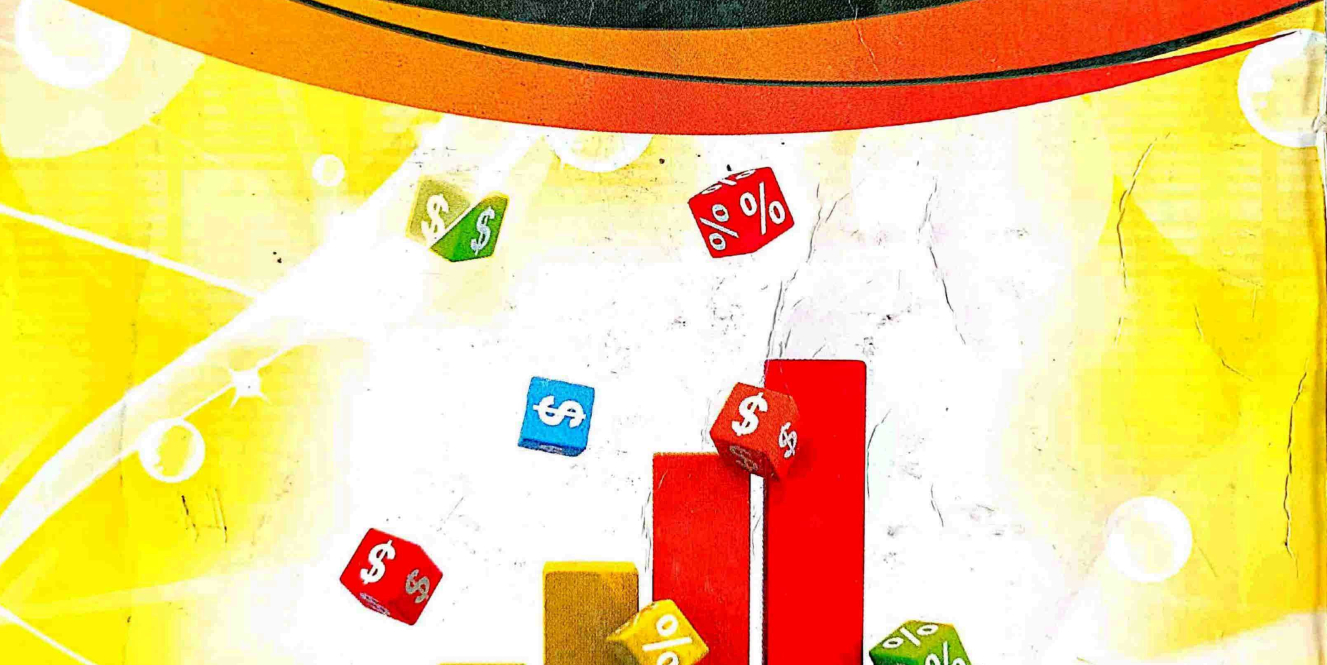
Business Statistics

T. R. Jain   S. C. Aggarwal



MBA

Contents

PART-I

1. Introduction to Statistics — 1
2. Collection of Data — 14
3. Classification of Data: Frequency Distribution — 25
4. Presentation of Data — 46
5. Measures of Central Tendency — 87
6. Measures of Dispersion — 212
7. Measures of Skewness — 282
8. Moments and Measures of Kurtosis — 307

PART-II

1. Correlation — 1
2. Linear Regression Analysis — 73
3. Index Numbers - I — 171
4. Index Numbers - II — 232
5. Time Series Analysis – I — 275
6. Time Series Analysis – II — 343
7. Probability — 367

8. Probability Distributions — Binomial and Poission 455
9. Probability Distribution — Normal 502

PART-III

1. Partial and Multiple Correlation and Regression 3
2. Sampling & Sampling Distribution 37
3. Tests of Hypothesis—Large Sample Tests 55
4. Tests of Hypothesis—Small Sample Tests 86
5. Chi-Square Test 129
6. F-Test and Analysis of Variance 168
7. Statistical Estimation Theory 199
8. Non-Parametric Tests 227
9. Statistical Decision Theory 252
10. Statistical Quality Control 276
11. Association of Attributes 305
Advanced Statistical Tables
Appendix: Descriptive Questions (VSA/SA)
University Question Papers

