		302.7	37.84
		149.2		
2009	41.1			
2010	33.8			



Q.16. Explain moving-average method of trend of a time series.

Ans.

Moving-Average Method

Moving average method is a simple and flexible device of reducing fluctuations and obtaining trend values with a fair degree of accuracy. It consists in obtaining a series of moving averages (arithmetic means) of successive overlapping groups or sections of the time series. For *example*, there are six years *a, b, c, d, e* and *f* and three year's moving average is to be computed. It will be done as follows :

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}, \frac{d+e+f}{3}$$

The basic question to be decided in this method is that what should be the period of moving average, i.e., three yearly, four yearly, five yearly, etc. This decision is taken on the basis of size of data and fluctuations therein. From the point of view of calculation of moving averages, the questions can be divided in two categories : (1) when period is odd, and (2) when period is even.

1. **Odd Period Moving Averages :** It means moving averages of odd period or years, i.e., 3, 5, 7, 9, 11.......years. Its procedure can be explained as below on the assumption that three yearly moving averages are to be calculated : (i) First of all, three yearly moving totals will be obtained. The total of first three years will be placed against the centre of three years, *i.e.,* second year. (ii) After it, total of next three years (second, third and fourth) will be placed against third year, total of succeeding three years (third, fourth and fifth) will be placed against fourth year and this process will continue till the value of the last year is included in the total. (iii) Moving averages will be obtained by dividing each moving total by 3. It is important that moving averages will not be obtained for first and last year in case of 3 yearly moving averages and first two and last two years in case of 5 yearly moving averages.

Example : Calculate trend values from the following data assuming 5-year and 7-year moving averages :

Year	1	2	3	4	5	6	7	8
Value	110	104	98	105	109	120	115	110
Year	9	10	11	12	13	14	15	16
Value	114	122	130	127	122	118	130	140

Sol. **Calculation of Trend Values By Moving-average Method**

Year	Value	Moving Totals		Moving Average (Trend)	
		5-year	7-year	5-year	7-year
1	110	—	—	—	—
2	104	—	—	—	—
3	98	526	—	105.2	—
4	105	536	761	107.2	108.71
5	109	547	761	109.4	108.71
6	120	559	771	111.8	110.14
7	115	568	795	113.6	113.57
8	110	581	820	116.2	117.14
9	114	591	838	118.2	119.71
10	122	603	840	120.6	12.00
11	130	615	843	123.0	120.43
12	127	619	863	123.8	123.29
13	122	627	889	125.4	127.00
14	118	637	—	127.4	—
15	130	—	—	—	—
16	140	—	—	—	—

2. **Even Period Moving Averages :** If the moving average is to be calculated on the basis of even period, *i.e.,* 2, 4, 6 years, then averages are calculated after centring the moving totals. Suppose, four yearly moving totals are to be calculated, the following procedure would be adopted :

   (i) First of all, four yearly moving totals will be obtained. The first total will be of first four years, the next total of four years excluding first year and this process will be repeated. The first total will be placed between second and third year, second total between third and fourth year and so on.

   (ii) After it, these moving totals will be centred. For this purpose two periods moving totals will be obtained.

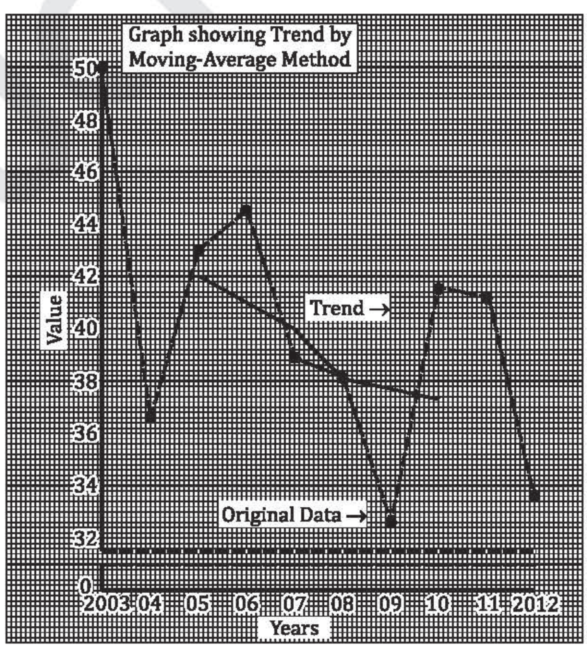
   (iii) Two period moving totals will be divided by 8.

