

STATISTICS FOR ECONOMICS

NAMAN SINGH
ANJALI VERMA



BlueRose ONE^{.com}
S t o r i e s M a t t e r
New Delhi • London

BLUEROSE PUBLISHERS

India | U.K.

Copyright © Dr. Naman Singh and Anjali Verma 2026

All rights reserved by author. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the author. Although every precaution has been taken to verify the accuracy of the information contained herein, the publisher assumes no responsibility for any errors or omissions. No liability is assumed for damages that may result from the use of information contained within.

BlueRose Publishers takes no responsibility for any damages, losses, or liabilities that may arise from the use or misuse of the information, products, or services provided in this publication.



BlueRose ONE
Stories Matter
New Delhi • London

For permissions requests or inquiries regarding this publication,
please contact:

BLUEROSE PUBLISHERS

www.BlueRoseONE.com

info@bluerosepublishers.com

+91 8882 898 898

+4407342408967

ISBN: 978-93-7542-582-3

Cover design: Deepak Katheria

First Edition: January 2026

Contents

1. ECONOMICS: AN INTRODUCTION	1
2. CENSUS AND SAMPLE METHOD.....	14
3. ORGANIZATION OF DATA	27
4. TEXTUAL AND TABULAR PRESENTATION OF DATA	49
5. DIAGRAMMATIC PRESENTATION OF DATA: BAR DIAGRAMS AND PIE DIAGRAMS.....	58
6. MEAN, MEDIAN, MODE.....	86
7. CORRELATION.....	132
8. INDEX NUMBER	159

[Chapter – 1]

ECONOMICS: AN INTRODUCTION

➤ ECONOMICS :

It is the study of how people and society choose to employ scarce resources that could have alternative uses in order to produce various commodities that satisfy their wants and to distribute them for consumption among various persons and groups in society.

(1) Economics as a science can be:—

- (i) **Positive Economics:**—It deals with what are the economic problems and how are they actually solved. Eg: India is a over populated country. It is based upon facts
- (ii) **Normative Economics:**—It deals with what ought to be or now the economic problems should be solved. *Eg: India should not be over populated country. It is suggested in nature.*

(2) Economics as an art:— Just like arts, economics also requires harmless in a questions for the practical application of laws, principles and theories.

❖ Human Activities

- (1) Economic Activities
- (2) Non-Economic Activities

(3) Economic Activities:—It refers to those activities where are takes to care a living.

Eg. Workers working in factory, etc.

❖ There mains economic activities:—

(i) **Consumption:** It is an economicactivities which deals with the use of goods and services for the satisfactions of human wants.

(ii) **Production:** It is the process of converting raw material into finished goods.

- ❖ Productions enhances the utility of the product by charging it is the force needed by the consumers.
- ❖ It includes all activities which are undertaken to produce goods for income and satisfying because wants.

- ❖ It involves deciding what to produce and how to produce.
- ❖ Land, labour, capital and entrepreneur are the four factors of production.

(iii) Distributions: It is that economic activity which studies how income generated is distributed among these factors of production. It deals with determining how GDP is distributed to the factors in the form of rent, wages, interest and profit respectively.

Non-Economic Activities :—Activities which are not concerned with creation of or wealth are known as non-Economic activities.

Eg. Teacher teaching his own son.

Definitions of Economics by Different Economists.

Wealth Definitions

- It is given by Adam Smith.
- In ordinary languages, Wealth means money, In economics, wealth refers to those goods which satisfy human wants.
- Adam said that economics is concerned with the problems arising from wealth – creating and wealth using activities of people.
- However, this definition was criticized as it equal human welfare and moral values.

Scarcity Definition

- It is given by Prof. Lionel Robbins.
- According to him, Economics is the science which studies human behavior as a relationship between ends and scarce means which have alternative uses.
- This definition of economics is given in his book, 'An essay on the nature and significance of economic science', **published in 1932.**
- Scarcity definition is based on following fundamental characteristics of human existence, which gives rise to economic problem.
 - (a) Unlimited Wants:—** Man has unlimited wants i.e. he can never be fully satisfied. No sooner a want is satisfied a new want
 - (b) Scarcity of Resources:—** Scarcity refers to limitations of supply in relation to demand for a commodity.
 - (c) Alternatives uses:** Resources are not only scarce but they can also be put to various uses.

Growth oriented Definition

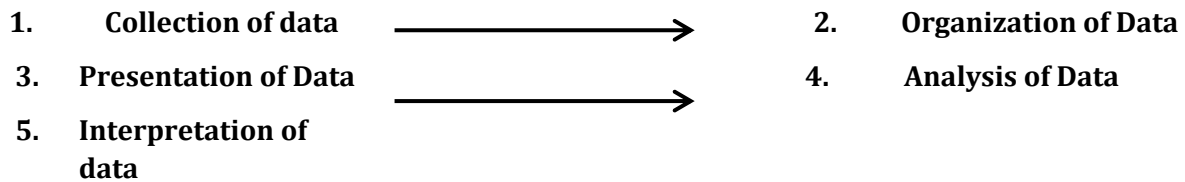
- It is given by Prof. Anthony Samuelson.

- According to him;
 'Economics is the study of how man and societies choose with or without the use of money to employ scarce productive resources, which would have alternative uses, to produce various commodities over time and distribute them for commodities now and in the future among various people and groups of society.
 Samuelson's definition tells us how economics is a social science and it is mainly covered with the way how society employs its limited resources for alternative uses.
 This definition is considered as most alternative and satisfactory.

Meaning and Importance of Statistics

- **Meaning of Statistics**
 In plural sense, it means a collection of numerical facts.
 In Singular sense, it deals with the collection, presentation, analysis and interpretation of the quantitative information.
- **Characteristics of Statistics in Plural Sense**
 - (i) **Aggregates of facts:-** Statistics are a number of facts. Useful and isolated figures are not statistics as such figures cannot be compared and no conclusion can be drawn from them.
 Eg. A single age of 30 years is not a statistics but a series relating to the age of group of persons will be called statistics.
 - (ii) **Affected by Multiplicity of Causes :-** Numerical figures (data) are influenced by a variety of factors. It is not an easy job to study the effects of any one factor separately by ignoring other factors.
 - (iii) **Statistics are stratified :- The statistical approaches to a subject is numerical so, any facts in a statistic must be numerically or quantitatively expressed.**
 - (iv) **Statistics should be collected with reasonable standard and of accuracy:-** In statistics, data is collected with reasonable standard of accuracy. A high degree of accuracy, as observed in accounting and mathematics, is not found in statistics, mass of data is involved.
 - (v) **Statistics are collected for a predetermined purpose:-** The purpose of collecting statistical data must be decided in advance, otherwise usefulness of the data collected would be negative. Data collected in an unsystematic manner and without complete awareness of the purpose will be misleading and cannot be made on this basis of valid conclusion.

➤ **Characteristics of Statistics in Singular sense**



1. **Collection of Data:**– It is the main and the first step in statistical enquiry. The technique of collections of data depends upon the objective of the study.
2. **Organization of Data:**– After collection, the data is organized in a proper forums which involves editing and classification.
3. **Presentation of Data:** After classification, the data is presented in some suitable manner, in the form of text, table, diagram, and graph.
4. **Analysis of Data:** After presentation of data, analysis is done with the help of simple mathematical techniques. It includes measures of central tendency, measures of dispersion, correlation, and regression, etc.
5. **Interpretation of Data:** It is the last step in the statistical methodology. It involves statistical thinking, skill, and experience, and final conclusions from the analyzed data.

Functions of Statistics

- (i) **To simplify complex facts:** It is very difficult for an individual to understand and conclude from huge numerical data. Statistical methods try to present the great mass of complex data into a simpler form.
- (ii) **To present facts in a determined form:** Quantitative facts can easily be presented in comparisons and to abstract & quantify facts. These facts are present in a definite form. For example: "Annual rate of inflation in the country is 10%" is more convenient than statements like "Prices are rising."
- (iii) **To make comparisons of facts:** Comparison is one of the main functions of statistics as absolute figures convey less concrete meaning.
- (iv) **To facilitate planning and policy formation:** Statistics help in the analysis of causes of economic problems and facilitate the formulation of policies for solving them.
- (v) **To help in forecasting:** As business is full of risks and uncertainties, correct forecasting is essential to reduce the uncertainties of business.
- (vi) **Formulation and Testing of Hypothesis:** Statistical methods are extremely useful in formulating and testing hypotheses.
- (vii) **To enlarge individual knowledge and experience:** Statistics enable people to enlarge their horizon. It sharpens faculties of rational thinking and reasoning and helps in propounding new theories and concepts.

Importance of Statistics

(i) Importance to Government:

In the present scenario, the government collects the largest amount of statistical data. This is required to fulfill the welfare objectives and for planning administration, making effective policies based on statistical methods. Such as index numbers, services analysis etc.

Importance in Economics

Every branch of Economics takes support from statistics in order to prove various economic theories. For example, the cost and agricultural production.

Importance of Statistics in Economic Planning

- At every stage of economic planning, we require statistics.
- Using statistical techniques, it is possible to assess the amount of various resources available in the economy and accordingly determine whether the specified rate of growth is sustainable or not.
- Study of Market Structure - Study of perfect competition, oligopoly, monopoly requires statistical comparison of market prices, costs, and profits of individual firms.

Importance of Statistics in Business

- (i) **For establishing a business unit** - Before starting a business, it is necessary to know its feasibility. It involves detailed information about location, size of output, availability of inputs, taxes, size of markets, etc., which may help to prove many key decisions.
- (ii) **For marketing strategy** - Before a product is launched, market research is a must. It makes use of various statistical techniques to analyze data.
- (iii) **For estimating the demand of product** - After launching of the business the next step is to estimate the present as well as the future demands of the product.
- (iv) **For production planning** - The businessman has to plan his production so that he is able to meet the demand of his product and ensure minimum losses on account of over or under production.
- (v) **For making quality control** - Statistical techniques can also be used to control the quality of the product manufactured by a firm.

Limitations of Statistics

- (i) **Statistics does not study qualitative phenomena** - Statistics can be applied in studying only those problems which can be expressed quantitatively or with characteristics such as honesty, poverty, welfare, beauty, patriotism cannot directly be measured in statistics.

- (ii) **Statistics does not deal with individuals** - Statistics deal only with aggregates of facts and no importance is attached to individual items.
- (iii) **Statistics can be misused** - It can be misused by wrongly motivated persons. Any person can misuse statistics and draw any type of conclusion he likes.
- (iv) **Statistical results are true only on average**—The results are true on an average as it is affected by large number of causes. Statistical laws are not universally true.
- (v) **Statistical laws are not exact**— As statistical laws are probabilistic in nature. It is based on only approximate and not exact.
- (vi) **Only experts can make the best possible use of statistics**— The techniques of statistics are not so simple to be used by any layman. These techniques can only be used by the experts as they are complicated in nature.

EXERCISE

Assertion— Reason questions

- (1) **Assertion:** Economics is the study of man in the ordinary business of life.

Reason: Economic activities are those that are undertaken for monetary gain.

Answer: (b) Both Assertion (A) and Reason (R) are true and Reason (R) is not the correct explanation of Assertion (A)

- (2) **Assertion:** Alternative uses of resources give rise to the problem of choice.

Reason: The resources which the producers have are unlimited.

Answer: (c) Assertion (A) is true but reason (R) is false.

- (3) **Assertion:** Economics only uses quantitative data

Reason: Qualitative facts can not be measured in quantitative terms.

Answer: (d) Assertion (A) is false but Reason (R) is true

- (4) **Assertion:** A single figure does not include statistics.

Reason: Single figures are unrelated and cannot be compared.

Answer: (A) Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A)

- (5) **Assertion:** Misuse of statistics is not possible.

Reason : Statistics may prove true what actually is not true.

Answer (d) Assertion (A) is false but Reason (R) is true.

One Word

Q. 1. Define economics according to Marshall.

Answer :According to Alfred Marshall, Economics is a study of man in the ordinary business of life.

Q.2. Define statistics in singular sense.

Answer: Statistics in singular sense means the collection, organization, presentation, analysis and interpretation of numerical data.

Q.3. Define the term scarcity in Economics.

Answer: Economics is the science which studies human behaviors as a relationship b/w ends scare means which have alternative uses.

Q. 4. Define economic activity with one example.

Answer: It refers to those activities which are undertaken to earn a living. Example — Teacher teaching in a school.

Q.5. Give the name of the book written by Adam Smith.

Answer: An inquiry into the nature and causes of the wealth of Nations which popularly known as wealth of nation.

Q. 6. State two limitations of Statistics.

Answer Two limitation are

- (a) It can be misused.
- (b) Results are true only on average.

Q.7. Who has given the growth oriented definition of Economics?

Answer: Prof. Anthony Samuelson has given the growth oriented definition of Economics.

Statement Based Questions

Q. 1. Statement 1 : Economic activities are undertaken to get satisfaction.

Statement 2: Economic activities are concerned with production, consumption and distribution of goods and services.

- (a) Both the statements are true.
- (b) Both the statements are false.
- (c) Statement 1 is true and statement 2 false.
- (d) Statement 2 is true and statement 1 is false.

(d) Answer

Q.2. Statement 1: Economics is concerned with the selection of resources under conditions of scarcity.

Statement 2: Scarcity of resources exist only in case of developing economics.

- (a) Both the statements are true.
- (b) Both the statements are false.
- (c) Statement 1 is true and statement 2 is false
- (d) Statement 2 is true and statement 1 is false.

(C) Answer

Q.3. Statement 1: Service provider is a person who works for some other person and get paid for it in the form of wages.

Statement 2: Service holder is a person who provides some kind of services to other for cr. Payment, like banker.

- (a) Both the statements are true.
- (b) Both the statements are false.
- (c) Statement 1 is true and statement 2 is false.
- (d) Statement 2 is true and statement 1 is false.

(b) Answer

Q.4. Statement 1 : Statistics in plural sense is descriptive in nature.

Statement 2: A high degree of accuracy is not insisted upon in statistics as mass of data is involved.

- (a) Both the statements are true.
- (b) Both the statements are false.
- (c) Statement 1 is true and statement 2 is false.
- (d) Statement 2 is true and statement 1 is false.

(A) Answer

Short Answer questions (Self Practice)

1. Distinguish between quantitative and qualitative data.
2. Explain any three functions of statistics
3. Discuss the importance of statistics in Business.
4. How is statistics importance for studying economics?
5. "Economics is a study of scarcity." Discuss.
6. Discuss the three main economic activities.
7. Discuss the nature of economics.

Multiple Choice Questions

1. **'Wealth Definition' of Economics was given by:**

(a) Adam Smith	(b) Samuelson
(c) Alfred Marshall	(d) Robbins

2. **The statistics is formed by:**

(a) A single figure	(b) An absolute descriptive fact
(c) a group of homogeneous facts and figures	(d) a group unrelated facts

3. **Statistics is:**

(a) An art	(b) a science
(c) both a science as well as an art	(d) neither a science nor an art

4. **Statistic in plural sense is:**

(a) Collection of data	(b) Organization of data
(c) Presentation of data	(d) None of the above

5. **Statistics is concerned with:**

(a) Qualitative information	(b) Quantitative information
(c) (a) or (b)	(d) Both (a) and (b)

6. **Statistics is applied in:**

(a) Economics	(b) Business management
(c) Commerce and industry	(d) All these

7. **Statistic is defined in terms of numerical data in the:**

(a) Singular sense	(b) Plural sense
(c) Either (a) or (b)	(d) Both (a) and (b)

8. **The word statistics is derived from the Latin word:**

(a) Status	(b) Stats
(c) Statistical	(d) Stata

9. Scarcity of resources applies to all:

- (a) Individuals
- (b) Organizations
- (c) Countries
- (d) All of the above

10. Welfare definition of Economics has been given by:

- (a) Lionel Robbins
- (b) Adam Smith
- (c) Alfred Marshall
- (d) Prof. Samuelson

11. Which definition was given by Adam Smith?

- (a) Welfare Definition
- (b) Wealth Definition
- (c) Scarcity Definition
- (d) Growth Oriented Definition

12. Which economic activity deals with the use of goods and services for the satisfaction of human wants?

- (a) Consumption
- (b) Exchange
- (c) Production
- (d) Distribution

13. Who expressed the view that "Economics is neutral between ends"?

- (a) Robbins
- (b) Marshall
- (c) Pigou
- (d) Adam Smith

14. The law of scarcity:

- (a) Does not apply to rich, developed countries
- (b) Applies only to the less developed countries
- (c) Implies that consumers want will be satisfied in a socialistic system
- (d) Implies that consumer's wants will never be completely satisfied.

15. Which of the following is a Non-economic Activity?

- (a) Housewife cooking food for her family
- (b) Doctor attending to patients in his clinic
- (c) Worker working in a factory
- (d) Chef cooking food at a Restaurant

16. Basis reason for existence of economic problem is:

- (a) Unlimited Wants (b) Scarcity
(c) Alternative uses (d) None of the above

17. What implication (s) does resource scarcity has for the satisfaction of wants?

- (a) Not all wants can be satisfied
(b) We will never be faced with the need to make choices
(c) We must develop ways to decrease our individual wants
(d) The discovery of new natural resources is necessary to increase our ability to satisfy wants

18. Which economist shifted the emphasis of economics from wealth to welfare?

- (a) Adam Smith (b) Samuelson
(c) Alfred Marshall (d) Lionel Robbins

19. The activity which is related to the use of scarce resources is called:

- (a) Non-economic Activity (b) Economic Activity
(c) Both (a) and (b) (d) Neither (a) nor (b)

20. Resources to satisfy human wants have _____ uses.

- (a) Alternative (b) No
(c) Few (d) None of the above

ANSWERS

1	2	3	4	5	6	7	8	9	10
A	C	C	D	B	D	B	A	D	C
11	12	13	14	15	16	17	18	19	20
B	A	A	D	A	B	A	C	B	A

Fill in the Blanks

1. A mother teaching her child is a _____ activity, While a teacher teaching the students is an _____ activity.
2. _____ definition of Economics has been given by Lionel Robbins.
3. _____ is an economic activity which deals with use of goods and services for satisfaction of human wants.
4. _____ activities are undertaken to earn a living.
5. In _____ activities, there is no expectation of any kind of monetary reward.
6. _____ it that economic activity which studies how income generated is distributed among the factors of production.
7. _____ is the root of all Economic Problems.

Fill in the blanks answers

1.	Non-economic, Economic	5.	Non-economic
2.	Scarcity	6.	Distribution
3.	Consumption	7.	Scarcity
4.	Economic		

NCERT Questions:—

1. Move at list of activities in a bus stand. How many of there are economic activities.

Answer:

- (i) Shops selling goods
 - (ii) Buses waiting for passengers
 - (iii) Conductor Selling Tickets
 - (iv) People sharing opinions
- (i) and (iii) are economic & (ii) and (iv) are non-economic.
2. “You have unlimited wants and limited resources the satisfy them,” Explain this statement by giving two examples.

Answer

Human wants are never ending, i.e. they can never fully satisfied. As soon as one want is satisfied, another new want

3. How will you choose the wants to be satisfied?**Answer**

Human wants differ in priorities, i.e. all wants are not of equal intensity. For every individual, some wants are more in portent and urgent as compared to others. As a result, we allocate our scarce resources in order of preference to satisfy some of our wants. So, we make choice between the alternative uses of the given

4. What are your reasons for studying economics?**Answer**

Economics enables us to understand different needs of the economy. However, the main reason to study of economics is scarcity, have economics is over covered with section of resources under conditions of scarcity.

NCERT Questions (Self Practice)

1. How will you choose the wants to be satisfied?
2. Statistical methods are no substitute for common sense. Comment
3. What are your reasons for studying Economics?
4. In a village of 200 farms, a study was conducted to find the cropping pattern. Out of the 50 farms surveyed 50% grew only wheat. Identify the population and the sample here.
5. Make a list of activities that constitute the ordinary business of life. Are these economic activities?
6. The Government and policy makers use statistical data to formulate suitable policies of economic development. Illustrate with two examples.
6. Mark the following statements as true or false:
 - (i) Statistics can only deal with quantitative data.
 - (ii) Statistics solves economic problems.
 - (iii) Statistics is of no use to Economics without data.
7. You have unlimited wants and limited resources to satisfy them. Explain by giving two examples.

[Chapter – 2]

CENSUS AND SAMPLE METHOD

❖ Concept of Census:

- ☞ It refers to an aggregate of items to be studied for an investigation.
- ☞ We can also say that it is a method in which **data are collected covering every item** of the universe relating to the problem under investigation.
- ☞ It is also known as '**Universe**', '**population**' and '**Complete Enumeration**'.
- ☞ It is the aggregate of **all items** about which we want to **obtain information**.

• Suitability:

1. Area of Investigation is very small.
2. Diverse items in the population and each have distinctive significance.
3. Requirement of intensive examination
4. High degree of accuracy and reliability is needed

• Merits

1. High degree of **reliability and accuracy**.
2. Results are **less biased**.
3. Extensive or **detailed information** are available of universe.
4. **Diverse characteristics** are studied
5. More important for **complex investigation** of data.

• Demerits:

1. **Costly or expensive** process of investigation.
2. **Time Consuming** process
3. Large **Qualified and skilled Man power** is required.
4. **Not suitable for Large investigation**

❖ **Sample Method:**

- A method of data collection or investigation in which data is collected about the **sample on a group of items taken from the population** for examination and conclusion are drawn on their basis.
- Sampling is the selection of a **subset or a statistical sample** of individuals **from within a statistical population** to estimate characteristics of the whole population.
- Widely used method of investigation in day to day life.

• **Suitability:**

1. Widely used when the **Area of investigation is very large**.
2. High degree **accuracy and reliability is not required**.
3. **Intensive and detailed examination is not required**.
4. **Lack of diverse characteristics** in the universe.

• **Merits:**

1. **Economical** or less expensive method of investigation.
2. **Time saving** process of investigation
3. **Errors** can be easily **identified**.
4. Relatively **larger area** it can cover for investigation.
5. **Administrative convenience** in handling a limited number of Data.
6. **Scientific technique** of investigation.

• **Demerits:**

1. **Partial investigation** of universe.
2. **Wrong conclusion** can be drawn when wrong subset of investigation is selected.
3. **Difficulty in selecting representative** sample.
4. **Bias** can be possible
5. **Difficult** when universe has **diverse characters**.
6. **Specialised knowledge is required** to conduct sampling (not useful for non-skilled person)

❖ **Essentials of a Sample:**

1. **Representative:** A sample must represent to all characters of the universe.
2. **Independent:** each unit of a sample must be independent from each other.
3. **Homogeneity:** if more than one sample are selected from a universe, these samples should be homogeneous to each other.

4. **Adequacy:** items in the sample should be fairly adequate so the investigator can draw a right conclusion.

❖ **Methods of Sampling:**

(I) **Random Sampling:**

- The method of sampling in which each and **every item** of the universe has **equal chance of being selected in the sample**.
- This method is used when **various items** of the universe are **homogeneous or identical**.
- **Merits:**
 - 1) **Free from personal bias** of investigator
 - 2) Each item of universe has **equal chance to select in sample**.
 - 3) **Fair representation** of universe
 - 4) **Simple and straightforward** method of sampling
- **Demerits:**
 - 1) Method does **not represent the proportion** of different items
 - 2) Method does **not give weightage** to certain important items
 - 3) **Not accurate** at the time when universe has **diverse characteristics**

➤ **Types of Random sampling:**

i) **Lottery method:**

- ☞ Paper – slips are made for each item of the universe and the **slips are shuffled** in a box then some of the **slips are drawn to form a sample** of the universe.
- ☞ This method is suitable when **universe not so much large**.

ii) **Tables of random numbers:**

- ☞ A sample is framed with reference to the tables.
- ☞ Tippet's table is most widely used.
- ☞ Using 41600 figures, tippet has involved 10,400 numbers comprising of four units.

Exit Polls:

- A percentage of the people exiting the poll booths are contacted and asked whom they voted for.
- From the sample, a prediction is made about the victory chances of different candidates

(II) Non-Random Sampling:

- All those methods of sampling in which **all the items of universe have not equal chance** to being selected as sample.
- **Convenience** and **personal judgement** have vital role to **select non-random sample**.

i) Purposive or Deliberate sampling:

- ☞ The investigator himself makes the **choice of the sample** items which in his **opinion are the best representative** of the universe.
- ☞ It is simply made by **choice of investigator**.
- ☞ This method is suitable when some of the **items have special significance** in the investigation.
- ☞ **Merits:**
 - Flexible to choose the items of special significance.
 - Deliberately tuned to the purpose of study.
 - Simple technique of selecting sample
- ☞ **Demerits:**
 - Possibility of personal bias.
 - Reliability of result is doubtful

ii) Stratified sampling:

- ☞ In this method, population is divided into **different groups (Strata) having different characteristics** and some the **items are randomly selected from each group** for sample.
- ☞ Method is suitable when the population has different groups with different characteristics.
- ☞ This method is also known as **Mixed technique of sampling** because it consists the features of both purposive sampling and random sampling.
- ☞ **Merits:**
 - Covers **diverse characteristics** of the population.
 - Comparative analysis** of the data becomes possible.
 - Reliable** as well as **meaningful results**.
- ☞ **Demerits:**
 - Only used by the investigator who have **complete knowledge** about the diverse characteristics.
 - Personal bias** is possible in two different strata.
 - It is **not useful for small size of population** because construction of the strata is not possible

iii) **Systematic sampling:**

- ☞ In this method units of the **population are numerically geographically and alphabetically arranged** and then every **nth item** is of the numbered items is **selected as a sample**.
- ☞ It is also known as **short-cut way of random sampling**.
- ☞ **Merits:**
 - i. Simple and easily determined.
 - ii. Hardly any possibility of personal bias.
- ☞ **Demerits:**
 - i. Not equal chance of each item being selected
 - ii. If items have identical characteristics, then this method doesn't have significance.

iv) **Quota sampling:**

- ☞ In this method population is divided in to different groups or classes according to their different characteristics and **some percentage from different group is fixed to be selected**.
- ☞ The investigator selects the **fixed number of items from each group** to frame a sample.
- ☞ **Merits:**
 - I. Not an expensive technique
 - II. Not possibility of personal bias
- ☞ **Demerits:**
 - I. Reliability of conclusions is doubtful
 - II. Specific knowledge is required

v) **Convenience sampling:**

- ☞ Sampling is done by the investigator in such a manner that **suits his convenience**.
- ☞ This method is the **simplest least expensive**.
- ☞ **Unscientific and unreliable** method of sampling.

❖ **Reliability of Sampling Data:**

- (I) **Size of Sample:** If size of sample is too small the conclusion can be doubtful so the size of sample should be adequate.
- (II) **Method of Sampling:** Each method of sampling has its own set of merit or demerit, so the suitable method should be opted according to investigation.
- (III) **Bias of correspondents and enumerators:** bias should be as less as possible because it hampers the reliability of investigation

(IV) Training of enumerators: the enumerator should be expert or well trained so he can choose the appropriate method of sampling.

❖ **Comparison between census and sample:**

Basis	Census	Sampling
Coverage	Covers all the items of universe	Only few items of universe is covered
Suitability	Area investigation is small	Area of investigation is large
Accuracy	High degree of accuracy	Low degree of accuracy
Cost	Highly Expensive	Less expensive
Time	Highly time consuming	Low time consuming
Nature of items	Diverse characteristics	Relatively identical characteristics
Verification	Either not possible or difficult	Easily verifiable

❖ **Statistical errors:**

1. **Sampling Error:**

- When the errors are related with the size and nature of the sample selected for the study.
- Either the sample size is too small then it becomes non-representative or wrong characteristics are chosen for the sample.

$$\text{Sampling Error} = \text{estimated value} - \text{true value}$$

2. **Non-Sampling Error:**

- The errors which **belong with collection of data**. Various type of non sampling errors:

i) **Error of measurement:**

- a) Difference in the scale of measurement
- b) Difference in the rounding off procedure adopted by different investigation

ii) **Error of non-responsive:**

The respondents do not offer the required information.

iii) **Error of misinterpretation:**

Respondent fails to interpret the questions in the questionnaire

iv) **Error in data Acquisition:**

Errors occur when responses are recorded incorrectly. Example: 05 is recorded as 50.

v) **Error of sampling bias:**

Targeted population is not being selected as sample.