~~Satish~~
Kashyap

SYLLABUS

MAHARISHI DAYANAND UNIVERSITY, ROHTAK

QUANTITATIVE ANALYSIS

MBA

Semester-I

Time Allowed: 3 Hours Max. Marks: 100
[External Assessment: 80 Marks
Internal Assessment: 20 Marks]

Note: *Eight questions shall be set in the question papers. The students will be required to attempt any five questions (selecting at least one question from each unit). All questions will carry equal marks.*

Unit-I

Construction of frequency distributions and their analysis in the form of measures of central tendency and variations; types of measures, their relative merits, limitations and characteristics; skewness: meaning and co-efficient of skewness.

Unit-II

Correlation analysis–meaning & types of correlation, Karl Pearson's coefficient of correlation and Spearman's rank correlation; regression analysis–meaning and two lines of regression; relationship between correlation and regression co-efficients. Time series analysis-measurement of trend and seasonal variations; time series and forecasting.

Unit-III

Probability: Basic concepts and approaches, addition, multiplication and Bayes' theorem. Probability distributions–meaning, types and applications, Binomial, Poission and Normal distributions.

Unit-IV

Tests of significance; Hypothesis testing; Large samples, Small samples: Chi-square test, Analysis of variance.

SYLLABUS

GURU JAMBHESHWAR UNIVERSITY, HISAR

BUSINESS STATISTICS

MBA

Semester-I

Time Allowed: 3 Hours

Max. Marks: 50

Objective: *The objective of this course is to make the students learn about the applications of statistical tools and techniques in decision making.*

Course Contents:

Definition, scope and limitations of statistics; Descriptive statistics: central tendency, dispersion; Probability theory: additive and multiplicative rules, conditional probability, Baye's theorem; Probability distributions; Binomial, Poisson, Normal distribution, their characteristics and applications.

Sampling and sampling methods: Basic sampling concepts, sampling and non-sampling errors; sampling distributions of mean and proportion; law of large numbers; central limit theorem; statistical estimation: point and interval estimation of population mean.

Inferential statistics: Hypothesis testing–formulation of hypothesis and types of errors; large and small sample tests – Z, t, F tests and ANOVA (One way); non-parametric tests: Chi-square test, Sign test, Wilcoxon signed rank test, Kruskal Wallis test.

Correlation and regression analysis-two variables case; time series analysis–meaning, importance and application, trend analysis using least square method.

Statistical quality control: Causes of variations, quality control charts, purpose and logic of constructing a control chart, types of control charts, computing the control limits (X and R Charts); control charts for attributes-fraction defectives and number of defects; acceptance sampling.

Index numbers: Meaning and types, weighted aggregative indices –Laspeyer's and Paasche's indices and their comparison, test of adequacy, problems of index number.

SYLLABUS

DEENBANDHU CHHOTU RAM UNIVERSITY OF SCIENCE AND TECHNOLOGY, MURTHAL

BUSINESS STATISTICS

MBA

Semester-I

Time Allowed: 3 Hours

Max. Marks: 100

[External Assessment: 70 Marks

Internal Assessment: 30 Marks]

Course Objective: The objective of this course is to enable candidates to develop a knowledge and understanding of some basic statistical techniques and ability to apply this knowledge and understanding in solving business problems.

Course Contents:

Unit-I

Definition and role of statistics: Application of inferential statistics in managerial decision-making; Measures of centra tendency: mean, median and mode and their implications; Measures of Dispersion: range, skewness,standard deviation and mean deviation.

Unit-II

Introduction, Objectives of Time Series, Identification of Trend – Variations in Time Series: Secular Variation, Cyclical Variation, Seasonal Variation, and Irregular Variation-Methods of Estimating Trend; Index Numbers: Definition; uses; types;Simple Aggregate Method and Weighted Aggregate Method – Laspeyre's, Paashe's, Fisher's and CPI. Construction of Index Numbers and their uses.