Example : From the following data, calculate the 4-yearly moving averages and determine the trend values. Plot the original data and the trend on a graph :

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Value	50.0	36.5	43.0	44.5	38.9	38.1	32.6	41.7	41.1	33.8

Sol. **Calculation of Trend Values by 4-yearly Moving Average**

Year	Value	4-yearly Moving Totals	2 Periods Moving Totals Centred	Moving Averages (Trend Values)
2003	50.0			
2004	36.5			
		174.0		
2005	43.0		336.9	42.11
		162.9		
2006	44.5		327.4	40.93
		164.5		
2007	38.9		318.6	39.83
		154.1		
2008	38.1		305.4	38.18
		151.3		
2009	32.6		304.8	38.10
		153.5		
2010	41.7		302.7	37.84
		149.2		
2011	41.1			
2012	33.8			



Q.17. Fit a straight line trend by the method of least squares to the following data :

Year	2009	2010	2011	2012	2013	2014	2015
Production ('000 Qntls)	80	90	92	83	94	99	92

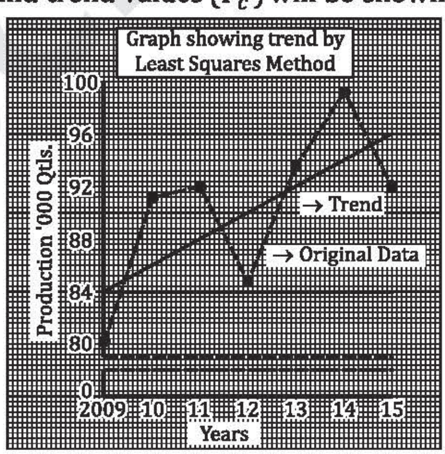
Sol. **Calculation of Trend Values by Least Squares Method**

Year	Production ('000 Qntls.)	Deviation from 2009	Squares	Product of X and Y	Trend Values
	Y	X	X^2	XY	$a+bX=Y_c$
2009	80	– 3	9	– 240	$90+2\times-3=84$
2010	90	– 2	4	– 180	$90+2\times-2=86$
2011	92	– 1	1	– 92	$90+2\times-1=88$
2012	83	0	0	0	$90+2\times0=90$
2013	94	1	1	94	$90+2\times1=92$
2014	99	2	4	198	$90+2\times2=94$
2015	92	3	9	276	$90+2\times3=96$
$N=7$	$\Sigma y=630$	$\Sigma X=0$	$\Sigma X^2=28$	$\Sigma XY=56$	$\Sigma Y_c=630$

$$a=\frac{\Sigma Y}{N}=\frac{630}{7}=90,\ b=\frac{\Sigma XY}{\Sigma X^2}=\frac{56}{28}=2$$

$\therefore$ $$Y_c=90+2X$$

The production figures (Y) and trend values (Y_c) will be shown on graph paper as follows :



❑

MODEL PAPER

Business Statistics

B.Com.-I (SEM-I) [M.M. : 75

Note : Attempt all the sections as per instructions.

Section-A : Very Short Answer Type Questions

Instruction : Attempt all **FIVE** questions. Each question carries **3 Marks.** Very Short Answer is required, not exceeding 75 words.

1. What do you know about Professor Prasanta Chandra Mahalanobis?
2. What are the limitations of graphs?
3. Calculate the Median when Mean and Mode of Distribution are 38.6 and 32.6 respectively.
4. Write the methods of computing correlation.
5. Write the importance of Index Number.

Section-B : Short Answer Type Questions

Instruction : Attempt all **TWO** questions out of the following 3 questions. Each question carries **7.5 Marks.** Short Answer is required not exceeding 200 words.

6. Discuss main limitations of statistics.

Or The real and estimated monthly income of three persons *A, B, C* are given below; calculate absolute error, relative error and percentage error :

Person	*A* ₹	*B* ₹	*C* ₹
Real monthly income	1,000	600	920
Estimated monthly income	1,100	500	920

7. Mean of 12 values is 128. While calculating this mean, one value which actually was 101 was wrongly read as 110. Find the actual mean of the values.

Or Define the importance and applications of correlation analysis.

8. Write the merits and limitations of Least Squares Method.

Or Construct Fisher's ideal Index from the following data and show that it satisfies time reversal and factor reversal tests :

Commodity	2013		2014	
	Price	Value	Price	Value
P	18	140	22	170
Q	22	90	28	120
R	12	150	15	180
S	26	90	32	110
T	38	440	60	500

Section-C : Long Answer Type Questions

Instruction : Attempt all **THREE** questions out of the following 5 questions. Each question carries **15 Marks**. Answer is required in detail, between 500-800 words.

9. Discuss the meaning and scope of statistics.

Or What do you mean by probability sampling method? Explain the different probability sampling methods.

10. What do you mean by measures of central tendency? Describe its properties, various types and importance.

Or Calculate mean by shortcut method from the following data :

Years	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Price of Rice (in ₹)	40	50	55	80	58	60	75	35	45	52

11. What do you mean by measure of dispersion. Discuss its types, objects and importance. Also write the methods of measuring dispersion?

Or Calculate quartiles from the following data :

Marks obtained : 06, 30, 37, 18, 14, 42, 34, 11, 09, 26, 22, 03, 28, 52, 48

12. What do you mean by Degree of correlation and correlation coefficient? Write the applications of correlation also.

Or Calculate Karl Pearson's Coefficient of Correlation. From the following data :

X	42	52	55	60	66	68	65	60	58	34
Y	11	13	18	22	26	40	31	27	24	18

13. Discuss the main characteristics and limitations of Index Number.

Or Deflate the per capita income shown in the following table or the basis of the rise in the cost of living index :

Year	2001	2002	2003	2004	2005	2006	2007	2008
Cost of Living Index	100	115	130	140	155	220	245	340
Per Capital Income (₹)	75	90	100	110	130	145	150	170

❑