Note: It is noted that non-sampling errors are more serious than sampling errors because sampling errors are minimised by enlarging the size of sample but it is not possible to cop-up in case of non-sampling errors.

EXERCISE

Assertion/Reason Based

1. **Assertion (A):** Primary data is original
Reason (R): Primary data is collected by the investigator himself
Ans (a): Both Assertion (A) & Reason (R) are true and Reason (R) is the correct explanation of Assertion (A)
2. **Assertion (A):** Sampling Method is an economical method
Reason (R): Under Sampling, analysis of data is confined only to a fraction of the population.
Ans (b): Both Assertion (A) & Reason (R) are true and Reason (R) is not the correct explanation of Assertion (A)
3. **Assertion (A):** Census cannot contain non-sampling errors
Reason (R): It is difficult to minimize non-sampling errors.
Ans (a): Both Assertion (A) & Reason (R) are true and Reason (R) is the correct explanation of Assertion (A)
4. **Assertion (A):** The questionnaire should not be too long
Reason (R): Long questionnaires discourage people from completing them.
Ans 4 (a) Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).
5. **Assertion:** Each value of a variable is an observation.
Reason: Variables vary over a period of time.
Ans 5 (a) Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).
6. **Assertion:** Sample survey includes every element of the population.
Reason: Most of the surveys are sample surveys.
Ans 6 (d) Assertion (A) is false but Reason (R) is true.
7. **Assertion:** The results of Indirect Oral Investigation are very accurate.
Reason: In case of Indirect Oral Investigation, information is obtained from sources.
Ans 7 (d) Assertion (A) is false but Reason (R) is true.

One Word

Q1 What is the meaning of Secondary data?

Ans: Data which are not directly collected but rather obtained from the published or unpublished sources are called Secondary Data.

Q2 State two merits of Primary data.

Ans: (i) High degree of accuracy
(ii) Need of less caution

Q3 In which type of data, precaution is highly required?

Ans: Secondary data

Q4 What is direct Personal investigation?

Ans: It is that method of investigation in which data are collected personally by the investigators by asking questions relating to enquiries.

Q5 What do you mean by a questionnaire?

Ans: It is a list containing a number of questions pertaining to the investigation.

Q6 When the multiple choice questions (MCQs) should be used?

Ans: When more than two answers are possible for a given question.

Q7 What do you mean by enumerators?

Ans: Enumerators are the persons who put up certain questions to the respondents and fill their answers in the questionnaire.

Q8 What is meant by sampling method?

Ans: It is a method of investigation in which only a few units of the universe are selected and studied.

Q9 What is Stratified Sampling?

Ans: It is a method of sampling in which universe is divided into different strata.

Q10 State the meaning of Purposive Sampling.

Ans: A method in which the investigator himself makes the choice of the sample items which in his opinion are best representative of the Universe.

Statement Based Questions

Question 1

Statement 1: Most of the surveys are census surveys.

Statement 2: Census surveys are preferred in statistics.

Ans: Both the statements are false.

Question 2

Statement 1: Pilot survey is the pre-testing of the questionnaire.

Statement 2: Pilot survey helps in assessing the suitability of questions and clarity of instructions.

Ans: Both the statements are true.

Question 3

Statement 1: Population or the universe in statistics means totality of the items under study.

Statement 2: Population or the universe is a group to which the results of the study are intended to apply.

Ans: Both the statements are true.

Question 4

Statement 1: Lottery method is an example of random sampling.

Statement 2: In random sampling the convenience or judgement of the investigator plays an important role in selection of the sample.

Ans: Statement 1 is true and Statement 2 is false.

Question 5

Statement 1: Primary data is based on first-hand information.

Statement 2: Use of primary data saves time and cost.

Ans: Statement 1 is true and Statement 2 is false.

Question 6

Statement 1: Information from correspondents is widely used by news channels to collect news information.

Statement 2: The method of information from correspondents is suitable when regular and continuous information is required.

Ans: Both the statements are true.

Question 7

Statement 1: Indirect oral investigation is a very costly method.

Statement 2: The degree of accuracy is extremely high in case of indirect oral investigation.

Ans: Both the statements are false.

Short Answer Questions (Self Practice)

- Q1.** What do you mean by primary data? Mention the different methods of collecting primary data.
- Q2.** How can data be collected through Telephonic Interview? Mention its merit and demerit.
- Q3.** Distinguish between census and sampling methods of data collection.
- Q4.** Discuss two merits and demerits of random sampling.
- Q5.** What is meant by statistical errors? Which factors give rise to statistical errors?

Q6. What is meant by Direct Personal Investigation? Give its merit and demerits.

Multiple Choice Questions (MCQs)

- 1. Which data is collected by the investigator himself?**
 - (a) Primary
 - (b) Secondary
 - (c) Both (a) and (b)
 - (d) None of these
- 2. Data collected from 'The Times of India' is an example of:**
 - (a) Primary Data
 - (b) Secondary Data
 - (c) None of these
 - (d) Census
- 3. Data collected on religion from the census reports are:**
 - (a) Sample data
 - (b) Secondary data
 - (c) Primary data
 - (d) Either (a) or (b)
- 4. When population under investigation is infinite, we should use:**
 - (a) Sample Method
 - (b) Census Method
 - (c) Either (a) or (b)
 - (d) Neither (a) nor (b)
- 5. Data from secondary source are:**
 - (a) Collected for other purposes than the current study
 - (b) Obtained from the newspaper
 - (c) More reliable than data from a primary source
 - (d) Both (a) and (b)
- 6. Primary data is preferred over secondary data where:**
 - (a) Time available is short
 - (b) Accuracy is important
 - (c) Sufficient finance is not available
 - (d) Much accuracy is not required
- 7. The data collected on the height of a group of students after recording their heights with a measuring tape are:**
 - (a) Primary data
 - (b) Continuous data
 - (c) Discrete data
 - (d) Secondary data

- 8. After every ten years, information regarding population of India is collected through:**
- (a) Census
 - (b) Sample
 - (c) Both (a) and (b)
 - (d) Neither (a) nor (b)
- 9. Sample study is useful:**
- (a) When population is not completely known
 - (b) As it is easy to handle samples
 - (c) As results are more reliable
 - (d) It is cheap
- 10. Which method of collection of data covers the widest area?**
- (a) Direct Personal Investigation
 - (b) Mailed Questionnaire Method
 - (c) Telephone interview method
 - (d) All of these
- 11. Which of the following is a limitation of Indirect Oral Investigation?**
- (a) Lack of Accuracy
 - (b) Costly
 - (c) Not suitable for wide areas
 - (d) All of these
- 12. It is best to use a census while conducting a survey if:**
- (a) The population is large
 - (b) The population is small
 - (c) Time is limited to conduct the survey
 - (d) Cheaper method is needed
- 13. A good questionnaire should have:**
- (a) Minimum questions
 - (b) Concise
 - (c) Clear
 - (d) All the above
- 14. In random sampling:**
- (a) Each element has equal chance of being selected
 - (b) Sample is always full of bias
 - (c) Cost involved is very less
 - (d) Cost involved is high

- 15. Direct personal investigation method suffers from:**
- (a) Personal Bias
 - (b) Excessive Expenses
 - (c) Time Consuming
 - (d) All the above
- 16. In a good questionnaire:**
- (a) Number of questions should be numerous
 - (b) Questions relating to mathematical computations should be asked
 - (c) Personal Questions should be preferred
 - (d) Questions using double negatives should be avoided
- 17. The quickest method to collect primary data is:**
- (a) Direct Personal Investigation
 - (b) Indirect Oral Investigation
 - (c) Telephone Interview
 - (d) Mailed Questionnaire Method
- 18. Stratified sample is preferred where:**
- (a) Population is perfectly homogeneous
 - (b) Population is non-homogeneous
 - (c) Random sampling is not possible
 - (d) Small samples are required
- 19. The method of sampling in which universe is divided into a number of groups and then certain numbers of items are taken from each group at random.**
- (a) Stratified Random Sampling
 - (b) Systematic Sampling
 - (c) Cluster Sampling
 - (d) Multistage Sampling
- 20. Which of the following is the method of collecting primary data?**
- (a) Indirect Oral Investigation
 - (b) Telephonic Interviews
 - (c) Information from Correspondents
 - (d) All of these

Multiple Choice Questions (MCQs) Answers

1	2	3	4	5	6	7	8	9	10
A	B	B	A	D	B	A	A	A	B
11	12	13	14	15	16	17	18	19	20
A	B	D	A	D	D	C	B	A	d

Fill in the Blanks

- _____ Data is originally collected by an investigator for the first time for some specific purpose.
- _____ Sampling Method refers to a method in which every item in the universe has a known chance of being chosen for the sample.
- Under the method of _____, the investigator approaches third parties, who are in the possession of information about the subject of enquiry.
- According to _____, if a random sample of adequate size is selected from a large population, it tends to possess the same characteristics as those of the population.
- Data obtained from published or unpublished sources is known as _____ Data.
- Under _____ Sampling, the investigator gives special attention to his convenience while selecting the sample units.
- _____ errors refer to the differences between the sample estimate and the actual value of a characteristic of the population.
- Quota sampling is a _____ sampling.
- External Data can be collected from _____ as well as _____ sources.
- _____ is more suitable if population is heterogeneous in nature, while _____ is more suitable in case of homogeneous population.
- _____ is also known as Second Hand Data.

Fill In TheBlanksAnswer

1	Primary	7.	Sampling
2	Random	8.	Non-random
3.	Indirect Oral Investigation	9.	Primary, Secondary
4.	Law of Statistical Regulatory	10.	Census Method, Sampling Method
5.	Secondary	11.	Secondary Data
6.	Convenience		

〔Chapter –3〕

ORGANIZATION OF DATA

- Classification of data means grouping of related facts into classes.
- Classification is defined as “the process of dividing data into mutually exclusive classes, so that data within each class is homogeneous and data between classes is heterogeneous”.

Objectives of Classification:

- (1) **Simplification:** To simplify complex data and make it easier to understand.
 - (2) **Easy Comparison:** Facilitate comparison of data within and between groups.
 - (3) **Identify Patterns:** To identify patterns and trends in data.
 - (4) **Reduce Complexity:** Reduce the complexity of large datasets.
- **Characteristics of a Good Classification.**
 - 1) **Exhaustive :** The classification should include all the data.
 - 2) **Homogeneous:** Data within each class should be homogenous.
 - 3) **Heterogeneous:** Data within each class should be heterogeneous
 - 4) **Relevant:** The classification should be relevant to the purpose of the analysis.
 - **Basis of Classification:**
 - 1) **Geographical Basis:** Classification based on geographical location, such as region, country or city.

Example:- Classification of sales data by region

Region	Sales (in `)
North	10,000
South	8,000
East	6,000
West	12,000

Chronological Classification: When data are classified on the basis of time.

Example: Production of company.

Year	Education (in units)
2002	300
2003	400
2004	500
2005	600

3. **Qualitative Classification:** Classification based on time such as qualities or attributes, such as, gender occupation, religion and education level.

Qualitative Classification

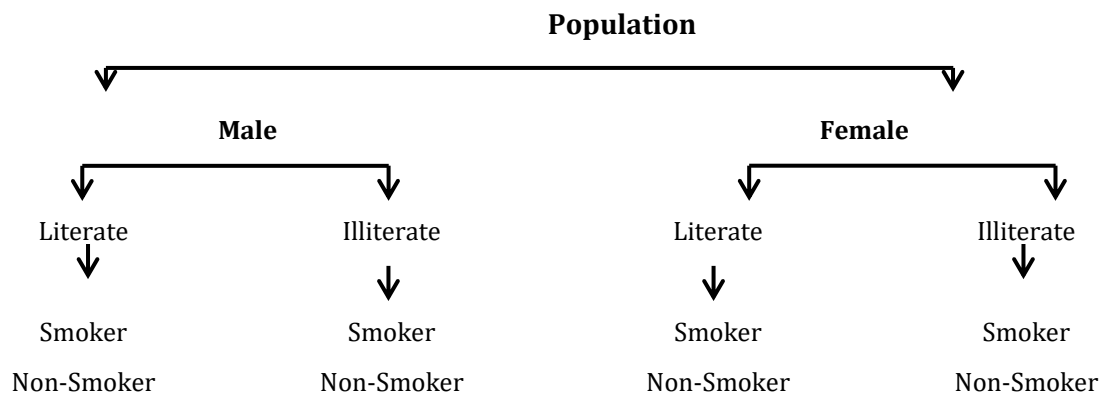
- (i) Simple Classification
- (ii) Manifold Classification

Simple Classification: it is called classification acc. to dichotomy.

The data are classified into two categories i.e. presence and absence of attributes.

Male – Female, Health – Unhealthy are example.

Manifold classification: In manifold or multi fold classification, the given data are classified into make them two categories.



4) **Quantitative or numerical classification:**–Classification based on numerical values such as height, age, income, expenditure, marks etc.

Example: Age Group and population.

Age Group	Population
20-30	4,25,000
31-40	5,35,000
41-50	6,45,000

Variable: An characteristics are a phenomenon which is capable of being measured and changes its value overtime is called a variable.

1. Types of Variable:

- 1) Discrete Variable
- 2) Continuous

(1) Discrete Variable: These are those variables that increase in jumps or in complete numbers for Example. 1,2,3,10, $1\frac{1}{4}$, $1\frac{1}{3}$ etc.

(2) Continuous Variable: Variables that assumes a range of value or increase not in jumps but continuously or in fraction.

For Example 5'1", 5'.2" and sq. on.

➤ **Raw Data:** An mass of data in its crude form called raw data. It is an unorganised mass of the various items.

40	10	15
35	5	20
25	3	35
15	2	10

➤ Conversion of raw data using statistical series.

1. Types of Statistical Series

- 1) Individual series or series without frequencies
- 2) Frequency series or series with frequencies

2. Frequency series or series with frequencies

- 1) Discrete series / Frequency Array
- 2) Frequency Distribution

1) Individual Series: These series in which the items are listed singly.

For Example :- Marks obtained by 5 students

8, 7, 5, 4, 2

(i) According to series number : The items are arranged in a serial number.

For Example — In the below table, marks obtain by students arranged according to their roll no.

Roll No.	Marks
1	5
2	7
3	8
4	4
5	2

(ii) Ascending or descending Order—In ascending order, the smallest value is placed first, while in descending order the highest value is placed first.

Data Arranged in Ascending Order

5	15	25	40
5	15	30	40
10	20	345	45
15	25	40	50

Data Arranged in Descending Order

50	40	25	15
45	35	20	10
40	30	15	5
40	25	15	5

Frequency Series:

(a) Discrete Series or Frequency Array

(b) Frequency Distribution

(a) Frequency—

Frequency is the number of items on items occurs c or repeats itself in the series.

(b) Class Frequency —

The number of items an items an item repeats itself corresponding to a range of value [or

class interval] is called class frequency.

For Example —If 3 students scoring makes between 10-15 then 3 will be class frequency.

Tally Bars

Marks	Tally Bars	Frequency
10	II	4
15	I	9
20	III	8
25	HHH	6
30	I	7

(i) Discrete series or frequency Array: That series in which data are presented in a way that enact measurement of items of items are clearly shown.

Marks	Tally Bars of the frequency	Frequency (Total)
11	III	3
12	HHH	5
13	IIII	4
14	II	2

ii. Frequency distribution: Data are classified under different classes with a range is called class interval.

Marks	Tally Bars	Frequency
15–20	IIII	4
20–25	III	3
25–30	I	1
30–35	II	2

Types of Frequency Distribution

1. Exclusive series
2. Inclusive series
3. Open end Series
4. Cumulative Frequency series

5. Mid-values frequency series

- 1) Exclusive Series:** In this series the upper limit of one class interval is the lower of the next class interval. It is allowed exclusive because frequencies of the upper limit of each class interval is not included in that class.

Marks	Frequency
10-15	8
15-20	4
20-25	5
25-30	8
Total= 25	

Inclusive series: Inclusive series is that series which include all items up to its upper limit.

Upper limit of the class interval class not repeat in the non-class interval as lower limit

Inclusive Series

Marks	Frequency
5 – 9	5
10-14	8
15-19	4
20-24	3
Total	20

Conversion of inclusive series into exclusive series

1. Half of the difference is added to the upper limit and half is subtracted from the lower limit. Conversion of the above inclusive series into on exclusive series.

Marks	Frequency
4.5 – 9.5	5
9.5 – 14.5	8
14.5-19.5	4
19.5 – 24.5	3

➤ **Difference b/w exclusive & inclusive series**

Base	Exclusive Series	Inclusive Series
Limit	Upper limit of one class is always equal to the lower limit of next class.	There is difference b/w UL of one & lower limit of next class
Inclusion	The upper limit of the class is not included	The UL of the class is included.
Conversion	Not need to convert into inclusive series	Need to convert into exclusive for calculation process.
Suitability	Suitable for all cases (in complete number or in fraction)	Suitable only for the value is a complete number.

3. **Open-end series:** It is that series in which lower limit of the first class interval and (or) the upper limit of last class interval is missing.

Open end series

Marks	Frequency
Below 5	1
5 -10	3
10-15	4
15-20	6
20 and above	1

Cumulative frequency series: CF series is that series in which the frequencies are continuously added corresponding to each class interval in the series.

2. Types of Cumulative Frequency.

- 1) Less than Cf
- 2) More than Cf

For example:—

Marks	No of students
0-5	9
5-10	5
10-15	4
15-20	3

'Less than' cumulative frequency

Marks	No. of students (CF)
Less than 5	9
Less than 10	14 (9+5)
Less than 15	18 (14+4)
Less than 20	21 (18 +3)

Marks	No. of students (CF)
More than 0	9
More than 5	5
More than 10	4
More than 15	3

- 5. Mid-Value frequency Series:** Mid values frequency series are these series in which we have only mid-class of the class intervals & the cross-ponding frequencies.

Mid Value	Frequency
10	9
15	8
20	11
25	12
30	10

Solved Practical's

(1) Following Descending order

Ascending Order		Descending Order	
Marks	Frequency	Marks	Frequency
3	2	25	1
4	1	22	3
5	1	21	1
6	2	19	2

7	1	18	4
8	3	17	2
9	1	16	2
10	4	15	2
11	2	14	4
12	1	13	1
13	1	12	1
14	4	11	2
15	2	10	4
16	2	9	1
17	2	8	3
18	4	7	1
19	2	6	2
21	1	5	1
22	3	4	1
25	1	3	2
	Total =40		Total =40

(2) Heights Series

Heights	55	58	60	62	65	66	68	72	
Tally Marks	HHH II	HHH H	HHH	I	I	III	HHH	HHH I	
Frequency	6	6	5	1	1	3	5	6	=35

(3) The following so on.

C-I	Frequency
10-20	4
20-30	5
30-40	7
40-50	7
50-60	4

60-70	3
-------	---

(4) Following statistics

0-10	0
10-20	2
20-30	3
30-40	6
40-50	12
50-60	14
60-70	9
70-80	2
80-90	2
Total	50

5. Prepare Data

C.I.	Tally Marks	Frequency
20-24	III	3
25-29	HHH	5
30-34	HHH-III	8
35-39	HHHHH I	11
40-44	HHH III	8
45-49	HHH-II	7
50-54	HHH II	7
55-59	1	1
		= 50

6. Fromlast class.

Size of each class = 20

First class = 100-120

The last class = 180 - 200

First class interval = $120 - 20 = 100$

Upper series

7. **Calculate the missing Distribution**

Answer: As the class intervals are not series C.I. cannot be

8. Convert the following distribution.

C.I.	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	5	12	23	29	24	19	10

Less than Cumulative frequency.

C.I.	Frequency
Less than 20	5
Less than 30	17
Less than 40	40
Less than 50	69
Less than 60	93
Less than 70	112
Less than 80	122
Less than 90	124

(9) Convert the following frequency series

Answer:

C.I.	Frequency
0-20	10
20-40	$18 - 10 = 8$
40-60	$25 - 18 = 7$
60-80	$45 - 25 = 20$
80-100	$55 - 45 = 10$

(10) Prepare frequencies data

Answer : Value for adjustment = $\frac{35-25}{2} = 5$

Upper limit = $25 + 5 = 30$

$$\text{Lower limit} = 25 - 5 = 20$$

Frequency distribution

Mid Value	C.F.	Frequency
25	20-30	6
35	30-40	10
45	40-50	12
55	50-60	12
65	60-70	10
		$\Sigma f = 50$

11. In a series 5-9, 10-14 etc.

Answer:

Quantities	Tally Marks	F
5-9	III I	6
10-14	IIII I	11
15-19	IIII II	13
20-24	IIII III	18
25-29	III III	8
30-34	III	5
35-39	III	3
40-44	0	-
		$\Sigma f 64$

You are given series

Answer - Value of adjustments $= \frac{25-15}{2} = 15$

L.L of F.C. = $15 - 5 = 10$

U.L. of F.C. = $15 + 5 = 20$

Mid value	Class Interval	Frequency
15	10-20	8
25	20-30	12
35	30-40	15
45	40-50	10
55	50-60	5
		$\Sigma f = 50$

13. For the data given etc.

Answer:

C.I.	100-110	110-120	120-130	130-140	140-150	150-160
Tally Marks	I	IIII	HHH	HHHH	HHI	IIII
F	1	4	5	10	6	4
					$\Sigma F = 30$	

EXERCISE

Assertion Reason Based

1. Assertion (A) : Nationality is an attribute.

Reason : (R) : Nationality cannot be measured.

Ans. (a) Both assertion (A) and reason (R) are true and reason (R) is the correct explanation of Assertion. (A)

2. Assertion (A): For a discrete variable the classification of its raw data is known as a frequency array.

Reason (R): Discrete variable takes all the values of observations.

Ans. (C) Assertion (A) is true but Reason (R) is false.

3. Assertion(A) : Classification is arranging, things into groups or classed based on some criteria.

Reason (R): Classification brings order to raw data.

Ans. (b) Both assertion (A) and Reason (R) are true and Reason (R) is not the correct explanation of Assertion (A)

4. Assertion (A): In case of inclusive series the value of upper limit of a class never equals the value of lower limit of the next class.

Reason (R) Class frequency are same in both exclusive and inclusive series, although the class intervals are apparently different in two cases.

Ans. (b) Both assertion (A) & reason (R) are true & reason (R) is not the correct explanation of assertion (A)

5. **Assertion (A):** Exclusive series ensures continuity of data.

Reason (R) : In case of exclusive series upper limit of one class is the lower limit of succeeding class.

Ans. (a) Both Assertion (A) & Reason (R) are true & Reason (R) is the correct explanation of Assertion (A)

One Word

1. **What is quantitative classification?**

Ans. When data is classified on the basis of the characteristics which can be measured it is known as quantitative classification.

2. **Mention the types of statistical series on the basis of Construction.**

Ans. (i) Individual series (ii) Discrete series (iii) Continuous series

3. What are class limits?

Ans. Upper limit and lower limit of a class are called class limits.

4. What is meant by mid-point?

Ans. It is the central point of a class-interval which is half way between lower and upper class limits.

5. What is meant by the term frequency?

Ans. It refers to number of times a given value appears in distribution.

6. What do you understand by raw data?

Ans. A mass of data in its original form is known as raw data.

7. Name the series which have class-interval.

Ans. Continuous series

8. What is meant by exclusive series?

Ans. The series in which the upper of preceding class becomes the lower limit of the succeeding class is known as exclusive series.

9. What is meant by Time series?

Ans. If the different values that a variable has taken in a period of time are arranged in a chronological order, the series is called a time series.

10. What is open-end series?

Ans. In a frequency distribution, if the lower limit of the first class and the upper limit of last class is not given, it is known as open-end series.

Statement Based Questions

1. Statement 1: Classification of data on attributes is called qualitative classification.

Statement 2: Attributes are expressed quantitatively.

Ans. Statement 1 is true and statement 2 is false.

2. Statement 1 : Discrete variable can take any numerical value.

Statement 2: Continuous variable can take only certain values.

Ans. Both the statements are false.

3. Statement 1: Class width is the middle value of a class.

Statement 2: Class mark lies halfway between the lower class limit and the upper class limit.

Ans. Statement 2 is true and statement 1 is false

4. **Statement 1: There is no class mark in a frequency array.**

Statement 2: Frequency array does not take any fractional value.

Ans. Both the statement are true.

5. **Statement 1: In case of open –end distribution, upper limit of first class and lower limit of last class is not given.**

Statement 2: Open-end-classes create problem in the graphic presentation of the data.

Ans. Statement 2 is true & statement 1 is false

Short Answer Questions (Self practice)

1. State the different method of classification.
2. What is meant by individual series? Give one example.
3. Give the meaning of discrete series with the help of an example.
4. Briefly discuss the following terms : (i) Class interval (ii) Range (iii) Frequency
5. Define 'statistical series'. State its main types.

NCERT Questions (Self Practice)

1. **Which of the following alternatives is true?**

(i) The class mid-point is equal to:

- (a) The average of the upper class limit and the lower class limit.
- (b) The product of upper class limit and the lower class limit.
- (c) The ratio of the upper class limit and the lower class limit.
- (d) None of these.

(ii) The frequency distribution of two variables is known as:

- (a) Univariate distribution
- (b) Multivariate distribution
- (c) Bivariate Distribution
- (d) None of the above.

(iii) Statistical calculations of classified data are based on:

- (a) The actual values of observations.
- (b) The upper class limits
- (c) The lower class limits

(d) The class mid-points

2. Can there be any advantages in classifying things?

Explain with an example from your daily life.

3. What is a variable? Distinguish b/w discrete and continuous variable.

4. Construct a frequency array of the output produced by a firm (in tonnes) in 20 months.

60,55,40,45,50,65,60,40,45,50,60,30,55,40,65,35,30,45,65,35.

5. Construct a frequency distribution from the following data by taking class interval of 6 (use both exclusive and inclusive methods):

69,90,76,80,63,58,51,72,71,89,57,79,79,62,66,59,52,51,48,75,90,76,70,90,58,54,78,87,70,91,67,80,68,54,81,80,45,70,90,76,91,70,71,63,54,65,89,90,62,80.

5. Following are the marks obtained in an entrance exam. Distinguish the data in the form of individual series, discrete series, continuous series and grouped cumulative frequency series.

35,35,38,36,30,31,45,45,35,36,38,32,44,45,40,38,42,45,45,43,40,35,36,37,39,38,74,2,42,45

6. Arrange the data and find range.

69, 37, 30,49, 7, 50, 23,69, 46, 44.

7. Convert the following simple frequency distribution into a cumulative frequency distribution "more than Method".

Wages (₹)	No. of Labours
0-6	8
6-12	12
12-18	7
18-24	15
24-30	20
30-36	26
36-42	19
42-48	18
48-54	20
54-60	17

8. Convert the following inclusive series into exclusive series.

Marks	7-13	14-20	21-27	28-34	35-41
No. of students	2	4	6	8	10

9. Convert the following series into cumulative frequency distribution by using less than and more than method.

Marks	8-16	16-24	24-32	32-40	40-48	48-56
No. of Students	8	6	2	4	15	5

10. Construct a grouped frequency distribution from the following data:

Wages (Mid-values)	No. of works
11	8
22	12
33	14
44	6
55	10

11. Prepare a simple frequency distribution from following data:

- 10 men get at the rate of `400 per man per day.
- 20 men get at the rate of `450 per man per day.
- 40 men get at the rate of `500 per man per day.
- 100 men get at the rate of `550 per man per day.
- 80 men get at the rate of `600 per man per day.
- 60 men get at the rate of `650 per man per day.
- 40 men get at the rate of `700 per man per day.

12. The following data is related to an survey for the number of appliances 50 families used in daily consumption:

33,21,40,41,47,14, 40,21,16,50,36,30,28,54,21, 37,41,42,37,44,30,42, 17,42,31,16,40,30, 14,32,30,30,19,17,42,37,53,30, 37,40,36,37,41,46,47,52,53,33,30,56.