Unit-III

Meaning of correlation, types of correlation: positive correlation, negative correlation, perfect correlation,linear and non-linear correlation; scatter diagram, Karl Pearson's coefficient of correlation, properties of correlation coefficient, probable error of correlation coefficient, meaning of multiple and partial correlations; multiple and partial correlation coefficients. Meaning of regression, types of regression:- simple and multiple regression, linear and non linear regression, statement of regression lines, definition of regression coefficients, properties of regression coefficients.

Unit-IV

Theory of Estimation, Testing of Hypothesis : Large Sample Tests, Small Sample Tests, (t,F,Z Test and Chi Square Test). Analysis of Variance – one way, two ways. Techniques of association of Attributes & Testing.

SYLLABUS

PUNJAB TECHNICAL UNIVERSITY, JALANDHAR

QUANTITATIVE TECHNIQUES

MBA

Semester-I

Max. Marks: 100

Time Allowed: 3 Hours

[External Assessment: 60 Marks
Internal Assessment: 40 Marks]

Objective: The objective of this paper is to acquaint the students with various statistical tools and techniques used to business decision making. The course aims at providing fundamental knowledge and exposure to the students to use various statistical methods in order to understand, analyse and interpret data for decision making.

Unit-I

Introduction to Statistics: Meaning, scope, importance and limitations, applications of inferential statistics in managerial decision-making. **Analysis of data**: Source of data, collection, classification, tabulation, depiction of data. **Measures of Central tendency:** Arithmetic, weighted, geometric mean, median and mode. **Measures of Dispersion:** Range, Quartile deviation, Mean deviation, Standard deviation, Co-efficient of variation, Skewness and Kurtosis.

Unit-II

Sampling and Sampling Distribution: Concept and definitions, census and sampling, probability samples and non-probability samples, relationship between sample size and errors, simple numerical only. Hypothesis Testing: Sampling theory; Formulation of Hypotheses; Application of Z-test, t-test, F-test and Chi-Square test, techniques of association of attributes & testing. Test of significance for small sample.

Unit-III

Correlation Analysis: Significance, types, Methods of correlation analysis: Scatter diagrams, Graphic method, Karl Pearson's correlation co-efficient, Rank correlation coefficient, Properties of Correlation. **Regression analysis:** Meaning, application of regression analysis, difference between correlation & regression analysis, regression equations, standard error and regression co-efficients. **Index Number:** Definition, and methods of construction, tests of consistency, base shifting, splicing and deflation, problems in construction and importance of index number.

Unit-IV

Time Series Analysis: Meaning, components and various methods of time series analysis Trend Analysis: Least Square method - Linear and Non- Linear equations, Applications in business decision-making. **Theory of Probability:** Definition, basic concepts, events and experiments, random variables, expected value, types of probability, classical approach, relative frequency and subjective approach to probability, theorems of probability, addition, Multiplication and Bays Theorem and its application. **Theoretical Distributions:** Difference between frequency and probability distributions, Binomial, Poisson and normal distribution.

Note: Relevant Case Studies should be discussed in class.

SYLLABUS

MAHARISHI MARKENDSWAR UNIVERSITY, MULLANA

QUANTITATIVE METHODS FOR DECISION MAKING

MBA

Semester-I

Course Contents

1. Arranging data to convey meaning; constructing frequency distribution.
2. Measure of central tendency and dispersion in frequency distribution.
3. History and relevance of probability theory; basic concepts in probability and probability rules.
4. Probability distributions - Normal, Binomial and Poisson; choosing the correct probability distribution.
5. Sampling and sampling distribution: random sampling, various sampling plans; relationship between sample size and standard error.
6. Testing hypothesis; basic concepts, hypothesis testing of means of samples with population standard deviation known and power of a hypothesis test.
7. Hypothesis testing of proportions and means under different conditions.
8. Chi-square test and analysis of variance.
9. Simple regression.
10. Correlation analysis.
11. Introduction to time series, variations in time series and trend analysis.
12. Applications of time series to forecasting.
13. Index number: definition, unweighted and weighted aggregate indices, quantity and value indices.
14. Statistical Package for Social Sciences (SPSS).

SYLLABUS

HIMACHAL PRADESH UNIERSITY, SHIMLA

BUSINESS STATISTICS

MBA

Semester-I

Part-B (Business Statistic Portion)

Unit-III

Frequency distribtion and their Analysis – measures of Central Tendency, Measures of dispersion.