- (i) Obtain the range of appliances used.
- (ii) Classify this range appropriately and obtain the frequency distribution of appliances for domestic use?
- (iii) What kind of class interval will be appropriate for this classification.
- (iv) Find out the number of families whose number of usage is
 - (a) Less than 35
 - (b) More than 42

13. Convert the following into exclusive series

(a)

X	1-9	10-19	20-29	30-39
F	8	12	10	6

(b)

X	1-4.9	5-9.9	10-14.9	15-19.9
F	8	15	6	4

(c)

X	1-9.99	10-19.99	20-29.99	30-39.99
F	12	20	16	14

(d)

Marks	Below 10	Below 15	Below 20	Below 25	Below 30
No. of students	8	17	32	42	50

(e)

Wages (₹)	Less than 100	100-200	Less than 300	300-400	Less than 500
No. of	10	8	25	5	50

Workers					
---------	--	--	--	--	--

(j)

Size	0–5	0–10	5–15	Upto 20	20 and above	25 –30
No. of workers	3	8	12	25	10	6

16. Marks obtained by 20 students of Class XI are given. Prepare a bivariate frequency table:

Marks in Economics	7	9	8	10	7	6	5	7	8	9	7	6	5	10	5	7	9	5	10	6
Marks in Accountancy	5	6	7	9	6	5	7	9	10	8	7	6	9	10	5	8	7	6	9	8

MCQs QUESTIONS

1. Which data is collected by the investigator himself?

- (a) Primary (b) Secondary
(c) Both (a) and (b) (d) None of these

1. Data collected from ‘The Times of India’ is an example of:

- (a) Primary Data (b) Secondary Data
(c) None of these (d) Census

3. Data collected on religion from the census reports are:

- (a) Sample data (b) Secondary data
(c) Primary data (d) Either (a) or (b)

4. When population under investigation is infinite, we should use:

- (a) Sample Method (b) Census Method
(c) Either (a) or (b) (d) Neither (a) nor (b)

5. Data from secondary source are:

- (a) Collected for other purposes than the current study
- (b) Obtained from the newspaper
- (c) More reliable than data from a primary source
- (d) Both (a) and (b)

6. Primary data is preferred over secondary data where:

- (a) Time available is short
- (b) Accuracy is important
- (c) Sufficient finance is not available
- (d) Much accuracy is not required

7. The data collected on the height of a group of students after recording their heights with measuring tape are:

- (a) Primary data
- (b) Continuous data
- (c) Discrete data
- (d) Secondary data

8. After every ten years, information regarding population of India is collected through:

- (a) Census
- (b) Sample
- (c) Both (a) and (b)
- (d) Neither (a) nor (b)

9. Sample study is useful:

- (a) When population is not completely known
- (b) As it is easy to handle samples
- (c) As results are more reliable
- (d) It is cheap

10. Which method of collection of data covers the widest area?

- (a) Direct Personal investigation
- (b) Mailed Questionnaire Method
- (c) Telephone interview method
- (d) All of these

11. Which of the following is a limitation of Indirect Oral Investigation?

- (a) Lack of Accuracy
- (b) Costly
- (c) Not suitable for wide areas
- (d) All of these

12. It is best to use a census while conducting a survey if:

- (a) The population is large (b) The population is small
(c) Time is limited to conduct the survey (d) Cheaper method is needed

Answer MCQs Questions

1	2	3	4	5	6	7	8	9	10
A	B	B	A	D	B	A	A	A	B
11	12								
A	B								

Missing Words

Think carefully about the most appropriate word or words to complete the given statements

- _____ data are original in nature.
- _____ of data refers to systematic recording of results either by counting or by enumeration.
- _____ is suitable for intensive field surveys.
- Under _____ method, investigator collect data by contacting with the third parties.
- _____ is generally adopted in those cases where the information is to be obtained at regular intervals from a wide area.
- To obtain different opinions from the different respondents on specific issues _____ questions should be asked.
- Personal interests of investigator does not affect the selection of the sample in _____ sampling.
- _____ errors can occur in census survey or sample survey.
- National sample Survey was set up in _____.
- The data collected by NSSO surveys are released through its quarterly journal _____.

Answer

1	2	3	4	5
Primary	Collection	Direct Personal Investigation	Indirect Oral Investigation	Information from correspondents
6	7	8	9	10.
Open	Random	Non-Sampling	1950	Sarvekshna .

〔Chapter 4〕

TEXTUAL AND TABULAR PRESENTATION OF DATA

Introduction

After classification of statistical data, it is required to arrange data in a presentable form.

3 forms of presentation of data

1. Textual or Descriptive presentation
2. Tabular Presentation
3. Diagrammatic Presentation

Textual Presentation of Data (Descriptive Presentation)

When data are described within the text.

Example: Cricket is the favorite sport of 25 students

Football is the favorite sport of 18 students.

Advantage: In this method of presentation, specific emphasis can be laid on certain items which require to be compared. Reader can understand statement with their explanations.

Drawback: It becomes difficult to identify patterns trends or comparisons when data is presented.

In Text Form

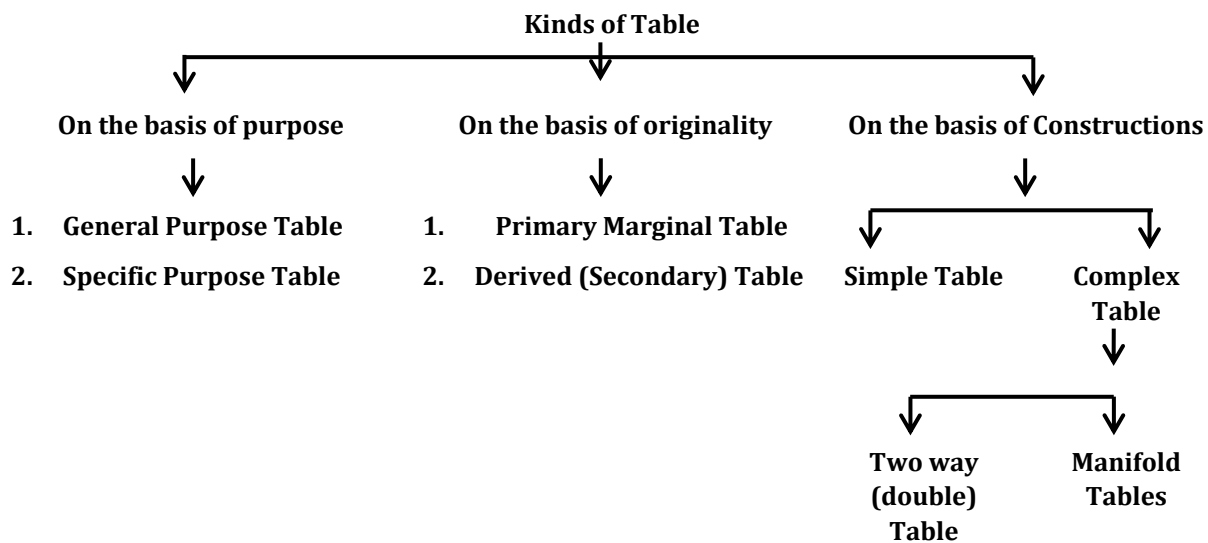
- Textual form takes more time read compared to tabular or graphical presentation.
- Presenting a large amount of data textually can lead to confusion and make it hard to analyze.

Tabular presentation of data

- Tabulation refers to systematic and logical arrangement of data in columns (vertically) and rows (horizontally) in accordance with some salient features and characteristics.

Essential Parts of Table

- 1) **Title:** An brief description of the data presented in the table.
- 2) **Headnote:** An note or explanation that provides additional information about the data.
- 3) **Box Head:** The topmost row of the table that contains the column heading.
- 4) **Column Headings:** The names or labels of the columns that describe the data.
- (5) **Rows:-** The horizontal arrangements of data that provide information about lach variable or characteristic.
- (6) **Columns:** The vertical arrangements of data that provide information.
- (7) **Captions:** Brief descriptions or explanation that provide additional information about the data in each rout or column.
- (8) **Footnote:**An note or explanation that provides additional information about the data or sources.
- (9) **Source:** The origin or source of the data presented in the table.



1) On the basis of purpose

These table are of 2 types

(a) **General purpose Table:** It is used to present data in a simple and concise manner.

- These tables can be seen in Govt. publications
- Also known As “references Tables’

(b) **Specific purpose table :** It provide specific information related to particular objective.

- These are prepared from general purpose table also known as ‘Summary Table’ ‘Analysis Table’

(2) On the basis of origin: These are of 2 types.

- (i) Primary (Original Table):** A primary table is constructed from original or raw data collected by researcher.

Also known as 'Classification Table'.

- (ii) Derived Table :** A derived table is constructed by manipulating or combining data from one or more primary or secondary tables.

Also known as 'Primary Table'.

On the basis of construction:- These are of 2 types

- 1) Simple Table (one way table)-** An table which shows only one characteristics of the data is known as simple table.

Also known as one variable table

Example :— No of students in the school

Class	No. of Students
V	50
VI	80
VII	60
VIII	40

- 2) Complex Table :** An Table which is classified on the basis of more one characteristics of the data:

This table is of following 2 types :

Double or two way table

Example or two way table

Class	No of Students		Total
	Boys	Girls	
9 th	40	32	72
10 th	30	9	39
11 th	25	25	50
12 th	27	33	60

3) Manifold Table

Example: No of students in collage.

Class	Boys			Girls			Total	
	Rural	Urban	T	Rural	Urban	T	Rural	Urban
B.A. I	10	20	30	10	30	40	20	50
B.A. II	20	30	50	15	45	60	35	75
B.A. III	30	40	70	5	15	20	35	55
Total	60	90	150	30	90	120	90	180

EXERCISE

Assertion – Reason based

1. Assertion: The title of the table has to be very clear, brief and carefully worded

Reason: The interpretations made from the table are clear and free from any ambiguity.

Ans. Both assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A)

2. Assertion : Body of a table is the main part.

Reason: Body of the table contains the actual data.

Ans. Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of assertion (A)

3. Assertion: In textual presentation, data are described within the text.

Reason : Textual presentation enables one to emphasize certain points of the presentation.

Ans. Both Assertion (A) and Reason (R) are true and Reason (R) is not correct explanation of Assertion (A).

4. Assertion: Captions refer to the headings of horizontal rows.

Reason: Source note refers to the source from which information has been taken.

Ans. Assertion (A) is false but Reason (R) is true.

5. Assertion (A) : Figures in case of original table are not rounded up.

Reason (R): The original table contains the statistical facts in their original form

Ans. Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A)

One word

1. What is meant by tabulation?

Ans. It is a systematic presentation of numerical data into rows and columns.

2. State two objectives of tabulation.

Ans. (i) To simplify the complex data.

(ii) To economize space.

3. What do you mean by stub?

Ans. It refers to the headings of horizontal rows.

4. What is the main part of the table called?

Ans. Body.

5. What is a general purpose table?

Ans. A general purpose table represents the raw data in great detail and covers a variety of information on the same subject.

6. What is the heading of rows called?

Ans. Stub

7. How are data presented in a table?

Ans. In tabulation, data are presented in the form of captions and stubs.

8. What does the body of a table contain?

Ans. It contains all the numerical information (data)

9. Write three essentials of a good table.

Ans. (i) Alternative

(ii) Management's size.

10. What is meant by caption?

Ans. It refers to the headings or designations of vertical columns.

11. Statement 1: Stubs are heading for the vertical columns in a table.

Statement 2: Captions are heading of horizontal rows in a table.

Ans. Both the statements are false.

12. Statement 1: Textual presentation of data enables one to emphasize certain points of the presentation

Statement 2: When the quantity of data is not too large textual presentation is more suitable.

Ans. Both the statements are true.

Short Answer Questions (Self Practice)

1. Distinguish between 'tabulation' and 'classification'
2. Discuss the essentials of a good table.
3. Briefly discuss (i) Title (ii) Body of the table
4. Discuss the merits of tabulation.

MCQs

1. **The most accurate mode of data presentation is:**
 - (a) Diagrammatic method
 - (b) Tabulation
 - (c) Textual presentation
 - (d) None of these
2. **Details are shown by:**
 - (a) Charts
 - (b) Tabular presentation
 - (c) Both (a) and (b)
 - (d) None of these
3. **Univariate frequency distribution is presented through:**
 - (a) Simple Table
 - (b) One-way Table
 - (c) Both (a) and (b)
 - (d) Two-way Table
4. **The headings of the rows given in the first column of a table are called:**
 - (a) Stubs
 - (b) Titles
 - (c) Captions
 - (d) Prefatory notes
5. **The column heading of a table is known as:**
 - (a) Stubs
 - (b) Sub-titles
 - (c) Reference notes
 - (d) Captions
6. **For tabulation, 'caption' is:**
 - (a) The lower part of the table
 - (b) The main part of the table
 - (c) The upper part of the table
 - (d) The upper part of a table that describes the column and sub-column
7. **Which of the following statements is not true for tabulation?**
 - (a) Complicated data can be presented
 - (b) Facilitates comparison between various observations
 - (c) Statistical analysis of data requires tabulation

- (d) Tabulation aims to present the classified data in a complex manner.
- 8. 'Stub' of a table is the:**
- (a) Right part of the table describing the columns
 - (b) Left part of the table describing the columns
 - (c) Right part of the table describing the rows
 - (d) Left part of the table describing the rows
- 9. Whether classification is done first or tabulation?**
- (a) Classification follows tabulation
 - (b) Classification precedes tabulation
 - (c) Both are done simultaneously
 - (d) No criterion
- 10. _____ is placed just below the title.**
- (a) Source
 - (b) Table Number
 - (c) Head Note
 - (d) Foot note
- 11. A table that shows two or more characteristics of data is known as:**
- (a) First Order Table
 - (b) Complex Table
 - (c) One-way Table
 - (d) General purpose table
- 12. In case of Tabulation, source of information is given:**
- (a) Just below the title
 - (b) Just below the Head Note
 - (c) Just above the Footnote
 - (d) Just below the Footnote
- 13. Simple Table is also known as:**
- (a) First Order Table
 - (b) One-way Table
 - (c) Both (a) and (b)
 - (d) Treble Table
- 14. In tabulation, source of the data, if any, is shown in the:**
- (a) Source Note
 - (b) Body
 - (c) Stub
 - (d) Caption
- 15. Special Purpose Table is also known as:**
- (a) Text Table
 - (b) Analytical Table
 - (c) Summary Table
 - (d) All of these
- 16. On the basis of Objective, tables are of two types:**
- (a) General Purpose Table and Special Purpose Table
 - (b) Original Table and Derived Table
 - (c) Simple Table and Complex Table

(d) None of these

17. Treble Table presents _____ characteristics.

- (a) Two (b) More than two
(c) Three (d) Four

18. General Purpose Table is also known as:

- (a) Reference Table (b) Derivative Table
(c) Repository Table (d) Both (a) and (c)

19. _____ Table contains data, which was initially collected from the primary source.

- (a) Summary (b) Derivative
(c) Reference (d) Original

20. The main objective of tabulation is:

- (a) To provide reference (b) To economize space
(c) To facilitate comparison (d) All of these

21. Number of students in different sections of Class XI will be shown by:

- (a) Treble Table (b) Double Table
(c) Simple Table (d) Manifold Table

MCQs ANSWERS

1	2	3	4	5	6	7	8	9	10
B	B	C	A	D	D	D	D	B	C
11	12	13	14	15	16	17	18	19	20
B	C	C	A	D	A	C	D	D	D
21									
C									

Fill in the blanks

- Three-way table is known as _____ table.
- In a Simple Table, data is presented according to _____ characteristic only.
- Headings and sub-headings given to columns of a table are called _____.
- _____ is a brief explanation of the contents of a table.
- In case of tabulation, _____ refer to the headings of horizontal rows.

Answers of fill in the blanks

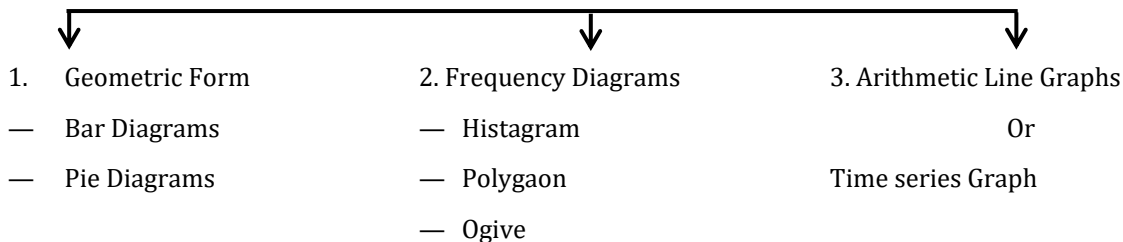
1. Treble
2. One
3. Captions
4. Title
5. Stubs

[Chapter – 5]

DIAGRAMMATIC PRESENTATION OF DATA: BAR DIAGRAMS AND PIE DIAGRAMS

- Data should be presented in a simple and attractive manner in the form of diagrams.
- Diagrammatic presentation is broadly classified as of 3 types.

Types of Diagrammatic Presentation



1. **Bar Diagrams** :—Bar diagrams are those diagrams in which data are presented in the form of bars or rectangles. It is also known as one dimensional diagram.

Main features of Bar diagram:

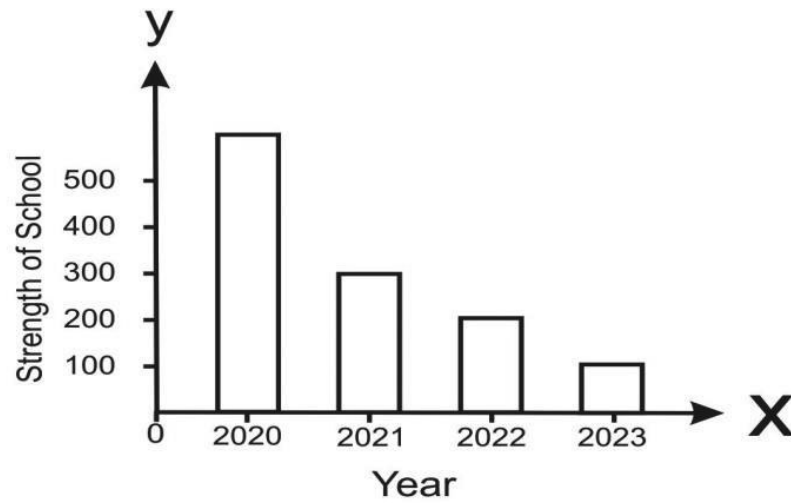
- 1) **Horizontal Axis** :—The horizontal axis (x-axis) represents the variable or categories being compares.
- 2) **Vertical Axis**:—The vertical axis (y-axis) represents the values or frequencies of the variables.
- 3) **Uniform Gap**:—There is a uniform gap between the bars, making the diagram easy to read.
- 4) **Clear Labels**:— Both axes are clearly labeled, and the diagram has a title that describes the data being represented.

Types of Bar Diagrams:—

1. **Simple Bar Diagram**:— These are those diagrams which are based on a single set of numerical data.

Example:—

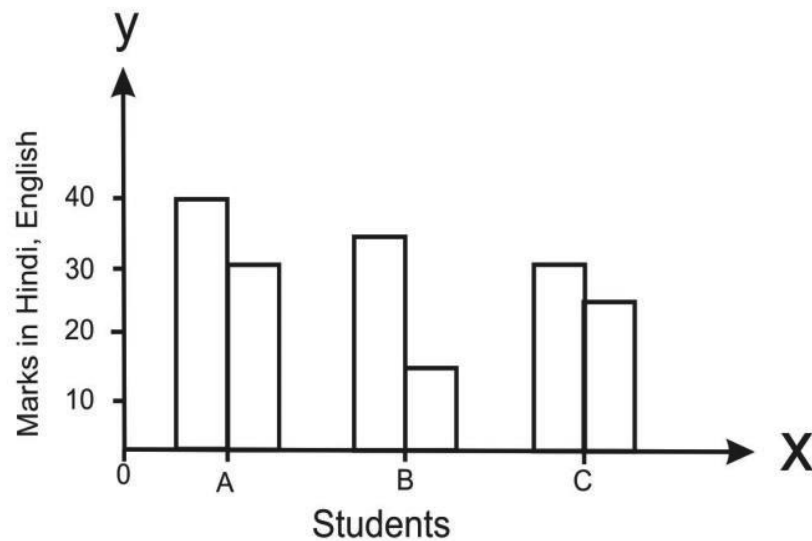
Year	2020	2021	2022	2023
Strength of school	500	300	200	100



2. Multiple Bar Diagram :—Compares the values of multiple variables across different categories.

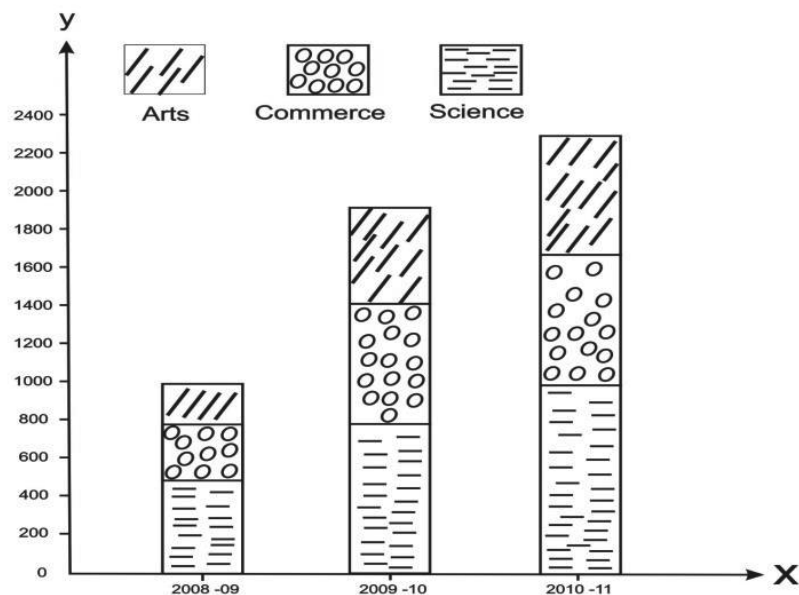
Example

Students	A	B	C
Marks in Hindi	40	35	30
In English	30	15	25



3. **Component Bar Diagram (Sub-divided or differential bar diagram):**—These diagrams which simultaneously present, total values as well as part values of a set of data.

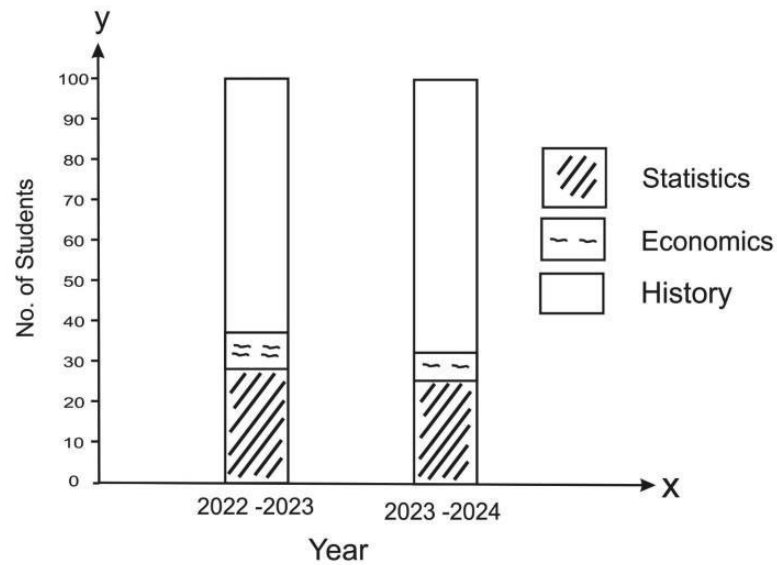
	Year		
	2008-09	2009-10	2010-11
Faculty			
Sciences	500	800	1000
Arts	300	600	750
Commerce	200	450	550



4. **Percentage Bar Diagrams:-** Percentage bar diagrams are these diagrams which show simultaneously, different parts of the values of a set of data in terms of percentages.

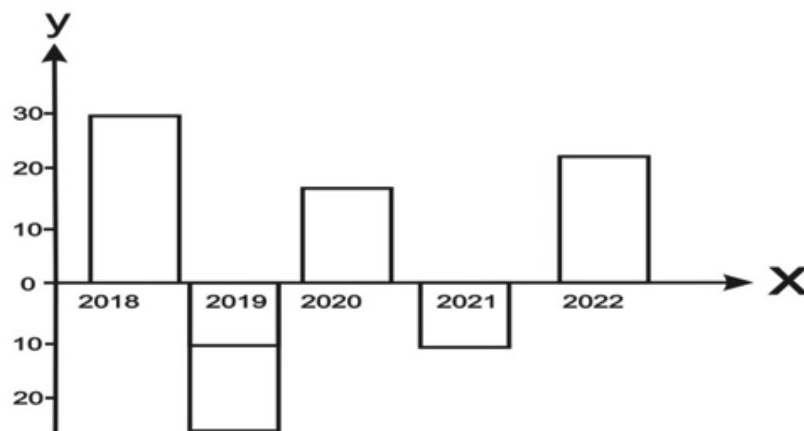
Example:—

Subjects	No. of students	
	2022-23	2023-24
Statistics	25	30
Economics	40	42
History	35	28



5) **Deviation Bar Diagram:-** Deviation Bar diagram are generally used for representing net quantities or net changes in data or excess – deficit .

Year	2018	2019	2020	2021	2022
Savings	30	-20	15	-10	20



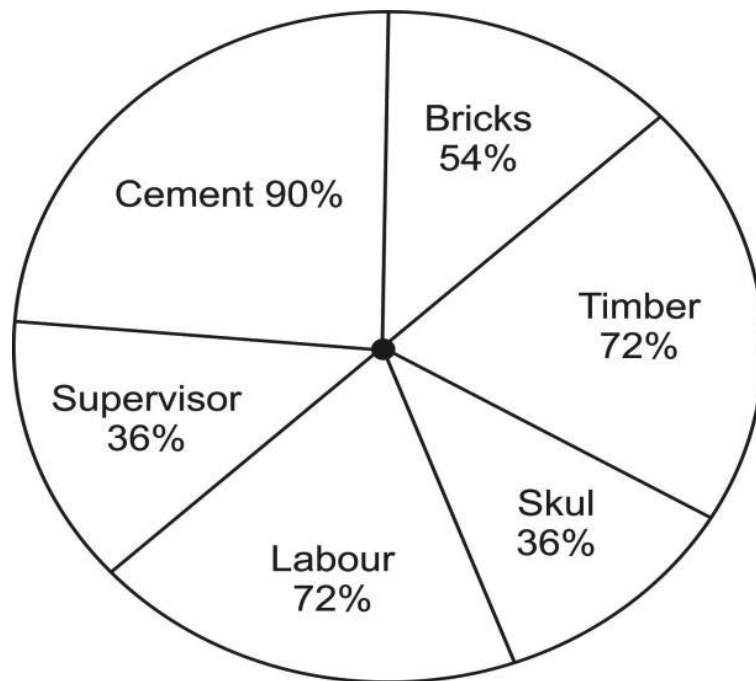
Pie Diagram

- A pie diagram, also known as a pie-chart, is a circular statistical graphic divided into slice to illustrate numerical preparation.
- In a pie chart the are length of each sector and is proportional to the quantity it represents.

Example

Items	% Expenditure	Percentage
Labour	20	72

Bricks	15	54
Cement	25	90
Stcul	10	36
Timber	20	72
Supervision	10	36



Histogram:—It is a graph of a grouped frequency distribution in which class-intervals are plotted on X-axis and their respective frequencies are Y -axis.

Histogram of frequency distribution are of two types.

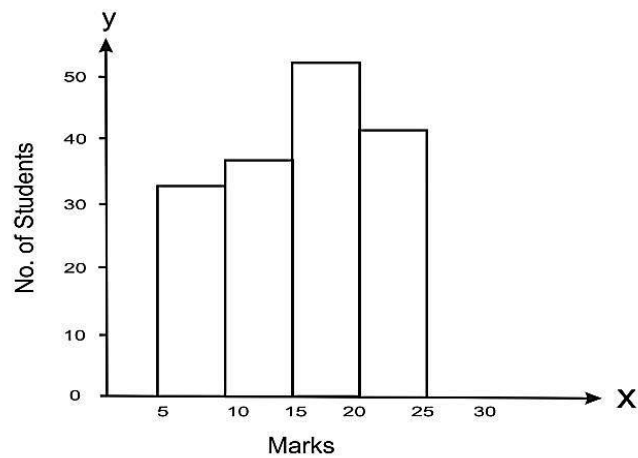
Histogram of equal class interval:These are those which are based on the data with equal class intervals.

In other words: In case of equal class intervals:—These are those which are based on the data with equal class intervals.

In other words: IN case of equal class intervals, the area of the rectangles are proportional to their respective frequencies.

All rectangular would have the same base in their width.

Marks	0-5	5-10	10-15	15-20	20-25
No. of students	25	30	35	50	40



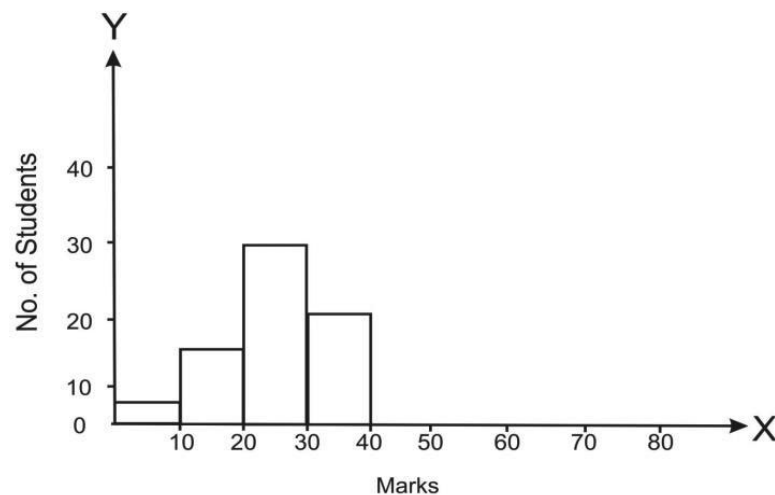
Histogram with unequal class intervals

When unequal class intervals of a frequency distribution are given, frequencies are first adjusted before constructing the histogram.

Example:—

Wages	0-10	10-20	20-30	30-50	50-80
No. of workers	5	15	30	40	30

Wages	Frequency	Adjusted frequencies
0-10	5	5
10-20	15	15
20-30	30	30
30-50	40	$40 \div 2 = 20$
50-80	30	$30 \div 3 = 10$



Histogram when Mid-values are given:—When mid values of class intervals are given first of all these mid-values must be converted into class intervals & then construct the histogram in the same manner.

Example

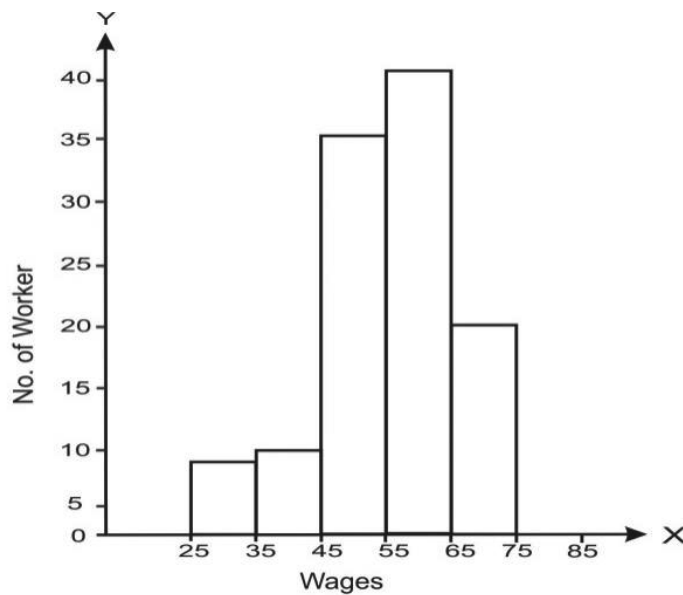
Mid values (Wages)	30	40	50	60	70
Frequency (Workers)	8	10	36	40	20

Solution :— The difference between two mid point = 10

Half of difference = $10/2 = 5$

5 will be add or sub from each mid value to get the class intervals.

Wages	25-35	35-45	45-55	55-65	65-75
No. of workers	8	10	36	40	20



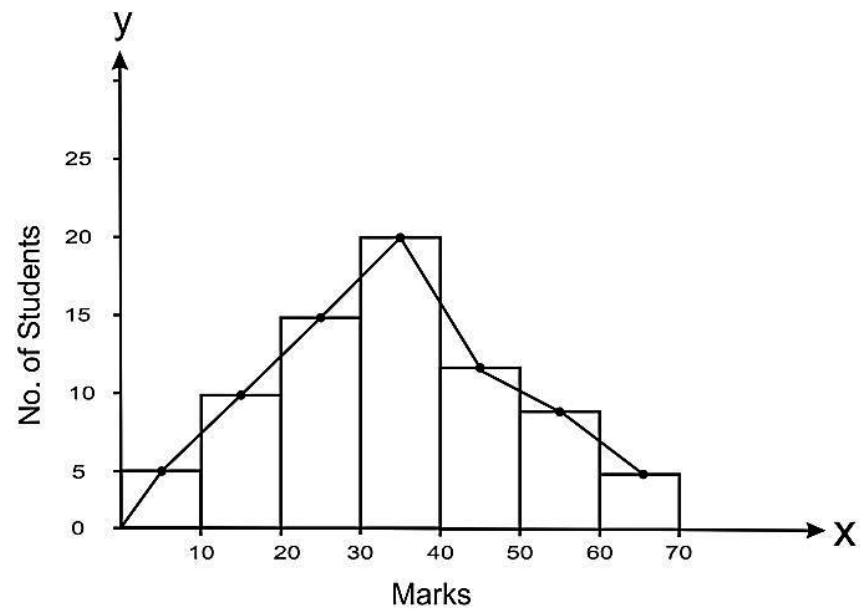
2. Frequency polygon:—It is the most common method of presenting grouped frequency distribution.

There are 2 methods of constructing frequency polygon:—

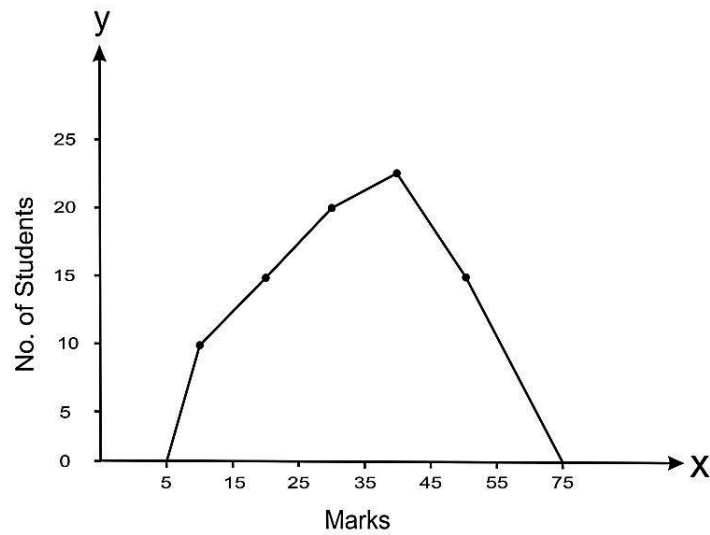
- (a) By Histogram
- (b) Without Histogram

Example:

Marks	No. of students
0-10	5
10-20	10
20-30	15
30-40	20
40-50	12
50-60	8
60-70	5

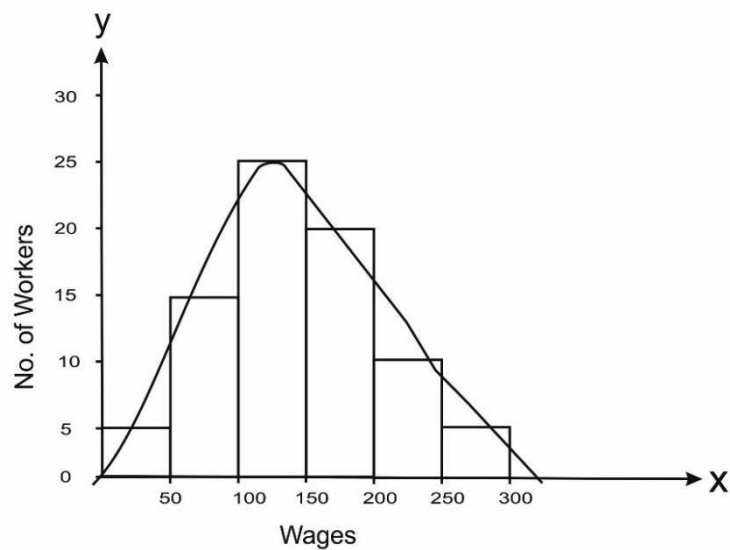
**Example:—**Frequency polygon without histogram

Marks	No. of Students	Mid-Value
10-20	10	15
20-30	15	25
30-40	20	35
40-50	22	45
50-60	15	55
60-70	10	65



4. **Frequency Curve:** It is the smoothed form of a frequency polygon when the verticals of a frequency polygon are joined by a smooth curve the resulting figure is called frequency curve.

In Wages	No. of workers
0-50	5
50-100	15
100-150	25
150-200	20
200-250	10
250-300	5



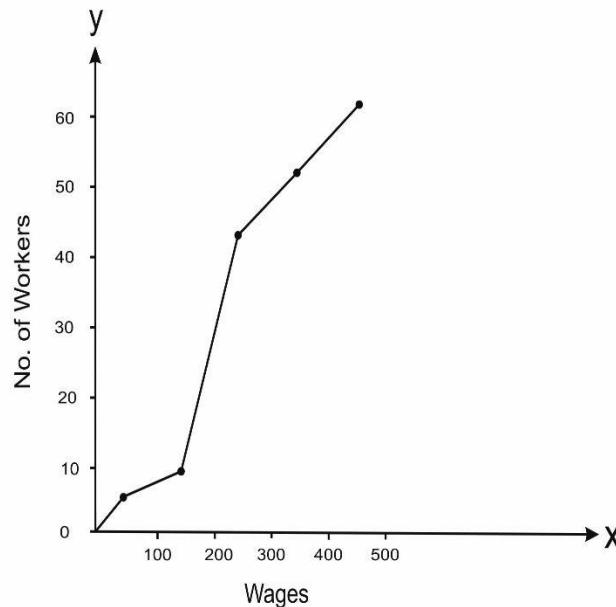
5. Ogive or Cumulative Frequency Curve.

It is the curve which is constructed by plating cumulative frequency data on graph paper in the form of a smooth curve.

Less than method: In less than ogive the cumulative frequency are platted against the respective upper limits of the class intervals.

Example:

Less than 100	6
Less than 200	11
Less than 300	44
Less than 400	58
Less than 500	64

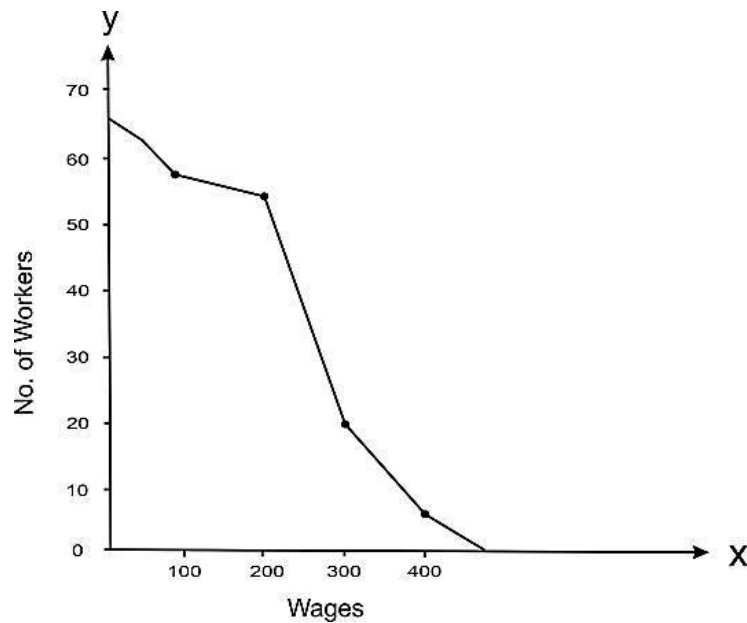


More than ogive & the cumulative frequencies are platted against the respective lower limits of the class interval.

Example

Wages	No. of worker
More than 0	64
More than 100	58
More than 200	53

More than 300	20
More than 400	6



More than ogive & the cumulative frequencies are plotted against the respective limits of the class interval.

En.	Wages	No of Workers	
More	Than	0	64
		100	58
		200	53
		300	20
		400	6

EXERCISE

Assertion: Reason Based

1. **Assertion (A)** : Bar diagrams are one- diversional

Reason (R) : In Case of Bar Diagram, only the length of the boar is kept in mind not the width

Ans: Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).

2. **Assertion (A)** : In case of one-dimensional diagrams, length or height of the bar remains the same.

Reason (R) : Bars need to be equidistant from each other in case of one – dimensional diagrams.

Ans: Assertion (A) is false but Reason (R) is true

3. **Assertion (A)** : While preparing sub-divided bar diagram, the various Components in each bar need to be kept in the same order.

Reason (R) : In a percentage bar diagram, all are of height 100.

Ans: Both Assertion (A) & Reason (R) are True & Reason (R) is not the correct explanation of Assertion (A)

4. **Assertion (A)** : Histogram is never drawn for a discrete variable.

Reason (R) : Histogram can also be to determine the value of mode.

Ans: Both Assertion (A) & Reason (R) are True & Reason (R) is not the correct explanation of Assertion (A).

5. **Assertion (A)** : Pie chart is a frequency diagram.

Reason (R) : Data in the form of grouped frequency distributions are generally represented by frequency diagram.

Ans: Assertion (A) is false but Reason (R) is true

6. **Assertion (A)** : A histogram is a two dimensional diagram.

Reason (R) : The width of a bar is not important in histogram.

Ans: Assertion (A) is true but Reason (R) is false.

7. **Assertion (A)** : Frequency polygon is the most common method of presenting grouped frequency distribution.

Reason (R) : Frequency polygon is an alternative to histogram.

8. **Assertion (A)** : Pie charts are not useful when the values of each component are similar.

Reason (R) : It is difficult to see the differences between Slice Sizes.

Ans: Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).

9. **Assertion (A) :** Pie chart are one of the most commonly used chart.

Reason (R) : Pie chart are simple to use.

Ans: Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).

10. **Assertion (A) :** The graphical representation of a cumulative frequency distribution is termed as a ogive curve.

Reason (R) : A Cumulative frequency distribution can be of 'less than' or more than type.

Ans: Both Assertion (A) and Reason (R) are true and Reason (R) is not a correct explanation of Assertion (A).

One Word

1. Name the graph that can determine the value of mode.

Ans: Histogram

2. Which graph can determine the value of median?

Ans: Cumulative frequency curve or ogive

3. What are frequency distribution graphs?

Ans: These are those graphs which are based on frequency distribution of the data

4. What is an Ogive?

Ans: An Ogive is a curve which is constructed by plotting cumulative frequency data on the graph paper.

5. Which type of Ogive rises from left to the right?

Ans: Less than Ogive

6. Define graphic presentation.

Ans: It is the method of presenting data in the form of curve or lines on graph paper.

Standard Based Questions

1. **Statement 1:** Histograms can also be used to determine the values of Median.

Statement 2: Histograms is a one –dimensional diagram s

Ans: Both the Statement are False

2. **Statement 1:** Simple Bar diagram depict only one characteristic of the data for various years.

Statement 2: Simple Bar diagram can be either vertical and horizontal.

Ans: Both the Statement are Ture

3. **Statement 1:** Bar diagram is an alternative to histogram

Statement 2 : Bar diagram can also derived from.

Ans: Both the Statement are False

4. **Statement 1:** Ogive is also called cumulative frequency curve

Statement 2: Ogive gives value of mode of the frequency distribution.

Ans: Statement 1 is true and Statement 2 is false

5. **Statement 1:** Time series graph is also known as Histogram

Statement 2: In more than Ogive the cumulative frequency are plotted against.

Ans: Both the statement are true.

Short Answer Questions

1. What is meant by sub-divided bar diagram?
2. Give the meaning of simple bar diagram.
3. Distinguish between Histogram and bar diagram
4. Mention three merits of graphic presentation.

5. State the steps for drawing frequency polygon in case of discrete series

Multiple Choice Questions

1. **The most attractive method of data presentations is:**
 - (a) Diagrammatic
 - (b) Textual
 - (c) Tabular
 - (d) Either (a) or (b)
2. **In a bar diagram, the bars are:**
 - (a) Horizontal
 - (b) Vertical
 - (c) Either (a) or (b)
 - (d) None of the above
3. **Diagrammatic representation of data is done by:**
 - (a) Pictures
 - (b) Charts
 - (c) Diagrams
 - (d) All these
4. **Sub-divided bar diagram is used to:**
 - (a) Study relation between different components
 - (b) Compare different components of a variable
 - (c) Either (a) or (b)
 - (d) Both (a) and (b)
5. **The most appropriate diagram to represent the data relating to the monthly expenditure on different items by a family is:**
 - (a) Histogram
 - (b) Pie diagram
 - (c) Angular Circle Diagram
 - (d) Both (b) and (c)
6. **When for some countries, the magnitudes are small and for other, the magnitudes are very large, to portray the data, it is preferred to construct:**
 - (a) Deviation bar diagram
 - (b) Duo-directional bar diagram
 - (c) Broken-Scale bar diagram
 - (d) Any of the above

7. In case of _____, all the bars are of equal height and width.
- (a) Multiple Bar Diagram (b) Percentage Bar Diagram
(c) Sub-divided Bar Diagram (d) Simple Bar Diagram
8. In a Pie Diagram, 1% value of data is represented by an angle at the center equal to:
- (a) 3.6° (b) 36°
(c) 360° (d) 0.36°
9. The diagram which represents data in a circle is known as:
- (a) Pie Diagram (b) Multiple Diagram
(c) Bar Diagram (d) Sub-divided bar Diagram
10. If a household spends 70% of his income on food, then degree measure of an angle in the pie diagram will be:
- (a) 200° (b) 210°
(c) 252° (d) 70°
11. Bar diagram is a:
- (a) Two Dimensional Diagram (b) One Dimensional Diagram
(c) Three Dimensional Diagram (d) None of these
12. The other name of Pie Diagram is:
- (a) Bar Diagram (b) Histogram
(c) Angular Circle Diagram (d) Polygon
13. Which of the following statements is true?
- (a) Breadth of the bar is same
(b) Bars may be drawn either vertically or horizontal
(c) Bars need to be equidistant from each other
(d) All of these

14. _____ is used to represent net changes in data like net profit, net exports, etc.
- (a) Pie Diagram (b) Deviation Bar Diagram
(c) Multiple Bar Diagram (d) Simple Bar Diagram
15. In case of one-dimensional diagrams, the magnitude of the characteristics s shown by:
- (a) Length of the bar (b) Width of the bar
(c) Either (a) or (b) (d) Both (a) and (b)
16. Year wise production of rice, wheat and maize for the last live years can be displayed by:
- (a) Sub-divided Bar Diagram (b) Simple Bar Diagram
(c) Broken Bar Diagram (d) Multiple Bar Diagram
17. One- dimensional diagrams are also known as:
- (a) Pie Diagram (b) Histograms
(c) Bar Diagrams (d) None of these
18. A good diagram should have:
- (a) Scale (b) Accuracy
(c) Index (d) All of these
19. Percentage bar diagram is a:
- (a) One-dimensional diagram (b) Two-dimensional diagram
(c) Three-dimensional diagram (d) None of these
20. Histogram is prepared in case of:
- (a) Individual Series (b) Discrete Series
(c) Continuous Series (d) None of these

21. Graphs can be useful for:

- (a) Showing trends in data
- (b) Adding visual appeal to business reports
- (c) Making comparisons
- (d) All of the above

22. A comparison among the class frequencies is possible only in:

- (a) Ogive
- (b) Histogram
- (c) Frequency polygon
- (d) Either (a) or (b)

23. The breadth of the rectangle is equal to the length of the class-interval in:

- (a) Ogive
- (b) Histogram
- (c) Both (a) and (b)
- (d) None of these

24. A frequency polygon is obtained by:

- (a) Constructing a frequency histogram
- (b) Constructing a cumulative frequency histogram
- (c) Linking mid-points from a frequency histogram
- (d) Using a line graph

25. While preparing time series graph, we show _____ on the X-axis:

- (a) Time
- (b) Income
- (c) Expense
- (d) All of the above

26. Ogive curve occur for:

- (a) More than type distribution
- (b) Less than type distribution
- (c) Bot (a) and (b)
- (d) None of (a) and (b)

27. It is always possible to construct a histogram from the:

- (a) Data
- (b) Frequency polygon
- (c) Both (a) and (b)
- (d) None of these

28. Mode is found graphically by:

- (a) Frequency Polygon
- (b) Ogive
- (c) Histogram
- (d) None of these

29. To find the number of observations less than any given value, we use:

- (a) Grouped frequency distribution
- (b) Single frequency distribution
- (c) Cumulative frequency distribution
- (d) None of these

30. A simple frequency distribution of a discrete variable is represented by:

- (a) Ogive
- (b) Histogram
- (c) Frequency Polygon
- (d) None of these

31. Median of a distribution can be obtained from:

- (a) Histogram
- (b) Frequency polygon
- (c) Less than type ogives
- (d) None of these

32. Diagrammatic representation of the cumulative frequency distribution is:

- (a) Frequency Polygon
- (b) Ogive
- (c) Histogram
- (d) None of these

33. Graph of successive points of a distribution joined by straight lines in statistical terminology is known as

- (a) Frequency Distribution
- (b) Frequency Polygon
- (c) Trend
- (d) Cumulative distribution curve

34. If you are interested in how earnings of a company have fluctuated over time, it would be best to use

- (a) Bar graph
- (b) Time series graph
- (c) Pie chart
- (d) Histogram

MCQs ANSWERS

1	2	3	4	5	6	7	8	9	10
A	C	D	D	D	C	B	A	A	C
11	12	13	14	15	16	17	18	19	20
B	C	D	B	A	D	C	D	A	C
21	22	23	24	25	26	27	28	29	30
D	B	B	C	A	C	C	C	C	C
31	32	33	34						
C	B	B	B						

Answer the following questions, 1 to 10, choosing the correct answer.

- Bar diagram is a**
 - One-dimensional diagram
 - Two-dimensional diagram
 - Diagram with no dimension
 - None of the above
- Data represented through a histogram can help in finding graphically the**
 - Mean
 - Mode
 - Median
 - All of the above
- Ogives can be helpful in locating graphically the**
 - Mode
 - Mean
 - Median
 - None of the above
- Data represented through arithmetic line graph help in understanding**
 - Long term trend
 - Cyclicality in data
 - Seasonality in data
 - All the above
- Width of bars in a bar diagram need not be equal. (True/False)**
- Width of rectangles in a histogram should essentially be equal. (True/False)**

7. Histogram can only be formed with continuous classification of data. **(True/False)**
8. Histogram and column diagram are the same method of presentation of data. **(True/False)**
9. Mode of a frequency distribution can be known graphically with the help of histogram **(True/False)**.
10. Median of a frequency distribution cannot be known from the ogives. **(True/False)**
11. What kind of diagrams are more effective in representing the following?
 - (i) Monthly rainfall in year
 - (ii) Composition of the population of Delhi by religion
 - (iii) Components of cost in a factory
12. Suppose you want to emphasize the increase in the share of urban non-workers and lower level of urbanization in India as shown in Example 4.2. How would you do it in the tabular form?
13. How does the procedure of drawing a histogram differ when class intervals are unequal in comparison to equal class intervals in a frequency table?
14. The Indian Sugar Mills Association reported that, 'Sugar production during the first fortnight of December 2001 was about 3,87,000 tonnes, as against 3,78,000 tonnes, during the same fortnight last year (2000). The off-take of sugar from factories during the first fortnight of December 2001 was 2,83,000 tonnes for internal consumption and 41,000 tonnes for exports as against 1,54,000 tonnes for internal consumption and nil for exports during the same fortnight last season.
 - (i) Present the data in tabular form.
 - (ii) Suppose you were to present these in diagrammatic form which of the diagrams would you use and why?
 - (iii) Present these data diagrammatically.
15. The following table shows the estimated sectoral real growth rates (percentage change over the previous year) in GDP at factor cost.

Year	Agriculture and allied sectors	Industry	Services
1994-95	5.0	9.2	7.0
1995-96	-0.9	11.8	10.3
1996-97	9.6	6.0	7.1
1997-98	-1.9	5.9	9.0
1998-99	7.2	4.0	8.3

1999-2000	0.8	6.9	8.2
-----------	-----	-----	-----

Line Frequency Graph

1. Draw a line frequency graph of the following data:

Marks	10	20	30	40	50	60	70
Frequency	3	7	9	11	12	14	15

Histogram

Equal Class-Intervals are given

2. Represent the following data by a histogram:

Marks in Statistics	0-20	20-40	40-60	60-80	80-100
No. of Students	12	28	60	48	30

Unequal Class-Intervals are given

3. The following data relates to the marks in Economics of 70 students. Depict it through Histogram.

Marks in Statistics	0-10	10-20	20-50	50-70	70-80
No. of Students	7	10	24	18	11

Mid-points are given

4. Represent the following frequency distribution by a histogram.

Mid-Values	2.5	7.5	12.5	17.5	22.5
Frequency	5	10	30	15	6

Inclusive Class-Intervals are given

5. The frequency distribution of marks obtained by 60 students of class in a college is given below:

Marks	30-34	35-39	40-44	45-49	50-54	55-59	60-64
No. of Students	3	5	12	18	14	6	2

Frequency Polygon

Discrete Series

6. Draw a frequency polygon for the following data:

Items	4	5	6	7	8	9	10
Frequency	4	6	10	25	22	18	12

7. Present the following data in the form of frequency polygon, using histogram.

Daily Wages (₹)	60-80	80-100	100-120	120-140	140-160	160-180	180-200
Frequency	4	6	10	25	22	18	12

8. The frequency distribution of marks obtained by students in a class test is given below:

Marks (mid-points)	45	55	65	75	85
No. of Students	5	9	12	8	2

Draw a histogram and frequency polygon to represents the frequency distribution of marks.

9. You are given the following marks secured by 45 students in an examination:

Marks	20-29	30-39	40-49	50-59	60-69	70-79
No. of Students	4	10	12	9	7	3

Represent the given data with the help of histogram and frequency polygon.

Continuous Series (without Histogram)

7. Present the following data in the form of frequency polygon, using histogram.

Mid-values	5	15	25	35	45	55	65	75
Frequency	4	10	16	22	25	12	7	2

Frequency Curve

11. Make a frequency curve of the following data:

Class Interval	20-40	40-60	60-80	80-100	100-120	120-140	140-160
Frequency	4	10	16	22	25	12	7

Ogive or Cumulative Frequency Curve

12. From the following information, construct, less than and more than ogive.

Daily Wages (in ₹)	60-90	90-120	120-150	150-180	180-210
No. of Workers	11	14	25	12	8

13. The table given below shows the amount of sales of 100 companies:

Sales (₹ in crores)	20-30	30-40	40-50	50-60	60-70	70-80
No. of Companies	7	12	15	30	22	14

With the help of information given above, prepare: (i) 'Less than' ogive; (ii) 'More than' ogive; and (iii) 'Less than' and 'More than' ogive.