Probability theory and Probability distributions – Binomial, Normal and Poission.

Unit-IV

Estimation – Point Estimation and Internal Estimation.

Hypothesis Testing – One Sample Test, two sample test, Z-test, t-test, Chi square test.

Simple Regression and Correlation; Estimaton using Regression line, Correlation Analysis; Introduction to Multiple and Partial Correlation.

Unit-V

Time Series: Variations in Time Series, Trend Analysis, Cyclical Variation, Seasonal Variation Index Numbers.

Unweighted Aggregate, Weighted Aggregated Index, Average of Relative methods, Quantity and Value Indices.

SYLLABUS

GAUTAM BUDH TECHNICAL UNIVERSITY, LUCKNOW
MAHAMAYA TECHNICAL UNIVERSITY, NOIDA

BUSINESS STATISTICS

MBA

Semester-I

Role of Statistics: Applications of inferential statistics in managerial decision-making; Measures of central tendency; Mean, Median and Mode and their implications; Measures of Dispersion: Range, Mean of deviation, Standard deviation, Coefficient of Variation (C.V.), Skewness, Kurtosis.

Time Series Analysis: Concepts, Additive and Multiplicative models, Components of time series, Trend analysis: Least Square method-Linear and Non-linear equations, Applications in business decision-making.

Index Numbers: Meaning, Types of index numbers, Uses of index numbers, Construction of Price, Quantity and Volume indices, Fixed base and Chain base methods.

Correlation: Meaning and types of correlation, Karl Pearson and Spearman rank correlation.

Regression: Meaning, Regression equations and their applications, Partial and Multiple correlation and regression: An overview.

Probability: Concept of probability and its uses in business decision-making; Addition and multiplication theorems; Bayes' Theorem and its applications.

Probability Theoretical Distributions: Concept and application of Binomial; Poisson and Normal distributions.

Estimation Theory and Hypothesis Testing: Sampling theory; Formulation of Hypotheses; Application of Z-test, t-test, F-test and Chi-Square test. Techniques of association of Attributes and Testing.

PART–I

Introduction to Statistics

1

INTRODUCTION

An educated citizen needs an understanding of basic statistical tools to function in a world that is becoming increasingly dependent on quantitative information. Statistics means numerical description to most people. In fact, the term statistics is generally used to mean numerical facts and figures such as per capita income, agricultural production during a year, rate of inflation and so on. However, as a subject of study, statistics refers to the body of principles and procedures developed for the collection, classification, summarization and interpretation of numerical data and for the use of such data. Without the assistance of statistical methods an organisation would find it impossible to make sense of the huge data. The purpose of statistics is to manipulate, summarize, and investigate data so that useful decision-making information results. In fact, every business manager needs a sound background of statistics. Statistics is a set of decision-making techniques which aids businessmen in drawing inferences from the available data.

ORIGIN OF STATISTICS

The term statistics has its origin in Latin word *Status*, Italian word *Statista* or German term *statistik*. All the three terms mean "Political State". In fact, the beginning of statistics was made to meet the administrative needs of the state. In ancient periods, the states were required to collect statistical data mainly for two purposes. One, concerning population so that state may come to know of the number of youngmen in the country that can be recruited in the army. Two, concerning land holdings so that state may calculate the total amount of land revenue that can be collected. It is because of this reason that statistics is called **"Science of Statecraft"** or **"Political Arithmetic"**. In modern times, statistics is not related to the administration of the state alone, but it has close relation with almost all those activities of our lives which can be expressed in **quantitative terms**.

MEANING OF STATISTICS

Broadly speaking, the term statistics has been generally used in two senses—**(1) Plural sense**, and **(2) Singular sense**. In the plural sense, the term statistics refers to numerical statements of facts relating to any field of enquiry such as data relating to production, income, expenditure, population, prices, etc. In other words, the term statistics in its plural sense refers to **numerical data or statistical data**. In its singular sense, the term statistics refers to a science in which we deal with the techniques or methods for collecting, classifying, presenting, analysing and interpreting the data. In other words, the concept in its singular sense, refers to **statistical methods**. Thus, the word 'statistics' refers either to the data themselves or to the methods dealing with numerical data.