Time Series Graphs

14. Represents the following data relating to annual profits of a company with the help of suitable graph:

Year	2017	2018	2019	2020	2021	2022	2023
Profit (in ₹ cores)	25	37	45	35	50	54	60

Simple Bar Diagram

1. From the following data relating to cement production of a factory (2018-23), prepare a simple bar diagram.

Year	2018	2019	2020	2021	2022	2023
Production (in M. Tonnes)	125	175	260	310	450	520

2. Average wages of some firms are given below. Represent this by Simple Bar Diagram.

Firma	A	B	C	D	E
-------	---	---	---	---	---

Average Wages (₹)	2,725	3,250	3,750	4,680	5,320
--------------------------	-------	-------	-------	-------	-------

Multiple Bar Diagram

3. Present the following data by multiple bar diagram.

Year	B. Com. (H)	Eco. (H)	B. Com. (IP)
2021-22	420	200	140
2022-23	320	240	300
2023-24	380	360	480

Sub-Divided Bar Diagram or Component Bar Diagram

4. Represent the following data with the help of sub-divided bar diagram:

Year	Production (in '000 tonnes)		
	Wheat	Rice	Cotton
2021	35	22	10
2022	40	25	16
2023	40	12	20

Percentage Bar Diagram

5. Represent the data relating to the cost of production in a factory by means of percentage diagram:

Element of Cost (in '000 tonnes)	Items		
	A	B	C
Raw Materials	75	55	60
Wages	45	30	25
Factory Overheads	15	10	30
Office Overheads	5	15	5

Deviation Bar Diagram

5. Represent the following data relating to net export of a company by deviation bar diagram:

Year	2017	2018	2019	2020	2021	2022	2023
Profit (in ₹ cores)	25	37	45	35	50	54	60

Deviation Bar Diagram

6. Represent the following data relating to net exports of a company by deviation bar diagram.

Year	2018	2019	2020	2021	2022	2023
Net Exports (in ₹ cores)	25	37	45	35	50	54

Pie Diagram

7. The estimated percentage break-up of the cost of construction of a house (excluding cost of land) s as follows:

Labour	Bricks	Cement	Steel & Timber	Marble	Miscellaneous
20%	12%	25%	15%	13%	15%

Construct a pie diagram to represent the given data.

8. Draw two pie diagrams to represent the marks obtained by Isha and Utkarsh (out of 50) in an examination.

Subject	Isha	Utkarsh
Maths	40	37
Economics	35	45
Accounts	45	40
Commerce	46	34
English	34	44

Miscellaneous Questions

9. Represent the following data with help of pie diagram.

Items	Rent	Food	Clothing	Transportation	Recreation	Savings	Others
Expenditure (in percent)	20%	25%	22%	15%	8%	6%	4%

10. The of tyres manufactured by a company for different years.

Year	No of Tyre Produced
2019-20	5,000
2020-21	7,500
2021-22	11,000
2022-23	15,600
2023-24	16,000

Present the data with the help of a suitable diagram.

12. The following table gives the monthly expenditure of two families A and B. Represent them by means of sub-divided bar diagram and Percentage diagram.

Items of Expenditure	Family A (income: ₹ 50,000)	Family B (income: ₹ 80,000)
Food	15,000	23,000
Clothing	9,000	17,000
House Rent	11,000	11,000
Education	5,000	9,000
Fuel and Lighting	4,000	13,000
Miscellaneous	6,000	7,000

13. Draw a pie diagram to represent the area of the oceans of world.

Ocean	Pacific	Atlantic	Indian	Antarctic	Arctic
Area (millions s.q. km)	70.8	41.2	28.5	7.6	4.8

14. The following data shows the profit earned by Company A and Company B for the year 2020-21 to 2023-24. Present the data with the help of suitable diagram.

Year	Company A (in ₹ '000)	Company B (in ₹ '000)
2020-21	225	160
2021-22	310	240
2022-23	450	380
2023-24	575	600

15. Represent the following data by Pie diagram.

Items of Expenditure	Family A	Family B	Family C
Food	12,000	30,000	40,000
Clothing	3,500	14,000	10,000
Education	2,000	7,000	3,000
House Rent	1,500	3,000	5,000
Miscellaneous	5,000	6,000	5,000

〔Chapter – 6〕

MEAN, MEDIAN, MODE

Arithmetic Mean: It is defined as the sum of observation divided by the number of observation.

- It is the most popular, simplest and basically used measure of central Tendency.
- The arithmetic mean is 2 type
 1. Simplest Arithmetic Mean
 2. Weighted Arithmetic Mean
 - (a) **Individual Series:** In individual series where frequencies are not given: We calculate arithmetic mean by following methods:
 1. Direct Method
 2. Short Cut Method
 3. Step deviation Method
- 1. **Direct Method:** In direct method, mean is obtained by dividing the sum of observation (Σx , ie. Summation of x) by the no of observation (N)

Formula: $\bar{X} = \frac{\Sigma X}{N}$

$\bar{X}$ = Arithmetic Mean

Σ = Summation

X = Value of Size

ΣX = Sum of all the observations

N = Number of items or observations.

Example: Calculate arithmetic mean of the following data.

Worker	Income
A	100

B	200
C	150
D	300
E	100
N = 5	$\Sigma X = 850$

$$\bar{X} = \frac{\Sigma X}{N}$$

$$\bar{X} = \frac{850}{5} = 170$$

2. **Short Cut Method (Assumed Mean Method):** In this method any value whether exist in the data or not, can be taken as assumed mean and then alleviation are taken from that value.

In order to simplify the calculation centrally located value in the data can be selected as assumed mean.

$$\text{Formula } \bar{X} = A + \frac{\Sigma d}{N}$$

Example:- Find mean by the short cut

Worker	Income	D=x-A (A=150)
A	100	-50
B	200	+50
C	150(A)	0
D	300	+150
E	100	-50
N = 5		$\Sigma d = 100$

- (iii) **Step Deviation Method:** The short cut method can be further simplified. The calculation can be further simplified by dividing all the calculations taken from assumed mean by the common factor (I or i) to reduce their size.

The divided deviations are called step deviations.

Example: Calculate mean by step deviation method.

Worker	Income	D=X-A (A = 150)	D=d/I(I = 50)
A	100	-50	-1

B	200	+50	+1
C	150(A)	0	0
D	300	+150	+3
E	100	-50	-1
N= 5			$\Sigma d' = 2$

Formula: $\bar{X} = A + \frac{\Sigma d'}{N} \times 1$

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

$$\bar{X} = 150 + \frac{100}{5}$$

$$\bar{X} = 150 + 20 = 170$$

Arithmetic Mean ($\bar{X}$) = 170

(b) Discrete Series: In case of discrete series frequencies are grouped according to their size.

1. **Direct Method:**
2. **Short cut Method**
3. **Step deviation Method**

1. **Direct Method:** Frequency against each of the observation is multiplied by value of the observation.

Obtained values added together and divided by the no. of items i.e. Σf

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

Σfx = Sum of the product of variables frequencies.

$\Sigma f = N$ = Sum of frequencies.

Ex: Calculate arithmetic mean by the following data.

Marks	No. of Students	Fx
2	6	12
4	8	60
8	7	56
10	9	90
	2f= N = 40	$\Sigma fx = 250$

$$\bar{X} = A + \frac{\Sigma d'}{N} \times 1$$

$$\bar{X} = \frac{2fx}{N}$$

$$\bar{X} = \frac{250}{40} = \bar{X} = 6.25 \text{ Marks}$$

ii. Short Cut Method

Marks	No. of Students	d= X-A(A = 6)	Fd
2	6	-4	-24
4	8	-2	-16
6(A)	10	0	0
8	7	+2	14
10	9	+4	36
	$\Sigma f = N = 40$		$\Sigma fd = +10$

$$\bar{X} = A + \frac{\Sigma fd}{N}$$

$$\bar{X} = 6 + \frac{10}{40}$$

$$\bar{X} = 6 + 0.25$$

$$\bar{X} = 6.25 \text{ Marks}$$

(iii) Step Deviation Method

$$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times I$$

			(I = 2)	
Marks	No. of Students	D=X -A (A=6)	D' = d/I	Fd'
2	6	-4	-2	-12
4	8	-2	-1	-8
6(A)	10	0	0	0
8	7	+2	+1	+7
10	9	+4	+2	+18
	$\Sigma f = N = 40$			$\Sigma fd' = 5$

$$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times I$$

$$\bar{X} = 6 + \frac{5}{40} \times 2$$

$$\bar{X} = 6 + 0.25$$

$$\bar{X} = 6.25 \text{ Marks}$$

Continuous Series: In case of continuous series or frequency distribution, class intervals are given.

Mid Values (Mid - points) of various class intervals are taken.

$$\text{Mid Value} = \frac{L_1 + L_2}{2}$$

The arithmetic mean in continuous series:

i. Direct Method

ii. Short cut method

iii. Step deviation method.

i. Direct Method In this method we convert the series in a discrete form by finding the mid-values and rest of the procedure is the same as under discrete series.

$$\text{Formula :-} \frac{\sum fx}{\sum f}$$

Example: Calculate arithmetic mean by direct method

Marks	No. of students (f)	M.V. (X)	f×x ₂
0-10	4	5	20
10-20	7	15	105
20-30	10	25	250
30-40	5	35	175
40-50	4	45	180
	$\Sigma f = N = 30$		$\Sigma fx = 730$

$$\text{M.V.} = \frac{L_1 + L_2}{2}$$

$$\bar{X} = \frac{\Sigma fx}{\Sigma f} = \frac{730}{30} = 24.33$$

$$\bar{X} = 24.33$$

(ii) Short cut method

$$\text{Formula : - } \bar{X} = f + \frac{\Sigma fd}{\Sigma f}$$

Example: Calculate Arithmetic mean

Marks	No. of students	Mid value	d=	fd
0-10	4	5	-20	-80
10-20	7	15	-10	-70
20-30	10	25(A)	0	0
30-40	5	35	+10	50
40-50	4	45	+20	80
	$\Sigma f = N = 30$			$\Sigma fd = -20$

$$\bar{X} = A + \frac{\Sigma fd}{\Sigma f}$$

$$\text{M.V.} = \frac{L_1 + L_2}{2}$$

$$\bar{X} = 25 + \frac{(-20)}{30}$$

$$\bar{X} = 25 - 0.666$$

$$\bar{X} = 25 - 0.67$$

$$\bar{X} = 24.33$$

(iii) Step Deviation Method

$$\text{Formula : } \bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times I$$

Marks	MV (x)	No. of students (f)	D = X-A A = 25	d' = d/I	fd
0-10	5	4	-20	-2	-8
10-20	15	7	-10	-1	-7
20-30	25(A)	10	0	0	0
30-40	35	5	+10	+1	+5
40-50	45	4	+20	+2	+8
		$\Sigma f = N = 30$			$\Sigma fd' = -2$

$$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times I$$

$$\text{M.V.} = \frac{L_1 + L_2}{2}$$

$$\bar{X} = 25 + \left(\frac{-2}{30}\right) \times 10$$

$$\bar{X} = 25 - \frac{2}{3}$$

$$\bar{X} = 25 - 0.666$$

$$\bar{X} = 25 - 0.67$$

$$\bar{X} = 24.33$$

Arithmetic mean in different frequency distribution

1) Mean in Inclusive series.

For Calculating mean, inclusive series are not required to be converted into exclusive.

The M.V. remain same

Example: Calculate mean

Size	F	MV (x)	D=x-A (A=24.5)	d' = d/I (I = 10)	Fd'
0-9	2	4.5	-20	-2	-4
10-19	3	14.5	-10	-1	-3
20-29	1	24.5(A)	0	0	0
30-39	3	34.5	+10	+1	+3
40-49	1	44.5	+20	+2	+2
	$\Sigma f = N = 10$				$\Sigma fd' = -2$

$$\text{Formula: } \bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times I$$

$$\bar{X} = 24.5 + \frac{(-2)}{10} \times 10$$

$$\bar{X} = 24.5 - 2$$

$$\bar{X} = 22.5$$

Mean in cumulative frequency distribution

➤ Cf distribution ('less than or more than') must be converted into an ordinary or simple frequency distribution.

Example:

Wages	No. of Worker
More than 0	50

More than 100	45
More than 200	38
More than 300	20
More than 400	8

Wages	No. of Worker (f)	M.V.	d = x-A	d = d/I (I = 100)
0-100	5	50	-200	-2
100-200	7	150	-100	-1
200-300	18	250(A)	0	0
300-400	12	350	+100	1
400-500	8	450	+200	2
	$\Sigma f = N = 50$			

$$\Sigma fd' = 11$$

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times I$$

$$\bar{X} = 250 + \frac{11}{50} \times 100$$

$$\bar{X} = 250 + 22 = 272$$

$$\bar{X} = 272$$

Example → Calculate mean

Income	No. of persons
Less than 100	5
Less than 200	12
Less than 300	30
Less than 400	42
Less than 500	50

Answer:

Income	No. of person	M.V. (x)	D=x-A A = 250 -200	d' = d/I I = 100
0-100	5	50	-200	-2

100–200	7	150	–100	–1
200–300	18	250(A)	0	0
300–400	12	350	100	1
400–500	8	450	200	2

$$\Sigma fd' = 1$$

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times I$$

$$= 250 + \frac{11}{50} \times 100$$

$$\bar{X} = 250 + 22 = 272$$

3) When mid values are given

MV	F	fx
2	3	6
4	5	20
6	7	42
8	3	24
10	3	20
	$\Sigma f = N = 20$	$\Sigma fx = 112$

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

$$\bar{X} = \frac{112}{20} = 5.6$$

$$\bar{X} = 5.6$$

4) Unequal class Interval

Marks	No. of students	M.V.	Fx
0–10	7	5	35
10–30	6	20	120
30–40	10	35	350
40–80	5	60	300
80–100	2	90	180
	$\Sigma f = N = 30$		$\Sigma fx = 985$

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

$$\bar{X} = \frac{985}{30} = 32.83$$

$$\bar{X} = 32.83$$

5. Arithmetic mean in open - end series

Size	f	Mv(x)	d = X-A	d'=d/I	fd'
		A = 125	I = 50		
0-50	5	25	-100	-2	-10
50-100	7	75	-50	-1	-7
100-150	18	125(A)	0	0	0
150-200	6	175	50	1	+6
200-250	4	225	100	2	+8
	$\Sigma f = W = 40$				$\Sigma fd' = -3$

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times I.$$

$$= 125 + \frac{(3)}{40} \times 50$$

$$= 125 - 3.75 = 121.25$$

$$\bar{X} = 121.25$$

Median :- Median is a positional average because its value depends upon the position and not on magnitude of the items.

Calculation of median in individual

Arrange the data in ascending or descending order.

Apply the formula

$$M = \left[\frac{N+1}{2} \right]^{th} \text{ item}$$

M is median

N = Number of item.

Find out median → 100, 75, 96, 150, 85, 109, 35

Arranging → 35, 75, 85, 96, 100, 109, 150.

$$N = 7$$

$$= \frac{N+1}{2}; \frac{7+1}{2}; \frac{8}{2}$$

M = Size of the 4th item = 96]

$$M = 96$$

2) Calculate median in discrete series—

Wages (x)	No. of workers(f)	cf
100	3	3
200	8	11
300	10	21
400	4	25
500	2	27
	$\Sigma f = N = 27$	

14th item is located in the cf 21 size of which is 300.

$$M = \frac{N+1}{2}$$

$$M = 27 + \frac{1}{2}$$

M = Size of the 14th item

$$M = `300$$

Calculate Median in continuous series

- We have to located the median lays where median lies.
- The value of the median can be calculated from the median class with the help of formula.
- $M = L_1 + \frac{l}{f}(m - cf)$ in ascending order
- $M = L_2 - \frac{l}{f}(m - cf)$ in descending order

Example: Calculate median

Marks	No. of students	Cf
0-20	7	7
20-40	5	12(cf)
(L ₁)40-60	20(f)	32 (median)
60 -80	10	42

80–100	8	50
	$\Sigma f = N = 50$	

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

m = size of $\left(\frac{50}{2}\right)^{\text{th}} = 25^{\text{th}}$ item

25th item lives in cf 32 median group is 40 – 60

M = $L_1 + \frac{I}{f}(m - cf)$ in descending order

Example: **Calculate median**

Marks	No. of Students	cf
0–20	7	7
20–40	5	12(cf)
(L ₁) 40–60	20(f)	32 (median)
60–80	10	42
80–100	8	50
	$\Sigma f = N = 50$	

M = Size of $\left(\frac{N}{2}\right)^{\text{th}}$ item

M = Size of $\left(\frac{50}{2}\right)^{\text{th}} = 25^{\text{th}}$ item

25th item lives in cf 32 median group is 40–60

M = $L_1 + \frac{I}{f}(m - cf)$

$L_1 = 40, I = 20, f = 20, m = 25, cf = 12$

M = $40 + \frac{20}{20}(25 - 12)$

M = $40 + 13$

M = 53 Marks

Calculation of Median in different frequency distribution.

- Calculate in inclusive series
- Before the computation of median, inclusive series should be converted into exclusive series
- The frequencies of all classes will remain same.

Example:

Value	1-9	10-19	20-29	30-39	40-49
Frequency	5	7	11	5	2
Value (x)	F			Cf	
0.5-9.5	35			5	
9.5-19.5	7			12(cf)	
(L ₁) 19.5-29.5	11(f)			23 Median class	
29.5-39.5	5			28	
39.5-49.5	2			30	
	$\Sigma f = N = 30$				

m = size of $\left[\frac{N}{2}\right]^{th}$ item

m = Size of $\left[\frac{30^{th}}{2}\right]$ item = 15

15 item lies in cf23 so median group is (19.5-29.5)

$$M = L_1 + \frac{i}{f}(m - cf)$$

$L_1 = 19.5$, $I = 10$, $F = 11$, $m = 15$, $cf = 12$

$$M = 19.5 + \frac{10}{11}(15 - 12)$$

$$M = 19.5 + \frac{10}{11} \times 3$$

$$19.5 + 2.727 = 22.227$$

$$M = 22.23$$

2) Median in Cumulative Frequency

Cumulative frequency distribution class than or more than must be converted into an ordinary or simple frequency distributed

Income	No. of Persons
Less than 100	5
Less than 200	12
Less than 300	30
Less than 400	42

Less than 500	50
---------------	----

Income (x)	No of Persons (f)	Cf
0-100	5	5
100-200	7	12(cf)
L ₁ 200 -300	18(f)	30 median group
300 -400	12	42
400-500	8	50
	$\Sigma f = N = 50$	

M = Size of $\left[\frac{N}{2}\right]^{th}$ item

M = size of $\left[\frac{50^{th}}{2}\right]$ item = 25

25 item lies in cf 30, so median group is (200-300)

$M = L_1 + \frac{I}{f} (m - cf)$

$M = 200 + \frac{100}{18} (25 - 12)$

$200 + \frac{1300}{18}$

$M = 200 + 72.22 = 272.22$

(3) When Mid values are given

For calculating median, mid values must be converted into groups showing lower and upper limits.

Example:

Mu	10	20	30	40	50	60
F	7	10	12	21	12	8

Mid values need to convert into continuous series.

$$L_1 = M.V. - \frac{1}{2} \times I$$

$$10 - \frac{1}{2} \times 10 = 5$$

$$L_1 = 5$$

$$L_2 = M.V. + \frac{1}{2} \times I \quad 10 + \frac{1}{2} \times 10 = 15$$

$$L_2 = 15$$

Lower Limit $L_1 = 5$	Of first class
Upper Limit $L_2 = 15$	

X	F	Cf
5-15	7	7
15-25	10	17
25-35	12	29 (cf)
$L_1(35-45)$	21(f)	50 (Median group)
45-55	12	62
55-65	8	70
	$\Sigma f = N = 70$	

$M = \text{Size of } \left[\frac{N}{2} \right]^{th} \text{ item}$

$M = \text{size of } \left[\frac{70^{th}}{2} \right] \text{ item} = 35$

35 item lies in cf 50, so median group is 35-45

$$M = L_1 + \frac{I}{F}(m - cf)$$

35 item lies in cf 50, so median group is 35-45

$$M = L_1 + \frac{I}{f}(m - cf)$$

$$M = 35 + \frac{10}{21}(35 - 29) = 35 + \frac{10}{21} \times 6 = 35 + \frac{60}{21}$$

$$M = 35 + 2.857$$

$$M = 37.86$$

Quartile

- It is a statistical series is divided into four equal parts. The end value of each part is called quartile.
 - It is written as Q.
 - First quartiles also known as Q1 or lower quartile.
- $Q_2 =$ Is some as median of the series.

Q_3 = Upper quartile.

Calculation of quartile in individual & discrete of series.

Q_1 = Size of $\left(\frac{N+1}{4}\right)^{th}$ item of the series.

Q_3 = Size of $3\left(\frac{N+1}{4}\right)^{th}$ item of the series.

Formula:- $Q_1 = L_1 + \frac{\left(\frac{N+1}{4} - cf\right)}{f} \times i$

$Q_3 = L_3 + \frac{\left[3\left(\frac{N+1}{4}\right) - cf\right]}{f} \times i$

Q. Calculate Q_1, Q_3

The data is first arranged in ascending order:

S.No.	21	15	40	30	26	45	50	54	60	65	70
-------	----	----	----	----	----	----	----	----	----	----	----

$N = 11$

Q_1 = Size of $\left(\frac{N+1}{4}\right)^{th}$ item

= Size of $\left(\frac{11+1}{4}\right)^{th}$ item

= Size of 3rd item = 26

Q_3 = Size of $3\left(\frac{N+1}{4}\right)^{th}$ item

Size of $3\left(\frac{11+1}{4}\right)^{th}$ item

Size of 9th item = 60

Mode

According to the Kenny: — The value of the which occurs most frequency in a distribution is called the mode.

Croxtan and cowden state

The mode may be regarded as the most typical of a series of value.

Two ways of calculating mode

- 1) By Inspection
- 2) By converting individual series into discrete frequency series

Example

Age of 15 students of a class is reported below. Find the mode age

Age of series	22	24	17	18	17	19	18	21	20
	21	20	23	22	22	22			

Arrange the series in an ascending order :—

17	17	18	18	19	20	20	21	21	22
22	22	22	23	24					

Mode (Z) = 22

(2) Grouping Method:-

Size	2	3	4	5	6	7	8	9	10	11	12	13
Frequency	3	8	10	12	16	14	10	8	17	5	4	1

Size	I	II	III	IV	V	VI
Frequency		1+2	2+3	1+2+3	2+3+4	3+4+5

2	3	3+8=11	18	21	30	38
3	8					

4	10	10+12=22	28	42	40	32
5	12					

6	16	16+14=30	24	35	30	26
7	14					

8	10	10+8=18	25	10
9	8			

8	10	10+8=18	25	10
---	----	---------	----	----

9	8			
----------	----------	--	--	--

10	17	17+5=22	9
11	5		

12	4	4+15
13	1	

Inspection method in continuous series

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

C.I.	0-10	10-20	20-30	30-40	40-50
Frequency	2	5	7	5	2

C.I.	Frequency
0-10	2
10-20	5 (F ₀)
(L ₁)20-30	7(f ₁)
30-40	5
40-50	2

C.I.	Frequency
0-10	2
10-20	5
20-30	7
30-40	5
40-50	2

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

$$Z = L_2 20 + \frac{7-5}{2(7)-5-5} \times 10$$

$$= 20 + \frac{2}{14 - 10} \times 10$$

$$20 + \frac{2}{4} \times 10 = 20 + 5 = 25$$

$$Z = 25$$

(2) Grouping Method

$$Z = L_1 \frac{f_1 - f_0}{2f_1f_0 - f_2} \times i$$

C.I.	0-5	5-10	10-15	15-20	20-25	26-30	30-35
Frequency	1	2	10	4	10	9	2

Class interval	I	II	III	IV	V	VI
0-5	1	1+2=3	12	13		
5-10	2					
10-15	10	10+4=14			16	
15-20	4					
20-25	10	10+9=19	14	23		24
25-30	9					
30-35	2			11	21	

EXERCISE

Measure of Central Tendency

Assertion : Reason Based –

1. **Assertion** :Arithmetic mean is rigidly defined.

Reason: Everyone who computer the average get the same answer

Ans. Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A)

2. **Assertion** :Arithmetic mean is a calculated value.

Reason :Arithmetic mean is based on position in the series.

Ans. Assertion (A) is true but Reason (R) is false.

3. **Assertion** : The median refers to the middle is a distribution

Reason : Median is a positional overage

Ans. Both Assertion (A) and Reason (R) are true and Reason (R) is not the correct explanation of Assertion(A)

4. **Assertion** :Mode of a frequency distribution can be located graphically with the help of histogram.

Reason: Mode is preferable to mean and median when it is desired to know the matt typical value

Ans. Both Assertion (A) and Reason (R) are true and Reason (R) is not the correct explanation of Assertion. (A)

5. **Assertion:** Mode is the most typical value of a distribution.

Reason: The mode is affected by extreme values.

Ans. Assertion (A) is true but Reason (R) is false.

6. **Assertion:** The value of mode cannot always be determined.

Reason: Mode is capable of algebraic treatment.

Ans. Assertion (A) is true but Reason (R) is false.

7. **Assertion:** The second quartile is equal to median.

Reason: The second quartile has 50% of items below it and 50% of the observation above it.

Ans. Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A)

One Word

1. **What do you mean by measure of Central tendency?**

Ans. A measure of Central tendency is a single value, which represents an entire set of data.

2. **Give the meaning of arithmetic mean.**

Ans. It is defined as the sum of the values of all observations divided by the number of observations.

3. **If five students obtained 10,20,30,40 and 50 marks, what would be the arithmetic mean.**

Ans. 30 Marks

4. **Mention two drawbacks of arithmetic mean.**

Ans. (i) Affected by extreme values.

(ii) Not possible in case of qualitative characteristics.

5. Give the meaning of combined mean.

Ans. When mean of two or more than two series is computed collectively, it is known as combined mean.

6. Define Median.

Ans. Median may be defined as the middle value in the data set when its elements are arranged in a sequential order.

7. Which graph is used to locate median Graphically ?

Ans. Ogive curve.

8. Define Mode.

Ans. It is the value of occurring most frequently in a set of observations and around which other items of the set cluster most densely.

9. Give two merits of mode.

Ans. (i) Easy and simple to calculate
(ii) Not affected by the values of extreme items.

10. Which graph is used to compute mode?

Ans. Histogram.

Statement based Questions

1. **Statement 1:** Median is equal to middle item of distribution in case of even number of items.

Statement 2: To calculate median in an individual service the data needs to be first arranged in ascending or descending order.

Ans. Statement 2 is true and statement 1 is false.

2. Statement 1: Median can be easily located graphically with help of Ogives?

Statement 2: Median is not influenced by the extreme values.

Ans. Both the statements are true.

3. Statement 1: Arithmetic average is not affected by extreme value.

Statement 2: Arithmetic mean can be easily calculated in case of open-end classes without making any assumptions.

Ans. Both the statements are false.

4. Statement 1 Average needs to be calculated by considering each and every item of the series.

Statement 2 Average should be capable of further algebraic treatment.

Ans. Both the statements are true.

Short answer questions.

1. State three and two demerits of arithmetic mean.
2. Distinguish b/w mean and Median.
3. Mention the merits of mode.
4. Discuss the steps involved in calculating mode by Grouping method.

Fill in the blanks

1. Open-end class-intervals are those which do not have the _____ limit of the first class-interval and the _____ limit of the last class-interval.
2. Short-cut method is also called _____.
3. The sum of deviations of the observations from their arithmetic mean is always _____.
4. Weighted mean is _____ to simple arithmetic mean when equal weights are used for all the items in the series or distribution.
5. The arithmetic mean of 1, 4, 5, 7 and 8 will be _____.
6. _____ is preferable to mean and median when it is desired to know the most typical value.
7. According to Karl Pearson, the relationship between mean, median and mode in an asymmetrical distribution is given by $\text{Mode} = \text{_____}$.
8. The sum of deviation of items from median, ignoring signs, is the _____.
9. _____ can be easily located graphically with help of Ogives.
10. The value of the variable which occurs most frequently in a distribution is known as _____.
11. _____ is known to be the best average in open-end class-interval series.
12. Median of series 10, 14, 19, 25, 36 is _____.

Answers of Fill in the Blanks

1. Lower, Upper
2. Assumed Mean Method
3. Zero
4. Equal
5. 5
6. Mode
7. $3 \text{ Median} - 2 \text{ Mean}$
8. Minimum
9. Median
10. Mode
12. Median
12. 19

Unsolved Numerical

By using appropriate method solve the following numerical.