DEFINITIONS OF STATISTICS

Different writers have defined statistics differently. These are broadly divided into two categories:

(I) Definitions in the Plural Sense.

(II) Definitions in the Singular Sense.

(I) Definitions of Statistics in the Plural Sense or as Numerical Data

In the plural sense, the term Statistics is used for numerical data. Some of its important definitions are:

"*Statistics are numerical statements of facts in any department of enquiry placed in relation to each other.*" —**Bowley**

"*By Statistics we mean quantitative data affected to a marked extent by multiplicity of causes.*" —**Yule and Kendall**

"*Statistics are numerical descriptions of quantitative aspects of things and they take the form of counts or measurements.*" —**Wallis and Roberts**

The most popular definition of statistics in terms of numerical data has been given by **Horace Secrist** which is given below:

"*By statistics we mean aggregate of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a pre-determined purpose and placed in relation to each other.*"

It is a comprehensive definition that covers all aspects of any meaningful quantitative information.

Features or Characteristics of Statistics in terms of Numerical Data

Some of the important characteristics of statistics in terms of numerical data are as follows:

(1) Aggregate of Facts: A single number does not constitute statistics. No conclusion can be drawn from it. It is only the aggregate of facts capable of offering some meaningful conclusion that constitute statistics. Likewise, the ratio of radius and circumference of a circle cannot be called statistics. For example, if it is stated that there are 1000 students in our college, then it has no statistical significance. But if it is stated that there are 300 students in arts faculty, 400 in commerce faculty and 300 in science faculty in our college, it makes statistical sense as this data conveys statistical information. Similarly, if it is stated that population of India is 91.5 crore or that the value of total exports from India is Rs. 1,06,353 crore, then these aggregate of facts will be termed as statistics. It can, therefore, be said: *'All statistics are expressed in numbers but all numbers are not statistics'.*

(2) Numerically Expressed: Statistics are expressed in terms of numbers. Qualitative aspects like 'small' or 'big'; 'rich' or 'poor'; etc. are not statistics. For instance, the fact Kapil Dev is tall and Gavaskar is short, has no statistical sense. However, if the height of Kapil Dev is 6′ 2″ and that of Gavaskar is 5′4", it would be taken as statistical information.

(3) Affected by Multiplicity of Causes: Statistics are not affected by any single factor. These are influenced by many factors simultaneously. For instance, 30 per cent rise in prices may have been due to several causes, like reduction in supply, increase in demand, shortage of power, rise in wages, rise in taxes, etc.

(4) Reasonable Accuracy: A reasonable degree of accuracy must be kept in view while collecting statistical data. This accuracy depends on the purpose of investigation, its nature, size and available resources. For example, difference of one kg. of weight in five kg. of sweetmeat is a height of inaccuracy but if against the weight of one quintal of wheat there is difference of one kg. of wheat, the inaccuracy will be treated as negligence and insignificant.

(5) Placed in Relation to each other: Such numericals alone will be called statistics as are mutually related and comparable. Unless they have the quality of comparison, they cannot be called statistics. For example, if it is stated "Ram is 40 years old, Mohan is 5 ft. tall, Sohan has 60 kg. of weight", then these numbers will not be called statistics, as they are neither mutually related nor comparable. However, if the age, height and weight of all the three are inter-related, then the same will be considered as statistics.

(6) Pre-determined Purpose: Statistics are collected with some pre-determined objective. Any information gathered without any definite purpose will only be a numerical value and not statistics. If data pertaining to the farmers of a village are being collected, there must be some pre-determined objective. It may be to know the economic status of the farmers or distribution of land among them or to know their population or for any other purpose.

(7) Enumerated or Estimated: Statistics may be collected by enumeration or these may be estimated. If the field of investigation is vast, the procedure of estimation may be helpful. For example, one lakh people attended the rally addressed by the Prime Minister in Delhi and two lakh in Bombay. These statistics are based on estimation. As against it, if the field of enquiry is limited, the enumeration method is appropriate. For example, it can be verified by enumeration whether 20 students are present in the class ~~or 10 workers~~ are working in the factory.