Individual Series

1. Calculate Arithmetic Mean from the following data:

7, 13, 15, 20, 40, 30

2. Find out Arithmetic Mean from the following data:

30, 40, 50, 20, 12, 60, 8

3. Find out Arithmetic Mean from the following data:

5, 10, 8, 12, 10

4. Find out Arithmetic Mean from the following data:

25, 15, 12, 18, 15

5. Find out Arithmetic mean from the following data:

7, 6, 10, 4, 3

Discrete Series

6. Find out mean wages from the following:

Wages (₹)	160	170	180	190	200
Number of Workers	4	10	6	3	7

7. Calculate the average goals from the following:

Matches Played	4	5	6	7	8
Number of Goals	1	0	4	2	3

8. Compute the mean from the following:

X	2	4	6	8	10
F	2	1	5	1	1

9. Find out the mean from the following:

X	5	10	15	20	25	30
F	4	3	2	8	2	1

Continuous Series**10. Find out the mean from the following:**

X	0-2	2-4	4-6	6-8	8-10
F	5	2	3	8	2

11. Find out the mean from the following

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

12. Find out arithmetic mean from following data:

Class	Frequency
0-20	5
20-40	4
40-60	3
60-80	2
80-100	6

13. Find out average marks for class XI from following data:

Marks (Less than)	Number of Students
10	3
20	12
30	30
40	40
50	60
60	90

Missing Frequency**14. Find out the missing value if mean is 23:**

Class	0-10	10-20	20-30	30-40	40-50
Frequency	8	5	?	4	6

15. Find out the missing value if mean is 40 and N = 100.

X	5	15	25	35	45	55	65	75
F	5	7	?	25	?	15	8	5

16. Determine the Arithmetic Mean from the following data:

S.No.	1	2	3	4	5	6	7	8	9	10
Marks	30	28	32	12	18	20	25	15	26	14

17. Calculate the Arithmetic Mean:

Size	20	19	18	17	16	15	14	13	12	11
Frequency	1	2	4	8	11	10	7	4	2	1

18. Calculate the Arithmetic Mean from the following data:

Marks (less than)	80	70	60	50	40	30	20	10
No. of Students	50	45	40	30	16	10	7	3

Median and Mode

Individual Series

18. Find out Median and Mode from the following data:

5,10,8,7,10,15,10

19. Find out value of Median and Mode:

S.No.	1	2	3	4	5	6	7	8
Income	50	150	100	200	300	200	250	125

20. Calculate the value of Median and Mode:

Students	A	B	C	D	E	F	G	H	I	J
Marks	30	25	50	40	35	60	35	50	70	50

21. Calculate the value of Median and Mode:

5,10,8,12,10.

Discrete Series

22. Calculate the value of Median and Mode from the following data:

Size	5	10	15	20	25	30	35	40
Frequency	10	18	2	5	7	10	5	3

23. Find out the value of Median and Mode from the following:

Age	10	20	30	40	50	60	70
Number of Persons	40	20	25	100	20	40	25

24. Find out the value of Median and Mode from the following:

X	2	4	6	8	10
F	2	1	5	1	1

25. Find out the value of Median and Mode from the following:

X	5	10	15	20	25	30
F	4	3	2	8	2	1

26. Find out the value of Median and Mode from the following:

X	1	2	3	4	5
F	2	1	5	1	1

27. Calculate Mode from the following data:

Marks	10–20	20–30	30–40	40–50	50–60	60–70
Number of Students	7	10	30	5	12	8

28. Calculate Mode from the following frequency distribution:

Wages	100–200	200–300	300–400	400–500	500–600	600–700	700–800
No. of Workers	20	10	15	20	10	12	5

29. Calculate Median and Mode from the following data:

Size	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40
Frequency	10	3	7	5	10	6	3	6

Quartiles

Individual Series

30. Determine the value of Lower Quartile and Upper Quartile:

10	6	12	20	15	18	25	8
----	---	----	----	----	----	----	---

31. Calculate Lower and Upper Quartiles:

Roll No.	1	2	3	4	5	6	7	8	9	10
Marks	20	15	25	18	10	9	11	28	35	15

32. Calculate Lower and Upper Quartiles:

5,10,8,12,10

33. Calculate Lower and Upper Quartiles:

25, 15,12,18,15

34. Calculate Lower and Upper Quartiles:

7,6,10,4,3

Discrete Series

35. Calculate median, lower quartile and upper quartile from the following:

Marks	5	10	15	20	25	30	35	40
No. of	3	7	10	5	6	10	2	4

Students								
-----------------	--	--	--	--	--	--	--	--

36. Calculate lower quartile and upper quartile from the following:

Size	10	20	30	40	50	60
Frequency	7	10	15	25	10	5

37. Calculate lower quartile and upper quartile from the following:

X	2	4	6	8	10
F	2	1	5	1	1

38. Calculate lower quartile and upper quartile from the following:

x	5	10	15	20	25	30
f	4	3	2	8	2	1

39. Calculate lower quartile and upper quartile from the following:

x	1	2	3	4	5
f	2	1	5	1	1

40. Calculate upper and lower quartile from the following data:

Marks	0-20	20-40	40-60	60-80	80-100
No. of Students	6	4	15	8	7

41. Calculate upper and lower quartile from the following data:

x	0-2	2-4	4-6	6-8	8-10
f	5	2	3	8	2

42. Calculate upper and lower quartile from the following data:

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

Multiple Choice Questions (MCQs)

1. **Mean of 0.3,5,6,7,9,12,0.6 is:**
(a) 4.9 (b) 5.7
(c) 5.6 (d) None of these
2. **Simple average is sometimes called:**
(a) Unweighted average (b) Weighted average
(c) Relative average (d) None of these
3. **_____ is used when the sum of deviations from the average should be least.**
(a) Mean (b) Mode
(c) Median (d) None of these
4. **Mean should be:**
(a) Simple (b) Based upon all items
(c) Rigidly defined (d) All the above
5. **The algebraic sum of deviations of 8,1,6 from the A.M. viz., 5 is:**
(a) -1 (b) 0
(c) 1 (d) None of these
6. **Average value of given variables is known as:**
(a) Median (b) Mean
(c) Mode (d) Index Number
7. **Measures of central tendency are known as:**
(a) Difference (b) Averages
(c) Both (d) None of these

- 8. The A.M. of 1,3,5,6, x, 10 is 6. The value of x is:**
- (a) 10 (b) 11
(c) 12 (d) None of these
- 9. The values of all items are taken into consideration in the calculation of:**
- (a) Median (b) Mode
(c) Mean (d) None of these
- 10. Sum of the deviations about mean is:**
- (a) Zero (b) Minimum
(c) Maximum (d) One
- 11. Measures of central tendency for a given set of observations measure:**
- (a) The scatter ness of the observations
(b) The central location of the observations
(c) Both (a) and (b)
(d) None of these
- 12. If there are two groups containing 30 and 20 observations and having 50 and 60 as arithmetic means then the combined arithmetic mean is:**
- (a) 51 (b) 54
(c) 53 (d) 52
- 13. Which is the correct formula to determine Arithmetic Mean in case of individual series?**
- (a) $\bar{X} = A + \frac{\sum d'}{N} \times C$ (b) $\bar{X} = \frac{\sum X}{N}$
(c) $\bar{X} = A + \frac{\sum d}{N}$ (d) All of the above
- 14. The mean of 12 numbers is 24. If 5 is added in every number, the new mean is:**
- (a) 25 (b) 84

- (c) 29 (d) None of these

15. Sum of squares of the deviations about mean is:

- (a) Maximum (b) Minimum
(c) Zero (d) None of these

16. _____ average is obtained by dividing the total of set of observations by their number.

- (a) Weighted (b) Simple
(c) Both (a) and (b) (d) Neither (a) nor (b)

17. The mean of a set of observation is $\bar{X}$. If each observation is divided by β and then each observation is increased by 10, then the mean of the new set will be:

- (a) $\frac{\bar{X}}{\beta}$ (b) $\frac{\bar{X}+10}{\beta}$
(c) $\frac{\bar{X}}{\beta} + 10$ (d) $\beta\bar{X} + 10$

18. The most appropriate measure of central tendency in case of data of varying importance:

- (a) Combined Mean (b) Weighted Mean
(c) Assumed Mean (d) All of the above

19. Formula for calculating Combined Arithmetic Mean is:

- (a) $\bar{X}_{1,2} = \frac{\bar{X}_1 N_1 + \bar{X}_2 N_2}{N_1 + N_2}$ (b) $\bar{X}_{1,2,3} = \frac{\bar{X}_1 N_1 + \bar{X}_2 N_2 + \bar{X}_3 N_3}{N_1 + N_2 + N_3}$
(c) Both (a) and (b) (d) Neither (a) nor (b)

20. Which of the following statements is wrong?

- (a) Mean is not affected due to sampling fluctuations
(b) Mean is rigidly defined
(c) Mean has some mathematical properties
(d) All of these

21. The number of observations in a group is 50. If the average of first 20 is 5 and that of the remaining 30 is 4, then the average of the whole group is:

- (a) 4.1 (b) 4
(c) 4.2 (d) 4.4

22. When equal weights are given to all the observation of data then:

- (a) $\bar{X}_w > \bar{X}$ (b) $\bar{X}_w = \bar{X}$
(c) $\bar{X}_w < \bar{X}$ (d) None of these

23. Which of the following is a measure of central tendency?

- (a) Mean (b) Median
(c) Mode (d) All of these

24. Formula to determine Arithmetic Mean through short-cut method in case of individual series is:

- (a) $\bar{X} = A + \frac{\sum d}{N}$ (b) $\bar{X} = \frac{\sum fx}{\sum f}$
(c) $\bar{X} = A + \frac{\sum X}{N}$ (d) $\bar{X} = \frac{\sum X}{N}$

25. The mean weight of a group of 20 items is 25 and that of another group of n items is 34. The mean of combined group of 20+n items is found to be 32. The value of n is:

- (a) 60 (b) 10
(c) 70 (d) 50

26. Which of the following is a method of find Arithmetic Mean?

- (a) Short-cut Method (b) Direct Method
(c) Step Deviation Method (d) All of the above

27. Mathematical property of arithmetic mean is:

- (a) $\Sigma(X - \bar{X}) < 0$ (b) $\Sigma(X - \bar{X}) > 0$
 (c) $\Sigma(X - \bar{X}) = 0$ (d) None of these

28. Weighted Mean is less than simple arithmetic mean when:

- (a) Equal weights are used for all the items in the series or distribution
 (b) Larger weights are assigned to the larger items and smaller weights to the smaller items
 (c) Smaller weights are assigned to larger figures and larger weights to the smaller items
 (d) None of these

29. While computing mean from a grouped frequency distribution, it is assumed that:

- (a) The classes are of equal length
 (b) All the values of a class are equal to the mid-value of that class
 (c) The classes have equal frequency
 (d) None of these

30. The algebraic sum of deviation of observations from their arithmetic mean is:

- (a) 2 (b) -1
 (c) 1 (d) 0

31. Arithmetic Mean in case of individual series can be calculated by the following formula:

- (a) $\bar{X} = \Sigma X - N$ (b) $\bar{X} = \frac{N}{\Sigma X}$
 (c) $\bar{X} = \Sigma X$ (d) $\bar{X} = \frac{\Sigma X}{N}$

32. The mean of a set of numbers is $\bar{X}$. If each number is increased by λ , then mean of the new set is:

- (a) $\bar{X}$ (b) $\lambda\bar{X}$
 (c) $\bar{X} + \lambda$ (d) None of these

33. Which of the following statement is correct?

- (a) Sum of deviations of the observations from their arithmetic mean is always zero.
 (b) Sum of squares of the deviations of the items from their arithmetic mean is minimum
 (c) Weighted mean is less than simple arithmetic mean when larger weights are assigned to the larger.
 (d) Both (a) and (b)

34. Mean of 20 items was found to be 12. On verification, it was found that an item 15 was miscopied as 10. The correct mean is:

- (a) 12.10 (b) 12.25
 (c) 12.20 (d) None of these

35. Weighted mean is computed by the formula:

- (a) $\frac{\sum WX}{\sum W}$ (b) $\frac{\sum WX}{\sum X}$
 (c) $\frac{\sum W}{\sum X}$ (d) None of these

36. In a distribution, the value around which the items tend to be most heavily concentrated is called:

- (a) Median (b) Mean
 (c) Third Quartile (d) Mode

37. The sum of absolute deviations is minimum when taken from.

- (a) Arithmetic Mean (b) Median
 (c) Mode (d) None of these

- 38. The values of extreme items do not influence the average in case of :**
- (a) Median (b) Mean
(c) Mode (d) None of these
- 39. Median:**
- (a) Can be determined graphically (b) Affected by extreme items
(c) Cannot be determined (d) Involves complex calculations
- 40. Which of the following is not a measure of central tendency ?**
- (a) Mean (b) Median
(c) Mode (d) Range
- 41. The second quartile is known as:**
- (a) Median (b) lower quartile
(c) Upper Quartile (d) None of these
- 42. The number of observations smaller than, is the same as the number larger than it.**
- (a) Median (b) Mode
(c) Mean (d) None of these
- 43. Graphic location of mode is done with reference to:**
- (a) Cumulative frequency curve (b) Frequency Polygon
(c) Frequency Curve (d) Histogram
- 44. The most widely average used is:**
- (a) Median (b) Mean
(c) Third Quartile (d) Mode

- 45. For an open-end classification, which of the following is the best measure of central tendency:**
- (a) Arithmetic Mean (b) Geometric Mean
(c) Median (d) Mode
- 46. What do you call the partition value which divides the series into two equal parts?**
- (a) Upper Quartile (b) Median
(c) Mode (d) Lower Quartile
- 47. The middle most value of set of observations is:**
- (a) Median (b) Mode
(c) Mean (d) None of these
- 48. Mode of a series is**
- (a) An average value (b) A middle value
(c) Highest frequency value (d) None of above
- 49. If mean of a series is 32 and median is 40, what would be the value of mode?**
- (a) 54 (b) 58
(c) 56 (d) 38
- 50. Most frequent occurring value in a series is called:**
- (a) Mode (b) Median
(c) Mean (d) Quartiles
- 51. Median can be calculated from _____ series:**
- (a) Individual (b) Discrete
(c) Continuous (d) All of the above
- 52. A distribution with more than two modes is called:**

- (a) Bi-modal
- (b) Multi-modal
- (c) Uni-modal
- (d) None of these

53. In a frequency distribution of a large number of values, the mode is:

- (a) Smallest value
- (b) Largest observation
- (c) Observation with maximum frequency
- (d) Maximum frequency of an observation

54. To find the median, it is necessary to arrange the data in:

- (a) Descending order
- (b) Ascending order
- (c) Ascending or descending order
- (d) Any Random Order

55. For the observations 5,3,6,3,5,10,7,2, there are _____ modes.

- (a) 2
- (b) 3
- (c) 4
- (d) 5

56. A grouping table has:

- (a) 4 columns
- (b) 6 columns
- (c) 8 columns
- (d) None of these

57. The value of a variate that occurs most often is called:

- (a) Median
- (b) Mean
- (c) Mode
- (d) None of these

58. In case of an even number of observations, which of the following is median?

- (a) Any of the two middle-most value
- (b) The simple average of these two middle values
- (c) The weighted average of those two middle values
- (d) Any of these

59. 50% of actual values will be below and 50% of will be above:

- | | |
|----------|-------------------|
| (a) Mode | (b) Median |
| (c) Mean | (d) None of these |

60. Mode of 0,3,5,7,9,12,3 is

- | | |
|-------|-------|
| (a) 6 | (b) 0 |
| (c) 3 | (d) 5 |

61. A frequency distribution having two modes is said to be:

- | | |
|--------------|------------------|
| (a) Unimodal | (b) Bimodal |
| (c) Trimodal | (d) Without mode |

62. Histogram is useful to determine graphically the value of:

- | | |
|----------|-------------------|
| (a) Mean | (b) Median |
| (c) Mode | (d) All the above |

63. Half of the number in an ordered set has values less than:

- | | |
|----------|-------------------|
| (a) Mean | (b) Median |
| (c) Mode | (d) None of these |

64. If mode is ill defined then it is calculated with the help of formula:

- | | |
|---|---|
| (a) $\text{Mode} = 2 \text{ Median} - 3 \text{ mean}$ | (b) $\text{Mode} = 2 \text{ Median} + 3 \text{ mean}$ |
| (c) $\text{Mode} = 3 \text{ Median} + 3 \text{ mean}$ | (d) $\text{Mode} = 3 \text{ Median} - 2 \text{ mean}$ |

65. c.f. is used for :

- | | |
|-------------------|--------------------------|
| (a) Common factor | (b) Cumulative frequency |
| (c) Common value | (d) None of above |

66. When the distribution is symmetrical, mean, median and mode:

- | | |
|--------------|---------------------|
| (a) Coincide | (b) Do not coincide |
|--------------|---------------------|

- (c) Either (a) or (b) (d) None of these
67. **Median of 2, 5, 8, 4, 9, 6, 7 is :**
 (a) 9 (b) 8
 (c) 7 (d) 6
68. **In case of continuous frequency distribution, the size of _____ indicates class-interval in which the median lies.**
 (a) $\left[\frac{N+1}{2}\right]^{th}$ item (b) $\left[\frac{N}{2}\right]^{th}$ item
 (c) $\left[\frac{N-1}{2}\right]^{th}$ item (d) None of these
69. **Which one of the following average can be computed with the help of ogives?**
 (a) Simple Mean (b) Weighted Mean
 (c) Mode (d) Median
70. **In case of an even number of observations, which of the following is median?**
 (a) Weighted average of the two middle values
 (b) Simple average of the two middle values
 (c) Any of the two middle-most values
 (d) None of these
71. **To calculate _____, it is essential to make class-intervals equal and frequencies have to be adjusted.**
 (a) Mean (b) Mode
 (c) Median (d) None of these
72. _____ **is called a positional measure.**
 (a) Median (b) Mode
 (c) Mean (d) None of these

73. For ordering shoes of various sizes for resale, a _____ size will be more appropriate.

- | | |
|------------|-------------------|
| (a) Modal | (b) Mean |
| (c) Median | (d) None of these |

74. Quartiles divide a series into _____ parts:

- | | |
|-------|-------|
| (a) 2 | (b) 3 |
| (c) 4 | (d) 5 |

75. Which average is most suitable in case of calculating average intelligence of different employees in an office?

- | | |
|------------|---------------------|
| (a) Mode | (b) Mean |
| (c) Median | (d) Median and Mode |

76. The following value can be located through graph:

- | | |
|----------------------|-------------------|
| (a) Mode | (b) Median |
| (c) Both (a) and (b) | (d) None of these |

77. The most common value in a series is:

- | | |
|------------|------------------|
| (a) Mode | (b) Mean |
| (c) Median | (d) All of these |

78. Which of the following formula is used to find out median?

- | | |
|--|--|
| (a) $Me = I_1 + \frac{\frac{N}{2} - f}{c.f.} \times i$ | (b) $Me = I_2 + \frac{\frac{N}{2} - c.f.}{f} \times i$ |
| (c) $Me = I_1 + \frac{N - c.f.}{f} \times i$ | (d) $Me = I_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$ |

79. In case of discrete and continuous series, mode can be calculated by:

- | | |
|-----------------------|---------------------|
| (a) Inspection Method | (b) Grouping Method |
| (c) Both (a) and (b) | (d) None of these |

80. Which Quartile is called Median?

- (a) Q_1 (b) Q_2
 (c) Q_3 (d) None of these

Answer MCQs

1	2	3	4	5	6	7	8	9	10
B	A	A	D	B	B	B	B	C	A
11	12	13	14	15	16	17	18	19	20
B	B	D	C	B	B	C	B	C	A
21	22	23	24	25	26	27	28	29	30
D	B	D	A	C	D	C	C	B	D
31	32	33	34	35					
D	C	D	B	A					

36	37	38	39	40	41	42	43	44	45
D	B	A	A	D	A	A	D	A	C
46	47	48	49	50	51	52	53	54	55
B	A	C	C	A	D	B	C	D	A
56	57	58	59	60	61	62	63	64	65
B	C	B	B	C	B	C	B	D	B
66	67	68	69	70	71	72	73	74	75
A	D	B	D	B	B	A	A	C	C
76	77	78	79	80					
C	A	D	C	B					

FORMULAE

1. SIMPLE MEAN Individual Series Direct method	$\bar{X} = \frac{\Sigma X}{N}$	$\bar{X}$ = Arithmetic Mean ΣX = Summation of values of Variables X N = Number of observations
Short-cut Method	$\bar{X} = A + \frac{\Sigma d}{N}$	A = Assumed Mean Σd = Sum of deviations of variables from assumed mean
Step Deviation Method	$\bar{X} = A + \frac{\Sigma d'}{N} \times C$	$\Sigma d'$ = Sum of step deviations C = Common Factor
Discrete Series Direct Method	$\bar{X} = \frac{\Sigma fX}{\Sigma f}$	ΣfX = Sum of product of variables (X) and frequencies (f) Σf = Total of frequencies
Short-cut Method	$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$	Σfd = Sum of product of deviations (d) and respective frequencies (f)
Step Deviation Method	$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$	$\Sigma fd'$ = Sum of product of step deviation (d') and respective frequencies (f)
Continuous Series Direct Method	$\bar{X} = \frac{\Sigma fm}{\Sigma f}$	M = Mid –points Σfm = Sum of product of mid-points (m) and frequencies (f)
Short-cut Method	$\bar{X} = A + \frac{\Sigma fd}{\Sigma f}$	Σfd = Sum of product of deviations (d) from mid points with the respective frequencies (f)

Step Deviation Method	$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$	$\Sigma fd'$ = Sum of product of step deviation (d') and respective frequencies (f)
2. Combined Mean	$(\bar{X}_{1,2}) = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}$	$\bar{X}_{1,2}$ = Combined Mean $\bar{X}_1$ = Arithmetic Mean of first distribution $\bar{X}_2$ = Arithmetic Mean of second distribution N_1 = Number of items of first distribution N_2 = Number of items of second distribution
3. Weighted Mean	$\bar{X}_w = \frac{\Sigma WX}{\Sigma W}$	$\bar{X}_w$ = Weighted Mean ΣWX = Sum of product of items and respective weights ΣW = Sum of the weights
1. Median Individual Series	Me = Size of $\left(\frac{N+1}{2}\right)^{\text{th}}$ item Average of two items lying on either side of $\left(\frac{N+1}{2}\right)^{\text{th}}$	(In odd number series) (In Even Number series)
Discrete Series	Me = Size of $\left(\frac{N+1}{2}\right)^{\text{th}}$ item	
Continuous Series	Determine Median Class as $\left[\frac{N}{2}\right]^{\text{th}}$ item and apply the formula: $\text{Me} = I_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$	
2. Lower Quartile	Q_1 = Size of $\left(\frac{N+1}{4}\right)^{\text{th}}$ item	
Discrete Series	Q_1 = Size of $\left(\frac{N+1}{4}\right)^{\text{th}}$ item	

Continuous Series	Determine Quartile Class as $\left[\frac{N}{4}\right]^{th}$ item and apply the formula: $Q_1 = I_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$
Upper quartile individual series	$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)^{th} \text{ item}$
Discrete Series	$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)^{th} \text{ item}$
Continuous Series	Determine Quartile Class as $3\left[\frac{N}{4}\right]^{th}$ item and apply the formula : $Q_3 = I_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i$
4. Mode	
Individual Series	Mode is the value, which occurs largest number of times.
Discrete Series	If the frequencies are regular and homogeneous and there is a single maximum frequency , then Mode is the value corresponding to the highest frequency (Otherwise use Grouping Method)
Continuous Series	Step 1: Determine the Modal Class: (i) By inspection, if frequencies are regular homogeneous and there is a single maximum frequency; Otherwise (ii) Grouping Method.
	Step 2: Apply the following formula : Mode $(Z) = I_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$

[Chapter – 7]

CORRELATION

Meaning

Correlation indicates the relationship between two variables of a series so that changes in the values of one variable are associated with changes in the values of the other variable.

Correlation and Causation

Causation means that one variable causes another event to occur. If there is correlation between two variables, correlation does not imply causation.

(1) Effect of third Variable – It is possible that a high degree of correlation between the two variable may be due to the influence of a third variable not included in the analysis.

For Example – A high degree of correlation between the yield per acre of wheat and of maize may be due to the fact that both are related to the amount of rainfall. But none of the two variables is the cause of the other.

(2) Mutual Dependence – When two variables show a high degree of correlation then it may be difficult to explain which variable is influencing the other i.e. it is difficult to say which is the cause and which is effect because both reacting on each other.

(3) Pure Chance – The correlation between the two variables may be obtained due to pure chance. So while interpreting essential to see if there is likelihood of any relationship between variables under study.

Importance of Correlation

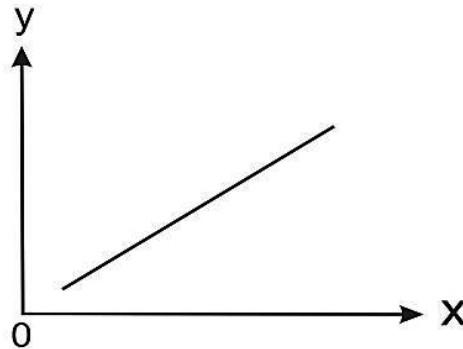
Causation means that one variable causes

- (1) It helps in measuring the relationship between two variables in one figure.
- (2) It analysis facilitates understanding of economic behavior.
- (3) When two variables are correlated then value of one variable can be estimated, given the value of another.

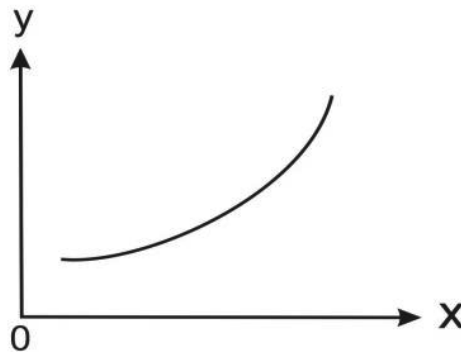
Types of Correlation

It can be classified into following categories.

1. **Positive correlation**– When two variables move in the same direction i.e. when one increases the other also increases and when one decreases the other also decreases.
2. **Negative Correlation**– When two variables move in opposite directions, i.e. when one increases the other decreases and when one decreases the other increases then such a relation is called negative correlation.
3. **Linear Correlation**– Linear correlation is said to exist if the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable.



4. **Non- Linear** – In non-linear correlation the amount of change in one variable does not bear a constant ratio to the amount of change in the other related variable.



5. **Simple Correlation** – When only two variables are studied it is a problem of simple correlation. For example, price and demand etc.
6. **Multiple Correlation** – When the relationship among three or more than three variables is studied simultaneously, then such relationship is called multiple correlation for example, output of maize due to rainfall etc.
7. **Partial Correlation** – The relationship between two variables is examined keeping other variables as constant.

Methods / Techniques of Measurement of correlation—

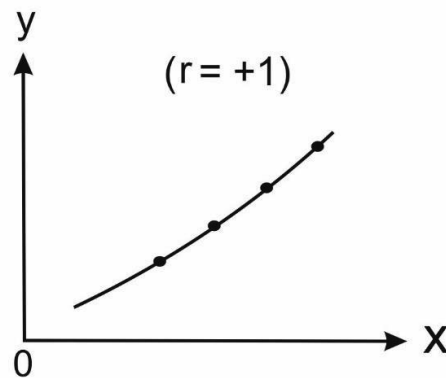
1. **Scatter Diagram**

2. **Karl Pearson's coefficient of Correlation**
3. **spearman's Rank Correlation Coefficient**

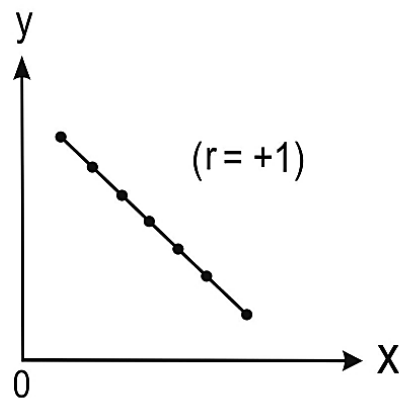
Scatter Diagram—

It is a simple and attractive method of diagrammatic representation of a bivariate distribution to determine the nature of correlation between the variables.

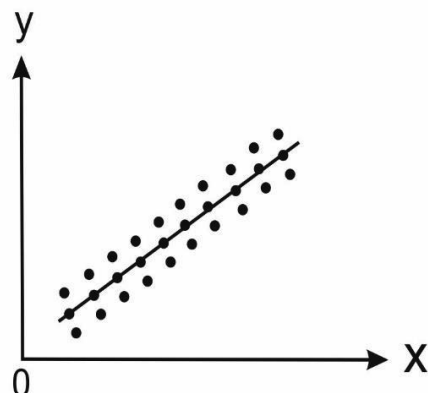
- (1) **Perfect positive Correlation** –If all the points of a scatter diagram fall on a straight line with positive slope.



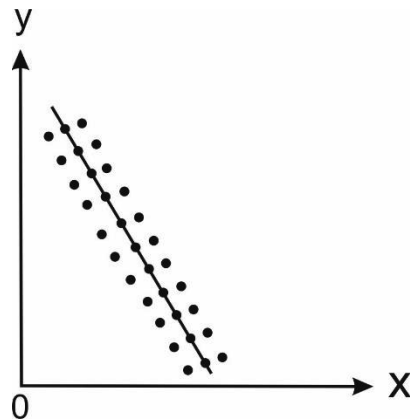
- (2) **Perfect Negative Correlation**— If all the points of a scatter diagram fall on a straight line with negative slope.



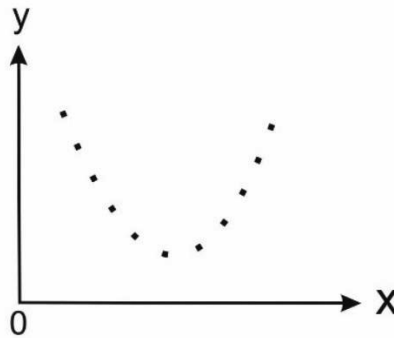
3. **Positive Correlation**— When all the points of a scatter diagram cluster around a straight line going upwards from left to right the correlation is positive correlation.



4. **Negative Correlation**—When all the points of a scatter diagram cluster around a straight line with negative slope the correlation is said to be negative.



5. **No Correlation**—If the points are scattered in a haphazard manner then it is a case of zero or no correlation.



Merits of Scatter Diagram

- (1) **Simplicity**—It is a simple method of studying correlation between two variables.
- (2) **Easily understandable**— It can be easily understood.

Demerits of Scatter Diagram

- (1) **Non-Mathematical Method**—This method does not indicate the exact numerical value of correlation which is possible by other mathematical methods of correlation.
- (2) **Rough Measures**— It gives only a broad and rough idea of the degree and nature of correlation between two variables.

Karl Pearson's Coefficient of Correlation

It is a mathematical technique which describes the co variation in ratio terms which was developed by Karl Pearson in 1890. It is also known as product moment correlation or simple coefficient of correlation.

It is a degree of linear association between the two variables X and Y is measured by the Karl Pearson's Coefficient of correlation. The Coefficient of correlation between X and Y will be

denoted by r_{xy} or r .

$$r = \frac{\Sigma dxdy}{N \sigma_x \times \sigma_y}$$

$\Sigma dxdy$ = Sum of product of deviations of x and y variable from their mean.

$$\Rightarrow dx = (x - \bar{x})$$

$$\Rightarrow dy = (y - \bar{y})$$

$\Rightarrow 6x$ = standard deviation of x series

$$\left[\sqrt{\frac{\Sigma dx^2}{N}} \right]$$

$\Rightarrow$ Standard deviation of y series.

$$\left[\sqrt{\frac{\Sigma dy^2}{N}} \right]$$

$\Rightarrow N$ = Number of pair of observations.

$$\text{Covariance} = \frac{\Sigma dxdy}{N} = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{N}$$

$$r = \frac{\Sigma dxdy}{N \times 6x \times \sigma_y} \dots\dots\dots (i)$$

$$\text{Covariance (x,y)} = \frac{\Sigma dxdy}{N}$$

$$r = \frac{\text{Cov.}(x,y)}{\sigma_x \cdot \sigma_y} \dots\dots\dots (ii)$$

$$\sigma_x = \sqrt{\frac{\Sigma dx^2}{N}}, \sigma_y = \sqrt{\frac{\Sigma dy^2}{N}}$$

From equation (i)

$$r = \frac{\Sigma dxdy}{N \times \sqrt{\frac{\Sigma dx^2}{N}} \times \sqrt{\frac{\Sigma dy^2}{N}}} = \frac{\Sigma dxdy}{N \times \frac{1}{N} \sqrt{\Sigma dx^2 \times \Sigma dy^2}}$$

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \times \Sigma dy^2}} \dots\dots\dots (iii)$$

EXAMPLE 1

From the following data, calculate coefficient of correlation between x and y.

	X Series	Y Series
Number of I Tons	20	20
Standard Deviation	5	3

Sum of product of deviations of x and y from their respective means = 200

SOLUTION

$$\sigma_x = 5, \sigma_y = 3, N = 20, \Sigma dxdy = 200$$

$$r = \frac{\Sigma dxdy}{N \times \sigma_x \times \sigma_y}$$

$$= \frac{200}{20 \times 5 \times 3} = \frac{200}{300} = 0.666$$

r = 0.67 Ans.

EXAMPLE 2

- (i) If Covariance between x and y variable is 10 and variance of x and y variables are 25 and 9 respectively, find coefficient of correlation.
- (ii) If the correlation coefficient between x and y is +0.62 and the variances of x and y series are respectively 30 and 108, find out the co-value between the series.
- (iii) Co-efficient of correlation between two variables x and y is 0.3, their covariance is 8 and the variance of x is 9, find the standard deviations of y – series.

SOLUTION

- (i) Covariance of x and y = 10

$$\text{Variance of } x = 25, \sigma_x = \sqrt{V} = \sqrt{25} = 5$$

$$\text{Variance of } y = 9, \sigma_y = \sqrt{V} = \sqrt{9} = 3$$

$$r = \frac{\text{Covariance of x and y}}{\sigma_x \cdot \sigma_y}$$

$$\frac{10}{5 \times 3} = \frac{10}{15} = 0.666$$

r = 0.67 Ans.

$$r = +0.62$$

- (ii) Variance of x = 30, $\sigma_x = \sqrt{V} = \sqrt{30} = 5.48$

$$\text{Variance of } y = 108, \sigma_y = \sqrt{V} = \sqrt{108} = 10.39$$

$$r = \frac{\text{Covariance of x and y}}{\sigma_x \cdot \sigma_y}$$

$$\frac{0.62}{1} = \frac{\text{Covariance of x and y}}{5.48 \times 10.38}$$

$r = 35.3$ Ans.

(iii) $r = 0.3$

Covariance = 8

Variance of $x = 9$, $\sigma_x = \sqrt{V} = \sqrt{9} = 3$

$\sigma_y = ?$

$$r = \frac{\text{Covariance of } x \text{ and } y}{\sigma_x \cdot \sigma_y}$$

$$= \frac{0.3}{1} = \frac{8}{3 \times \sigma_y}$$

$$= 0.9 \sigma_y = 8$$

$$\sigma_y = \frac{8}{0.9} = 8.89$$

$\sigma_y = 8.89$ Ans.

Karl Pearson's coefficient of correlation can be calculated by following methods:

- (A) Actual Mean Method
- (B) Assumed Mean Method / Short cut Method
- (C) Direct Method

(A) Actual Mean Method

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}}$$

EXAMPLE 3

Calculate coefficient of correlation from the following data:

X	2	3	5	6	9
Y	2	5	6	7	10

SOLUTION

x	$dx(X - \bar{X})$	dx^2	y	$dy(y - \bar{y})$	dy^2	dxd
3	-3	9	2	-4	16	12
2	-2	4	5	-1	1	2

5	0	0	6	0	0	0
6	1	1	7	1	1	1
9	4	16	10	4	16	16
$\sum x = \frac{25}{n=5}$	$\sum dx = 0$	$\sum dx^2 = 30$	$\sum y_2 30$	$\sum dy = 0$	$\sum dy^2 34$	$\sum d x dy 31$

$$\bar{X} = \frac{\sum x}{N} = \frac{25}{5} = 5$$

$$\bar{X} = 5$$

$$\bar{Y} = \frac{\sum y}{N} = \frac{30}{5} = 6$$

$$r = \frac{\sum dx dy}{\sqrt{\sum dx^2 \sum dy^2}}$$

$$r = \frac{31}{\sqrt{30 \times 34}} = \frac{31}{\sqrt{1020}} = \frac{31}{\sqrt{31.937}}$$

$$r = 0.97$$

(B) Short cut/Assured Mean Method

$$r = \frac{\sum dxdy \times N - (\sum dx \times \sum dy)}{\sqrt{\sum dx^2 \times N - (\sum dx^2)} \sqrt{\sum dx^2 \times N - (sdy)^2}}$$

EXAMPLE 4

Calculate Coefficient of Correlation of the data given in illustrations.

SOLUTION

x	$d(x - A)$	d^2x^2	y	$d(y - A)$	d^2y^2	$dxdy$
3	-4	16	2	-3	9	12
2	-3	9	5	0	0	0
5	-1	1	6	+1	1	-1
6	0	0	7	+24	4	0
9	+3	9	10	+5	25	15
$N = 5$	$\sum dx = 5$	$\sum dx^2 = 35$		$\sum dy = 5$	$\sum dy^2 = 39$	$\sum dxdy = 46$

$$r = \frac{\sum dxdy \times N - (\sum dx \times \sum dy)}{\sqrt{\sum dx^2 \times N - (\sum dx)^2} \sqrt{\sum dy^2 \times N - (\sum dy)^2}}$$

$$r = \frac{26 \times 5 - (-5 \times 5)}{\sqrt{35 \times 5 - (5)^2} \sqrt{39 \times 5 - (5)^2}}$$

$$r = \frac{130 + 25}{\sqrt{175 - 25} \sqrt{195 - 25}}$$

$$r = \frac{155}{\sqrt{150} \sqrt{170}}$$

$$r = \frac{155}{\sqrt{25500}} = \frac{155}{159.687} = 0.97$$

(C) Direct Method

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

EXAMPLE 5

Calculate Coefficient of Correlation of the data given in illustrations.

x	x^2	y	y^2	xy
2	4	2	4	4
3	9	5	25	15
5	25	6	36	30
6	36	7	49	42

9	81	10	100	90
$\Sigma x = 25$	$\Sigma x^2 = 155$	$\Sigma y = 30$	$\Sigma y^2 = 214$	$\Sigma xy = 181$

$$r = \frac{\Sigma xy \times N - (\Sigma x \times \Sigma y)}{\sqrt{\Sigma x^2 \times N - (\Sigma x)^2} \sqrt{\Sigma y^2 \times N - (\Sigma y)^2}}$$

$$r = \frac{181 \times 5 - (25 \times 30)}{\sqrt{155 \times 5 - (25)^2} \sqrt{214 \times 5 - (30)^2}}$$

$$r = \frac{905 - 750}{\sqrt{775 - 625} \sqrt{1070 - 900}}$$

$$r = \frac{155}{\sqrt{150} \sqrt{170}}$$

$$r = \frac{155}{25500}$$

$$r = \frac{155}{159.687}$$

$$r = 0.97$$

Assumption of Coefficient of Correlation

1. **Linear Relationship:** - It is assumed that there is linear relationship between the two series under study.
2. **Normality:** - Each of the variables in a series is being affected by a large number of independent contributory causes of such a nature which form a normal distribution.
3. **Causal Relationship:** - If there is cause and effect relationship between the for us affecting the distribution of items in the two series if there is no Correlation it is meaningless.

Merits

1. **Study of Degree:** - It Coefficient measures both the direction and degree of Correlation.
2. **Algebraic Treatment:** - It is used for further algebraic treatment.

Demerits

1. **Time Consuming Method:** - As Compared to other methods it more time consuming.
2. **Possibility of wrong interpretation:** - One has to be very careful in interpreting the value of the coefficient of correlation as very often the coefficient is misinterpreted.

Spearman's Rank Correlation

The measure of rank correlation was developed by British psychologist 'Charles Edward Spearman' in 1904. In this method various items are assigned rank according to their characteristics and a correlation is computed between these ranks.

Formula

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

r_k = Coefficient of rank Correlation

$\sum D^2$ = Sum of Squares of Rank Differences

N = Number of pairs of observations

Calculation of Rank Correlation

1. When Ranks are given
2. When Rank are not given
3. When Rank are equal

(1) When Ranks are Given**EXAMPLE 6**

In a singing competition two judges rand the 7 contestants as follows.

Judge 1	5	4	7	3	1	2	6
Judge 2	6	5	2	1	3	4	7

Calculate Coefficient of rank correlation

SOLUTION

Rank Judge 1 (R_1)	Rank Judge 1 (R_2)	($R_2 - R_1$)	D^2
5	6	-1	1
4	5	-1	1
7	2	+5	25
3	1	+2	4

1	3	-2	4
2	4	-2	4
6	7	-1	1
N = 7			$\Sigma D^2 = 40$

$$rk = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

$$rk = 1 - \frac{6(40)}{7^3 - 7}$$

$$rk = 1 - \frac{240}{336}$$

$$rk = 0.285.$$

(2) When Ranks are not given

EXAMPLE 7

Calculate rank correlation of coefficient by Judge A and B in a competition test as shown below:

Marks by Judge A	52	53	42	60	45	41	37	38	25
Marks by Judge B	65	68	43	38	77	48	35	30	25

SOLUTION

Marks by Judge A	Marks by Judge B	R₁	R₂	(R₂ - R₁) D	D³
52	65	3	3	0	0
53	68	2	2	0	0
42	43	5	6	-1	1
60	38	1	7	-6	36
45	77	4	1	3	9
11	48	6	5	1	1
7	35	8	8	0	0
8	30	7	9	-2	4
5	25	10	10	0	0

7	50	9	4	5	25
30					$\Sigma S^2 = 76$

$$rk = 1 - \frac{6\Sigma DN^2}{N^3 - N}$$

$$rk = 1 - \frac{6(76)}{(10)^3 - 10}$$

$$rk = 1 - \frac{456}{990}$$

$$rk = 0.539$$

(3) When Ranks are equal

$$rk = 1 - \frac{6 \left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m + \dots) \right]}{N^3 - N}$$

EXAMPLE 8

Calculate the Coefficient of Correction on of the following data by the Spearman's Rank Correlation method.

X	19	24	12	23	19	16
Y	9	22	20	14	22	18

SOLUTION

In X Series 1st & 2nd rank will assign to 12 and 16 respectively

As value 19 is repeated twice,

$$\frac{3+4}{2} = 3.5$$

Now, 5th & 6th will be assigned to 23 and 24 respectively

In y series As value 22 is repeated twice,

$$\frac{5+6}{2} = 5.5$$

X	Y	R₁	R₂	(R₂ - R₁) D	D²
19	9	3.5	1	2.5	6.25
24	22	6	5.5	0.5	0.25
12	20	1	4	-3	9

23	14	5	2	3	9
19	22	3.5	5.5	-2	4
16	18	2	3	-1	1
N = 6					$\Sigma S^2 = 29.50$

$$rk = 1 - \frac{6 \left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m + \dots) \right]}{N^3 - N}$$

$$rk = 1 - \frac{6 \left[29.50 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) \right]}{6^3 - 6}$$

$$rk = 1 - \frac{6(29.50 + 0.5 + 0.5)}{216 - 6}$$

$$rk = 1 - \frac{6 \times 30.5}{210}$$

$$rk = 0.128$$

Merits and Demerits of Rank Correlation

Merits

1. This methods is easy to calculate and simple to understand.
2. Rank method is very useful when the data is qualitative in nature like honesty, intelligence, voice quality etc.

Demerits

1. Rank method cannot be used for finding out correlation in a grouped frequency distribution.
2. Rank method lacks precision as composed to Karl Pearson's in method.

FORMULAE AT A GLANCE	
Karl Pearson's of Correlation	
1. Actual Mean Method	$r = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y} = \frac{\Sigma xy}{N \times \sqrt{\frac{\Sigma x^2}{N}} \times \sqrt{\frac{\Sigma y^2}{N}}} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$
2. Direct Method	$r = \frac{N \Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{N \Sigma X^2 - (\Sigma X)^2} \times \sqrt{N \Sigma Y^2 - (\Sigma Y)^2}}$
3. Short – Cut Method	$r = \frac{N \Sigma Dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N \Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N \Sigma dy^2 - (\Sigma dy)^2}}$
4. Step Deviation Method	$r = \frac{N \Sigma dx' dy' - \Sigma dx' \times \Sigma dy'}{\sqrt{N \Sigma dx'^2 - (\Sigma dx')^2} \times \sqrt{N \Sigma dy'^2 - (\Sigma dy')^2}}$

Spearman's Rank Correlation Coefficient	
When Ranks are not Equal	$r_k = 1 - \frac{6\sum D^2}{N^3 - N}$
When Ranks are Equal	$r_k = 1 - \frac{6\left(\sum D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) \dots\right)}{N^3 - N}$

Multiple Choice Questions

1. **A scatter diagram:**
 - (a) Is a statistical test
 - (b) Must be linear
 - (c) Must be curvilinear
 - (d) Is a graph of x and y values
2. **Maximum value of rank correlation coefficient is:**
 - (a) 0
 - (b) +1
 - (c) -1
 - (d) None of these
3. **If the relationship between x and y is positive, as variable y decreases, variable x:**
 - (a) Increases
 - (b) Decreases
 - (c) Remains same
 - (d) Changes linearly
4. **The correlation coefficient will be - 1 if the slope of the straight line in a scatter diagram is:**
 - (a) Positive
 - (b) Negative
 - (c) Zero
 - (d) None of these
5. **In a 'negative' relationship:**
 - (a) As x increases, y increases
 - (b) As x decreases, y decreases
 - (c) As x increases, y decreases
 - (d) Both (a) and (b)
6. **If there is a perfect disagreement between the marks in Geography and Statistics, then what would be the value of rank correlation coefficient?**
 - (a) 1
 - (b) Any value
 - (c) -1
 - (d) (b) or (c)
7. **The correlation between ages of husbands and wives is:**
 - (a) Positive
 - (b) Negative
 - (c) Zero
 - (d) None of these

8. Scatter diagram helps us to:

- (a) Find the nature of correlation between two variables
- (b) Obtain the mathematical relationship two variables
- (c) Compute the extent of correlation between two variables
- (d) Both (a) and (c)

9. The lowest strength of association is reflected by which of the following correlation coefficients?

- (a) 0.95
- (b) -0.60
- (c) -0.35
- (d) 0.29

10. Karl Pearson's coefficient is defined from:

- (a) Ungrouped data
- (b) Grouped data
- (c) Both (a) and (b)
- (d) None of these

11. When $r = 1$, all points in a scatter diagram would lie:

- (a) On a straight line directed from lower left to upper right
- (b) On as straight line
- (c) On a straight line directed from upper left to lower right
- (d) Both (a) and (c)

12. Tick the incorrect statement.

- (a) Correlation Coefficient lies between 0 and -1.
- (b) Karl Pearson's method calculates deviations from actual or assumed mean
- (c) Spearman's method of rank correlation measures coefficient of correlation for quantitative data
- (d) Both (a) and (c)

13. There is a high degree of negative correlation between 'overweight' and 'life expectancy'. A correlation coefficient consistent with the above statement is:

- (a) $r = 0.80$ (b) $r = 0.20$
(c) $r = -0.20$ (d) $r = -0.80$
- 14. Correlation coefficient is _____ of the units of measurement.**
(a) Independent (b) Dependent
(c) Both (a) and (b) (d) None of these
- 15. The correlation between sale of cold drinks and day temperature is:**
(a) Positive (b) Negative
(c) Zero (d) None of these
- 16. There is as high direct association between measures of 'cigarette smoking' and 'lung damage correlation coefficient with the above statement is:**
(a) 0.30 (b) 0.80
(c) -0.80 (d) -0.30
- 17. Simple correlation is called:**
(a) Linear correlation (b) Nonlinear correlation
(c) Both (a) and (b) (d) None of these
- 18. For finding the degree of agreement about beauty between two judges in a beauty contest,**
(a) Coefficient of correlation (b) Coefficient of rank correlating
(c) Scatter diagram (d) Coefficient of concurrent deviation
- 19. If all the plotted points in a scatter diagram lie on a single line, then the correlation is:**
(a) Perfect positive (b) Perfect negative
(c) Both (a) and (b) (d) Either (a) or (b)

- 20. If r is the correlation coefficient between the two variables, then:**
- (a) $-1 \leq r \leq 0$ (b) $1 \leq r \leq 2$
 (c) $-1 \leq r \leq 1$ (d) $0 \leq r \leq 1$
- 21. The correlation between sale of woolen garments and day temperature is:**
- (a) Positive (b) Negative
 (c) Zero (d) None of these
- 22. The correlation between shoe size and intelligence is:**
- (a) Zero (b) Negative
 (c) Positive (d) None of these
- 23. Which of the following statements is correct?**
- (a) Relationship between height and weight is an example of negative correlation
 (b) Coefficient of correlation is independent of change of scale
 (c) Coefficient of correlation is independent of change of origin
 (d) Both (b) and (c)
- 24. The highest strength of association is reflected by which of the following correlations coefficient:**
- (a) -1.0 (b) -0.95
 (c) 0.1 (d) 0.85

MCQs Answers

1	2	3	4	5	6	7	8	9	10	11	12
D	B	B	B	C	C	A	A	D	A	A	D
13	14	15	16	17	18	19	20	21	22	23	24
D	A	A	B	A	B	D	C	B	A	D	A

Fill in the Blanks

- _____ is said to exist if the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable.
- Under _____, the relationship between two variables is examined keeping other variables as constant.
- _____ is a diagrammatic representation of a bivariate distribution to determine the nature of correlation between the variables.
- The Karl Pearson's coefficient of correlation is also known as _____ or _____.
- Correlation coefficient ranges between _____ and _____.
- If two variables X and Y are independent, coefficient of correlation between them will be _____.
- If all the points of scatter diagram fall on a straight line with positive slope, then the correlation is said to be _____.
- When both the variables change in the same direction, then correlation is called _____.
- Maximum positive value of coefficient of correlation is _____.

Answer Key

- | | |
|-----------------------|---|
| 1. Linear Correlation | 2. Partial Correlation |
| 3. Scatter Diagram | 4. Product Moment Correlation, Simple Correlation Coefficient |
| 5. +1, - 1 | 6. Zero |
| 7. Perfectly Positive | 8. Positive |
| 9. + 1 | |

1. Find the coefficient of correlation between X and Y series from the data:

X	6	12	10	8	9	7	11
Y	8	12	11	9	10	13	7

2. Determine the coefficient of correlation form the following data between X and Y series:

X	8	6	9	11	13	7	12	10	9	8
Y	11	6	7	10	9	11	12	8	7	6

3. From the following data, calculate coefficient correlation between the variables X and Y using Karl Pearson's method:

X	10	6	9	10	12	13	11	9
Y	9	4	6	9	11	13	8	4

4. Calculate coefficient of correlation form the following data:

Height of fathers (inches)	66	68	69	72	65	59	62	67	61	71
Height of Sons (inches)	65	64	67	69	64	60	59	68	60	64

5. Making use of the data given below, calculate the coefficient of correlation.

X	10	6	9	10	12	13	11	9
Y	9	4	6	9	11	13	8	4

6. Calculate product moment correlation between the values of X and Y:

X	2	3	1	5	6	4
Y	4	5	3	4	6	2

7. Two judges in a beauty competition rank the 12 entries as follows:

X	1	2	3	4	5	6	7	8	9	10	11	12
Y	12	9	6	10	3	5	4	7	8	2	11	1

Calculate the Spearman's rank coefficient of correlation.

8. A group of ten workers of a factory is ranked according to their efficiency by two different judges as follows:

Name of Worker	A	B	C	D	E	F	G	H	I	J
Rank by Judge A	4	8	6	7	1	3	2	5	10	9
Rank by Judge B	3	9	6	5	1	2	4	7	8	10

Compute the coefficient of rank correlation.

9. Ten competitors in a debate contest are ranked by three judges in the following order.

Competitors	A	B	C	D	E	F	G	H	I	J
Rank by 1 st Judge	7	4	10	5	9	8	6	2	1	3
Rank by 2 nd Judge	4	1	9	10	7	3	2	5	6	8
Rank by 3 rd Judge	10	2	8	5	7	6	9	1	4	3

Use the ranking correlation method and state which pair of judges has the nearest approach.

Ranks are not given

10. A group of 8 students got the following marks in a test in Maths and Accountancy.

Marks in Maths	50	60	65	70	75	40	80	85
Rank by 1 st Judge	80	71	60	75	90	82	70	50

Compute the coefficient of rank correlation.

11. Calculate rank correlation between advertisement cost and sales as per the data given below:

Cost (in '000 ₹)	78	36	98	25	75	82	90	62	65	39
Sales (in '000 ₹)	84	51	91	60	68	62	86	58	53	47

Equal Ranks

12. Calculate the coefficient of rank correlation from the following data:

X	48	33	40	9	16	16	65	24	16
Y	13	13	24	6	15	4	20	9	6

Assertion Base Questions

1. **Assertion (A):** The Correlation between height and weight of a person is positive.

Reason (R): Positive Correlation means the movement of two variables in the same direction.

Ans: Both Assertion (A) of Reason (R) are true and Reason (R) is not the correct explanation of Assertion (A).

2. **Assertion (A):** Karl Pearson's method of calculating coefficient of correlation is based on covariance of the concerned variables.

Reason (R): The coefficient of correlation between two variable X and Y will be zero if they are independent.

Ans: Both Assertion (A) of Reason (R) are true and Reason (R) is not the correct explanation of Assertion (A).

3. **Assertion (A):** The scatter diagram gives a visual presentation of the relationship.

Reason (R): The scatter diagram is not confined to linear relations.

Ans: Both Assertion (A) and Reason (R) are true and Reason (R) is not the correct explanation of Assertion (A).

4. **Assertion (A):** If $r = 0$ the variables are uncorrelated.

Reason (R): There is a linear relationship b/w the variables.

Ans: Assertion (A) is true but Reason (R) is false.

5. **Assertion (A):** r has no unit

Reason (R): r is a pure number

Ans: Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).

6. **Assertion (A):** The value of ' r ' is affected by the change of origin.

Reason (R): The value of ' r ' is unaffected by the change of scale.

Ans: Assertion (A) is true but Reason (R) is false.

7. **Assertion (A):** Correlation measures causation.

Reason (R): Correlation studies and measures the direction and intensity of relationship

among variables.

Ans: Assertion (A) is false but Reason (R) is true.

One Word

1. Define Correlation

Ans: It indicates the relationship between two variables of a series so that changes in the values of one variable are associated with changes in the value of the other variable.

2. Give an example of simple correlation.

Ans: Relationship b/w price and demand.

3. What is meant by positive correlation?

Ans: When two variables move in the same direction then such a relation is called positive correlation.

4. What is negative Correlation?

Ans: When two variables move in opposite direction i.e. when one increases the other decreases and when one decreases the other increases then such a relation is called negative Correlation.

5. What is the graphical device for ascertaining whether two variables are related or not?

Ans: Scatter diagram.

6. What are the two mathematical methods for measuring coefficient of correlation?

Ans: (i) Karl Pearson's Method
(ii) Rank Correlation Method

7. State two merits of rank correlation.

Ans: (i) It is a simple and a non-mathematical method.
(ii) It can be easily understood and interpreted.

8. What is covariance?

Ans: It is a statistical representation of the degree to which two variables vary together.

Statement Based Questions

1. **Statement 1:** Correlation measures covariation.

Statement 2: Correlation measures causation.

Ans: Statement 1 is true and Statement 2 is false.

2. **Statement 1:** The Coefficient of Correlation has no unit.

Statement 2: The Coefficient of Correlation is a pure number.

Ans: Both the Statement are true.

3. **Statement 1:** Scatter diagram is not confined to linear relations.

Statement 2: Correlation means Causation.

Ans: Statement 1 is true and Statement 2 is false.

4. **Statement 1:** Repeated ranks need correction factors.

Statement 2: The value of 'r' lies between (-) 1 and 1.

Ans: Both the Statement are True

5. **Statement 1:** The calculation of rank correlation is not possible when ranks are not given.

Statement 2: Spearman's rank correlation is used only when rank are given.

Ans: Both the Statement are False.

Short Answer type Questions

1. What is meant by linear and non-linear correlation?
2. State the different methods for measuring correlation?
3. Draw a scatter diagram showing high degree of positive correlation.
4. How is the problem of equal rank solved?

- 1. The unit of correlation coefficient between height in feet and weight in kgs is**
 - (a) kh/feet
 - (b) Percentage
 - (c) Non-existent

- 2. The range of simple correlation coefficient is**
 - (a) 0 to infinity
 - (b) Mins one to plus one
 - (c) Minus infinity to infinity s

- 3. If r_{xy} is positive the relation between X and Y is of the type**
 - (a) When Y increases X increase
 - (b) When Y decreases X increases
 - (c) When Y increase X does not change

- 4. If $r_{xy} = 0$ the variable X and Y are**
 - (a) Linearly related
 - (b) Not linearly related
 - (c) Independent

- 5. Of the following three measures which can measure any type of relationship**
 - (a) Karl Pearson's coefficient of correlation
 - (b) Spearman's rank correlation
 - (c) Scatter Diagram

- 6. If precisely measured data are available the simple correlation coefficient is**
 - (a) More accurate than rank correlation coefficient
 - (b) Less accurate than rank correlation coefficient
 - (c) As accurate as the rank correlation coefficient

7. Why is r preferred to covariance as a measure of association?
8. Can r lie outside the -1 and range depending on the type of data?
9. Does correlation imply causation?
10. When is rank correlation more precise than simple correlation coefficient?
11. Does zero correlation mean independence?
12. Can simple correlation coefficient measure any type of relationship?
13. Collect the price of five vegetables from your local market every day for a week. Calculate their correlation coefficients. Interpret the result.
14. Measure the height of you classmates. Ask them the height of their benchmate. Calculate the correlation coefficient of these two variables. Interpret the result.
15. List some variables where accurate measurement is difficult.
16. Interpret the values of r as 1, -1 and 0.
17. Why does rank correlation coefficient differ from Pearsonian correlation coefficient?
18. Calculate the correlation coefficient between the heights of fathers in inches (X) and their sons (Y)

X	65	66	57	67	68	69	70	72
Y	67	56	65	68	72	72	69	71

(Ans: $r = 0.603$)

19. Calculate the correlation coefficient between X and Y and comment on their relationship:

X	-3	-2	-1	1	2	3
Y	9	4	1	1	4	9

(Ans: $r = 0$)

20. Calculate the correlation coefficient between X and Y and comment on their relationship:

X	1	3	4	5	7	8
Y	2	6	8	10	14	16

(Ans: $r = 1$)

[Chapter – 8]

INDEX NUMBER

Index number is an average change of various magnitudes expressed in different units. It is also known as the Barometers of Economic activity as they help to get an index of the present day situation with regard to changes in production consumption national Income exports and 'Imports, business level cost' of living etc.

Meaning of Index Numbers

It is a statistical device for measuring change in the magnitude of a group of related variables.

According to Edgeworth- Index number shows by its variation the change in a magnitude which is not susceptible either to accurate measurement in itself or of direct valuation in practice.

Absolute Change- It refers to the simple difference in the value over two periods in time.

Relative Change- It expresses the absolute change as a percentage of the value of the indicator in the earlier period.

Features/Characteristics of Index Numbers

- (a) **Index Numbers are Specialized Averages** – Average like mean median or mode can be used to compare two or more series however, if the units in which two or more series are expressed are different or if the series are composed of different types of items then averages cannot be used to compare them.
- (b) **Selection of Base year** – It is the year which is used as the basis for comparing changes.
- (c) **Selection of Number of Item or Commodities**– It depends on the purpose of index. In a general price index, a large number of commodities may be included.
- (d) **Selection of Data** –After deciding the items to be included in the construction of the index number the problem arises of how to obtain data on prices of selected goods together with their quantities consumed over a period of time.
- (e) **Choice of an Average** –Average play a vital role in arriving at a single index number summarising a large volume of information. Arithmetic mean and geometric mean are used in its construction.

Types or kinds or Index Numbers

1. **Price Index Numbers** – There are more frequently used by different experts such as statistician business man policy maker's researches. Price index numbers are further divided into the following two categories.
 - (a) **Wholesale Price Index Numbers** – They reflect the changes in general price level of the country.
 - (b) **Retail Price Index Numbers** – These index numbers show general changes in retails prices of various commodities such as consumption goods, stock and shares, bank deposits, consumer price index etc.
2. **Quantity Index Numbers** – These index numbers study the changes in the volume of goods produced consumed, distributed like the indices of industrial production agricultural production exports and imports etc., over a period.
3. **Value Index Numbers** – These index numbers are calculated on the basis of price and quantity both.

Methods of Calculation Price Index Numbers

It can be classified into following categories.

- (a) **Simple Index Numbers**
- (b) **Weighted Index Numbers**

- (a) **Simple Index Numbers** – It can be classified into two categories.

- (i) Simple Aggregative Method
- (ii) Simple Average of Price Relatives Method

- (i) Simple Aggregative Method – In this method, the total of current year prices of all the commodities are divided by the total of base year prices and the quotient is multiplied by 100.

Formula-

$$P_{01} = \frac{\sum p_1}{\sum p_0} \times 100$$

→ P_{01} = Index Number of the current year

→ P_{01} = Total of current years price

→ P_0 = Total of base years price

Example: 1

Calculate the index number of 2015 taking 2010 as the base year using simple aggregate method from the following information.

Commodity	Price in 2010 (₹ per unit)	Price in 2015 (₹ per unit)
A	3	6
B	5	6
C	5	6
D	3	4

Solution: 1

Commodity	Price in 2010 (p_0)	Price in 2015 (p_1)
A	3	6
B	5	6
C	5	6
D	3	4
Total	$\Sigma p_0 = 16$	$\Sigma p_1 = 22$

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100 = \frac{22}{16} \times 100$$

$$= 137.5$$

- (b) Simple Average of Price Relatives Methods** –Under this method, we convert the actual price for each variable into percentage of the base period. These percentages are called relatives because they are relative to the value for the base period.

Formula

Calculate Price Relative (R)

$$R = \frac{P_1}{P_0} \times 100$$

Calculate the Sum total of price relative's i.e. ΣR ,

$$P_{01} = \frac{\Sigma R}{N}$$

Example: 2

Calculate the price index number for 2015 taking 2010 as the base by Simple Average of price relative's method.

Commodities	A	B	C	D	E
Price (2010)	50	40	10	5	2
Price (2015)	80	60	20	10	6

Solution: 2

Commodities	P ₀ (2010)	P ₀ (2015)	$R = \frac{P_1}{P_0} \times 100$
A	50	80	160
B	40	60	150
C	10	20	200
D	5	10	200
E	2	6	300
N = 5			$\Sigma R = 1010$

$$P_{01} = \frac{\Sigma R}{N} = \frac{1010}{5} = 202$$

Price Index Number for 2015 = 202

Merits

- (a) It gives equal importance to each of the relative.
- (b) It is not influenced by unit price of commodities

Demerits

- (a) It is an unweighted method.
- (b) It is difficult to select the appropriate average.

(b) Weighted Index Numbers:-

To meet the weakness of the simple or unweighted method we assign weight to the price of each commodity by a suitable factor often taken as the quantity or the volume of the commodity sold during the base year.

It can be classified by two Methods.

- (a) Weighted Aggregative Method
- (b) Weighted Average of price Relatives Method

- (a) Weighted Aggregate Method** – Under this method, weights are assigned to different items and then weighted aggregate of the price are calculated.

The methods have been differently named after the name of their authors.

- (i) Laspeyre's Methods
- (ii) Paasche's Methods
- (iii) Fisher's Ideal Index Number

- 1. Laspeyre's Method** – This index was constructed by laspeyre in 1871. This index uses base year quantities as the weights of different items.

Formula

$$P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

p_0 = Base year prices

p_1 = Current year prices

q_0 = Base year quantity weights

- 2. Paasche's Methods** – This index was constructed by Paasche in 1874. This index used current year quantities as the weights of different items.

Formula

$$P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$$

- 3. Fisher's Ideal Index** – This index was constructed by prof. Irving fisher. Fisher's index is called ideal index number as it is the geometric mean of Laspeyres and Paaschis Index numbers and it considers both base year and current year quantities as weight.

$$P_{01} = \sqrt{\frac{\sum p_1 q_1}{\sum p_0 q_1} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

Why Fisher's Index is called on Ideal Method

→ It is based on variable or fluctuating weights.

→ The quantities of both base year and current year are used as weights.

Example: 3

From the following data calculate price index number for 2015 taking 2010 as base year by:

- (i) Laspeyres's Method
- (ii) Paasche's Method
- (iii) Fisher's Method

Commodities	Price (2010)	Quantity (2010)	Price (2015)	Quantity
A	4	20	6	30
B	5	10	4	8
C	3	8	4	6
D	2	10	2	15

Solution: 3

Commodities	p_0	q_0	p_1	q_1	$p_0 q_0$	$p_1 q_0$	$p_0 q_1$	$p_1 q_1$
A	4	20	6	30	80	120	120	180
B	5	10	4	8	50	40	40	32
C	3	8	4	6	24	32	18	24
D	2	10	2	15	20	20	30	30
					174	212	208	266

$$\Sigma p_0 q_0 = 174$$

$$\Sigma p_1 q_0 = 212$$

$$\Sigma p_0 q_1 = 208$$

$$\Sigma p_1 q_1 = 266$$

(a) Laspeyres-

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$P_{01} = \frac{212}{174} \times 100$$

$$P_{01} = 121.84$$

(b) Laspeyres –

$$P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_1} \times 100$$

$$P_{01} = \frac{266}{208} \times 100$$

$$P_{01} = 127.9$$

Fisher's –

$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

$$P_{01} = \sqrt{\frac{212}{174} \times \frac{266}{208}} \times 100$$

$$P_{01} = \sqrt{1.218 \times 1.2788} \times 100$$

$$P_{01} = \sqrt{1.5576} \times 100$$

$$P_{01} = 1.248 \times 100$$

$$P_{01} = 124.8$$

(b) Weighted Average of Price Relatives Method

Under this method weights may be determined by the proportion or percentage of expenditure on them in total expenditure during the base year.

$$R = \frac{p_1}{p_0} \times 100$$

Formula –

$$P_{01} = \frac{\sum RW}{\sum W}$$

Example: 4

Calculate the Index number by weighted average of relatives method from the following information.

Commodities	Price	Base Year Quantity	Current Year Price
A	10	10	12
B	5	4	8
C	2	10	4
D	4	20	8
E	8	10	10

Solution: 4

Commodities	p_0	q_0	p_1	$W = p_0 q_0$	$R = \frac{p_1}{p_0} \times 100$	RW
A	10	10	12	100	120	12000
B	5	4	8	20	160	3200
C	2	10	4	20	200	4000
D	4	20	8	80	200	16000
E	8	10	10	80	125	10000
						$\Sigma RW = 45200$

$$P_{01} = \frac{\Sigma RW}{\Sigma W}$$

$$P_{01} = \frac{45200}{300}$$

$$P_{01} = 150.67$$

→ **Consumer Price Index (CPI)**– It measures the average change in prices paid by the consumer for a specified list of goods and services consumed by them.

→ Method of Construction (CPI)

1. **Aggregate Expenditure Method**

2. **Family Budget Method**

1. **Aggregate Expenditure Method** – It is similar to the Laspeyre's method for construction of weighted index number.

Formula

$$\text{CPI} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

2. **Family Budget Method** – Under this method from family budget enquiry we have to estimate the aggregate expenditure of an average family on different commodities. These values are known as weights denoted by 'W'.

$$(W = p_0 q_0)$$

Price Relatives (R) are Calculated, i.e.

$$R = \frac{p_1}{p_0} \times 100$$

Formula –

$$\text{CPI} = \frac{\sum RW}{\sum W}$$

Use of CPI

1. **Measure the Real Wages-**

CPI is used in calculating the purchasing power of money and real wages.

$$(a) \text{ Purchasing Power of Money} = \frac{1}{\text{CPI}}$$

$$(b) \text{ Real wages} = \frac{\text{Money Wages}}{\text{CPI}} \times 100$$

CPI (Cost Living Index)

Example: 5

- If CPI (2002 = 100) is 425 in January, 2015 the equivalent of a rupee in January, 2015 is given by.

$$\frac{100}{425} = 0.24$$

- It means that it is worth 0.24 paise in 2002.
- If the money wages of the consumer is ₹ 5000 his real wages will be.

$$\frac{500}{425} \times 100$$

₹ 1176.47

Example: 6

Compute the Consumer price Index using Family Budget Method form the following information.

Commodities	Weight in %	Price (₹) 2012	Price (₹) 2014
Food	30	10	15
Fuel	15	12	15
Cloth	20	10	20
Rent	20	8	24
Miscellaneous	15	6	12

Solution: 6

Commodity	W	p ₀	p ₁	$R = \frac{p_1}{p_0} \times 100$	RW
Food	30	10	15	150	4500
Fuel	15	12	15	125	1875
Cloth	20	10	20	200	4000
Rent	20	8	24	300	6000
Miscellaneous	15	6	12	200	3000
	ΣW = 100				ΣRW = 19375

$$CPI = \frac{\Sigma RW}{\Sigma W} = \frac{19375}{100}$$

193.75

Index of Industrial Production

The index number industrial production measure change in the level of industrial production comprising many industries.

Formula –

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

Industrial Production

Example: 7

From the following data construct index of industrial production.

Industry	Output (units)		Weights
	2011-12	2023-24	
Mineral Production	125	190	35
Chemical Production	80	140	40
Electrical Products	170	272	10
Textile	220	308	15

Solution: 7

Industry					
	2011-12 (q_0)	2023-24 (q_1)	W	$\frac{q_1}{q_0} \times 100$	$\left(\frac{q_1}{q_0} \times 100\right)$
Mineral Production	125	190	35	152	5320
Chemical Production	80	140	40	175	7000
Electrical Products	170	272	10	160	1600
Textile	220	308	15	140	2100
			$\Sigma W = 100$		$\left[\frac{q_1}{q_0} \times 100\right] W$

= 16020

$$\begin{aligned}
 \text{Index Number of Industrial Production} &= \frac{\Sigma \left(\frac{q_1}{q_0} \times 100\right) W}{\Sigma W} \\
 &= \frac{16.20}{100} \\
 &= 160.20
 \end{aligned}$$

Uses of Index Numbers

1. **Helps in Policy Formulation-** Index numbers are indispensable tools for the management of any government or non-government organization. They are also used in planning and formulation various government and business policies.
2. **Index Number act as Economic Barometers-** It measure the pulse of an economy and act as a barometer to indicate fluctuations in general economic conditions of a country and to assess the development of economy.

3. **Help in Studying trends** – Index numbers are very useful in studying the trends of a series spread over a period of time.

For Example: If a businessman is interested in setting up a new industry then study of the trend of changes in the points, wages and income in different industries.

4. **To Measure and Compare Changes** – Index numbers help in comparative changes in two variables. It is not possible to measure change in absolute terms. The change over time and geographic location can be easily composed.

In..... of Index Numbers

1. **Provides relative changes only-** Index number are only estimates of relative changes in various events. They cannot speak the truth as they are only the approximate indicators.
2. **Lack of perfect Accuracy** - Index number ad based an sample items i.e., each and every item is not considered.
3. **Ignores qualitative changes** –while constructing the price or production index numbers no attention is paid to the changes in quality of the product.
4. **Manipulations are possible** - Index number can be constructed in such a manner so that the desired result can be obtained.

Weighted Aggregative Method (Laspeyre's Paasche's and Fisher's Ideal Method)

1. Calculate the index numbers form the following data using.
(i) Laspeyre's Method (ii) Paasche's Method (iii) Fisher's Ideal Method

Commodity	Base year		Current year	
	Price (in ₹) p_0	Quantity q_0	Price (in ₹) p_1	Quantity q_1
A	8	100	10	120
B	4	60	5	80
C	10	20	12	25
D	12	25	15	30

E	3	5	4	6
---	---	---	---	---

2. Calculate the Laspeyres's; Paasche's; Fisher's index numbers from the following data.

Commodity	Base year		Current year	
	Price (in ₹) p_0	Quantity q_0	Price (in ₹) p_1	Quantity q_1
A	10	30	12	50
B	8	15	10	25
C	6	20	6	30
D	4	10	6	20

3. The following table contains information from the raw material purchase records of a small factory for the year 2011-12 and 2023-24.

Commodity	2011-12		2023-24	
	Price (₹/Unit) p_0	Total Value q_0	Price (₹/Unit) p_1	Total Value q_1
A	5	50	6	72
B	7	84	10	80
C	10	80	12	96
D	4	20	5	30
E	8	56	8	64

Calculate Fisher's Ideal Index Number.

Weighted Average of Price Relatives Method

4. Calculate weighted average of price relative index number of prices for 2023-24 on the basis of 2011-12 from the following data:

Commodity	Quantity in 2011-12	Price (in ₹) 2011-12	Price (in ₹) 2023-24
A	20	20	35
B	12	15	18
C	8	10	11
D	4	5	5
E	6	4	5

Multiple Choice Questions

1. **Which of the following statement is correct?**
 - (a) Consumer Price Index is used to measure purchasing power of the consumer in rupees
 - (b) Index number are expressed in percentages
 - (c) Index Number measure the changes which are capable of direct measurement
 - (d) Both (a) and (b)

2. **Index number for the base period is always taken as:**
 - (a) 100
 - (b) 50
 - (c) 1
 - (d) 200

3. **Fishe's Ideal index is the:**
 - (a) Mean of Laspeyre's and Paasche's indexes
 - (b) Median of Laspeyre's and Paasche's indexes
 - (c) Geometric mean of Laspeyre's and Paasche's indexes
 - (d) None of the above

4. **The aggregate index formula using base period quantities is known as:**
 - (a) Laspeyre's index
 - (b) Fisher's Ideal index
 - (c) Bowley's index
 - (d) Paasche's index

5. **The index used to measure changes in total money value is called:**
 - (a) Price index
 - (b) Quantity index
 - (c) Value index
 - (d) None of these

6. **We use price index numbers:**
 - (a) To measure and compare prices
 - (b) To compare prices
 - (c) To measure prices
 - (d) None of these

7. The best average for construction an index numbers is:

- (a) Harmonic Mean (b) Arithmetic Mean
(c) Geometric Mean (d) None of these

8. P_1 in an index number indicated:

- (a) Price of the current year
(b) Average price of base year and current year
(c) Prince of the base year
(d) Quantity of the current year

9. Index Number by Laspeyre's Method is calculated with the help of following formula:

- (a) $P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$ (b) $P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$
(c) $P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$ (d) $P_{01} = \frac{\sum RW}{\sum W}$

10. Tick the incorrect statement

- (a) Index numbers are specialized averages
(b) Index numbers also consider qualitative changes
(c) Index number given perfect accuracy
(d) Both (b) and (c)

11. Whose method is considered ideal for the constriction of Index Numbers?

- (a) Laspeyre's Method (b) Paasche's Method
(c) Fisher's Method (d) None of these

12. Which of the following is used as weight under Paasche's method?

- (a) Quantity of the Base Year
(b) Quantity of the Current Year
(c) Average of Quantities of base year and current year
(d) None of these

13. Consumer Price Index is also known as:

- (a) Wholesale Price Index (b) Retail Price Index Numbers
 (c) Cost of Living Index (d) Both (b) and (c)

1	2	3	4	5	6	7	8	9	10	11	12	13
D	A	C	A	C	A	C	A	B	D	C	B	D

Fill in the Blanks

- Price Index Numbers measure general changes in prices between _____ and _____ year.
- Index numbers measure the effect of changes in relation to _____ or place.
- Index of industrial production is a _____ index.
- The index used to measure changes in total money value is called _____.
- Inflation is generally measured with the help of _____ index.
- Laspeyres' index number is based on the quantities of _____ year.
- Purchasing power of money is determined by _____ index.
- Index number for the base year period is _____.
- Fisher's Ideal Index is the _____ mean of the Laspeyres and Paasche

Answer Key

1	2	3	4	5	6	7	8	9
Current, Base	Time	Quantity	Value Index	Wholesale Price	Base	Consumer Price	100	Geometrical

Assertion: Reason Based

1. **Assertion (A)** : Price index number measure and Perinat Comparison of the prices of certain goods

Reason (R) :Price index numbers are more widely used

Ans: Both Assertion (A) and Reason (R) are tree and Reason (R) is not the correct explanation of Assertion (A).

2. **Assertion (A)** : Simple Aggregative price index is unweighted

Reason (R) : the relative importance of the items has not been properly reflected

Ans: Both Assertion (A) and Reason (R) are tree and Reason (R) is the correct explanation of Assertion (A).

3. **Assertion (A)** : In weighted aggregative index the items are treated as having equal importance.

Reason (R) : An index number becomes a weighted index when the relative importance of item is taken care of.

Ans: Assertion (A) is false but Reason (R) and True

4. **Assertion (A)** : Every index should have a base.

Reason (R) : Extreme values should not be selected as base period.

Ans: Both Assertion (A) and Reason (R) are True and Reason (R) is not the correct explanation of Assertion (A).

5. **Assertion (A)** : Due care should be taken in the collection of data.

Reason (R) : Data of poor reliability will give misleading results.

Ans: Assertion (A) is false but Reason (R) is true

6. **Assertion (A)** : Index numbers are extremely important in policy making.

Reason (R) : Index numbers are important in forecasting future economic activity.

Ans: Both Assertion (A) and Reason (R) are true and Reason (R) is not the correct explanation of Assertions (A).

One Word

1. Define Index Number

Ans: An index number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographical location or other characteristics.

2. State any one feature of Index number.

Ans: Index numbers are specialized average.

3. Define base period.

Ans: A base period is the reference date from which an index number is calculated.

4. State any one limitation of index numbers.

Ans: Index number provided relative changes only.

5. Mention one problem in construction index numbers

Ans: Selection of base year

6. Give two main methods to construct consumer price Index (CPI).

Ans: (i) Aggregative exp. Method

(ii) Family Budget Method

Standard Based Questions

1. **Statement 1:** Quantity Index number measure the changed in the prices of goods.

Statement 2: Quantity index number are more widely used.

Ans: Both the Statement are False

2. **Statement 1:** Consumer price index is also known as the cost of living index

Statement 2: Consumers price index measures the average change in wholesale price

Ans: Statement 1 is true and statement 2 is false

3. **Statement 1:** Wholesale price index include items pertaining to services.

Statement 2 :Wholesale price index does not have any reference.

Ans: Statement 2 is true and statement 1 is False

4. **Statement 1:** WPI is a useful guide for investors in the stock market.

Statement 2:The CPI is used to eliminate the effect of changes in price on aggregate

Ans: Both the statement are false

5. **Statement 1:** An index number take into account the quality of items

Statement 2: There is no unique index number that is acceptable to all.

Ans: Statement 2 is true and Statement 1 is false

6. **Statement 1:** The index numbers are indispensable in economic policy

Statement 2: Every formula of index number needs to be interpreted carefully.

Ans:Both the Statements are true.

EXERCISE

Answer the following questions, 1 to 10, choosing the correct answer.

1. **An index number which accounts for the relative importance of the items is known as**

- (a) Weighted Index
- (b) Simple aggregative index
- (c) Simple average of relatives

2. **In the case of the weighted index numbers the weight pertains to**

- (a) Base year
- (b) Current year
- (c) Both base and current year

3. **The impact of change in the price of a commodity with little weight in the index will be**

- (a) Small
- (b) Large
- (c) Uncertain

4. A consumer price index measures changes in

- (a) Retail prices
- (b) Wholesale prices
- (c) Producers prices

5. The item having the highest weight in consumer price index for industrial workers is

- (a) Food
- (b) Housing
- (c) Clothing

5. In general, inflation is calculated by using

- (a) Wholesale price index
- (b) Consumer price index
- (c) Producers price index

7. Why do we need an index number?

8. What are the desirable properties of the base period?

9. Why is it essential to have different CPI for different categories of consumers?

10. What does a consumer price index for industrial workers measure?

11. What is the difference between a price index and a quantity Index?

12. Is the change in any price reflected in a price Index number?

13. Can the CPI for urban non-manual employees represent the changes in the cost of living of the President of India?

14. The monthly per capita expenditure incurred by workers for an industrial centre during 1980 and 2005 on the following Items are given below. The weights of these items are 75.10. 5. 6 and 4 respectively. Prepare a weighted index number for cost of living for 2005 with 1980 as the base.

Items	Price in 1980	Price in 2005
Food	100	200
Clothing	20	25
Fuel & Lighting	15	20
House rent	30	40
Mise	35	65

15. Read the following table carefully and give your comments.

INDEX OF INDUSTRIAL PRODUCTION BASE 1993-94

Industry	Weight in %	1996-97	2003-2004
General index	100	130.8	189.0
Mining and quarrying	1073	118.2	146.9
Manufacturing	79.58	133.6	196.6
Electricity	10.69	122.0	172.6

16. Try to list to importance items of consumption in your family.

17. If the salary of a person in the base year is Rs 4,000 per annum and the current salary is Rs 6,000, by how much should his salary be raised to maintain the same standard of living if the CPI is 400?

18. The consumer price index for June, 2005 was 125. The food index was 120 and that of other items 135. What is the percentage of the total weight given to food?

19. An enquiry into the budgets of the middle class families in a certain city gave the following information:

Expenses on items	Food 35%	Fuel 10%	Clothing 20%	Rent 15%	Mise. 20%
Price (in Rs) in 2004	1500	250	750	300	400
Price (in Rs) in 1995	1400	200	500	200	250

What is the cost of living during the year 2004 as compared with 1995?

20. Record the daily expenditure, quantities bought and prices paid per unit of the daily purchases of your family for two weeks. How has the price change affected your family?

21. Given the following data-

Year	CPI of industrial workers (1982=100)	CPI of agricultural labourers (1997-87=100)	WPI (1993-94 =100)
1995-96	313	234	121.6
1996-97	342	256	127.2
1997-98	366	264	132.8
1998-99	414	293	140.7

1999-00	428	306	145.3
2000-01	444	306	155.7
2001-02	463	309	161.3
2002-03	482	319	166.8
2003-04	500	331	175.9

Source: Economic Survey, 2004-2005, Government of India.

- (i) Comment on the relative values of the index numbers.
- (ii) Are they comparable?

22. The monthly expenditure (Rs.) of a family on some important items and the Goods and Services Tax (GST) rates applicable to these items is as follows:

Items	Monthly Expense (Rs)	GST Rate%
Cereals	1500	0
Eggs	250	0
Fish. Meat	250	0
Medicines	50	5
Biogas	50	5
Transport	100	5
Butter	50	12
Babool	10	12
Tomato Ketchup	40	12
Biscuits	75	18
Cakes, Pastries	25	18
Branded Garments	100	18
Vacuum Cleaner, Car	1000	2+8

Calculate the average tax rate as far as this family is concerned.

The calculation of the average GST rate makes use of the formula for weighted average. In this case, the weights are the shares of expenditure on each category of goods. The total weight is equal to the total expenditure of the family. And the variables are the GST rates.

Category	Expenditure Weight (W)	GST Rate (x)	WX
Category 1	2000	0	0
Category 2	200	0.05	10
Category 3	100	0.12	12
Category 4	200	0.18	36
Category 5	1000	0.28	280
	3500		338

The mean GST rate as far as this family is concerned is $(338)/(3500) = 0.966$ i.e. 9.66%.