(8) Collected in Systematic Manner: Statistics should have been collected in a systematic manner. Before collecting them, a plan must be prepared. No conclusion can be drawn from statistics collected in haphazard manner. For instance, data regarding the marks secured by the students of a college without any reference to the class, subject, examination or maximum marks, etc., will only lead to confusion, not any meaningful conclusion.

In short, numerical data, in the absence of the above characteristics, would not be called statistics.

(II) Definitions of Statistics in Singular Sense or Statistics as a Subject

In the singular sense, Statistics means science of statistics or statistical methods. It refers to the study of the techniques relating to the collection, classification, presentation, analysis and interpretation of quantitative data. Some of the important definitions of the science of statistics are given below:

Dr. Bowley has given the following three definitions of statistics:

(1) Statistics is the Science of Counting: According to this definition, statistics is only a science of counting. But this is an incomplete definition of statistics mainly due to two reasons. (*i*) Statistics is not related only with the collection of data. It is also concerned with the presentation, tabulation, analysis and interpretation of data. This definition is incomplete since it emphasizes only one aspect of statistical methods, *i.e.*, counting. It ignores other methods like analysis and interpretation which are equally important. (*ii*) Counting is used in the collection of data, where field of enquiry is limited. But where the field is big or data are large, counting becomes impossible. Under such conditions estimates are made. For example, the production of wheat in India cannot be counted, it can only be estimated. **Bowley** himself observes, *"Great numbers are not counted, they are estimated"*.

(2) Statistics may rightly be called the Science of Averages: This is also an incomplete definition of statistics. This definition throws light only on an average which is one of the methods like dispersion, skewness, correlation, index numbers, etc. Hence, it limits the scope of science of statistics.

(3) Statistics is the science of measurement of social organism regarded as a whole in all its manifestations: This is another incomplete definition of statistics. Two reasons lead to the above conclusion (a) This definition limits the application of statistical methods to man's social activities or to social sciences only. (b) This definition mentions only one statistical method, *i.e.*, measurement and does not tell about other statistical methods. Hence, it cannot be regarded as a proper definition of statistics.

"Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data." **—Croxton and Cowden**

"Statistics is the science which deals with the collection, classification and tabulation of numerical facts as a basis for the explanation, description and comparison of phenomena." **—Lovitt**

"Statistics is the science which deals with the methods of collecting, classifying, presenting, comparing and interpretating numerical data collected to throw some light on any sphere of enquiry." **—Seligman**

▶ Features of Statistics as Science or Statistical Methods

According to all the above definitions of statistics as a science, it is clear that there are five stages of statistics which are given below:

(1) Collection of Data: Collection of relevant data concerning a problem is the first step in statistical method. Depending upon the problem under study, it is decided as to how, when and where and what kind of data are to be collected.

(2) Organisation of Data: The second step in statistical methods is to organize the collected data. With a view to rendering the collected data more comparable and simple, it is classified on the basis of time, place and quality, etc.

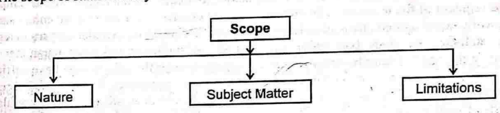
(3) Presentation of Data: To make the data intelligible, brief and attractive, it is presented in the form of tables, diagrams and graphs.

(4) Analysis of Data: The fourth step is the analysis of data. To draw conclusions, it is necessary to analyse the data. There are different methods of analysing the data, *e.g.*, measures of central tendency, measures of variation, correlation, etc.

(5) Interpretation of Data: Under this method, conclusions are drawn after analysing the data. Two or more kinds of data are compared and conclusions drawn. Even a layman may understand them. The conclusions are expressed in simple and easy language. On the basis of such conclusions future estimates are made.

SCOPE OF STATISTICS

The scope of statistics may be classified in the following parts:



(I) Nature of Statistics

The study of nature of statistics is to find out whether it is a Science or Art. As a science, statistics studies numerical data in a systematic manner and as an art, it makes use of the data to solve the problems of real life. Some scholars call it a study of Statistical Methods in preference to Statistics science, because its methods are used in all sciences.

Tippet says, *"Statistics is both a science and an art."* It is a science as its methods are basically systematic and have general applications. It is an art as its successful application depends to a considerable degree on the skill and special experience of a statistician.

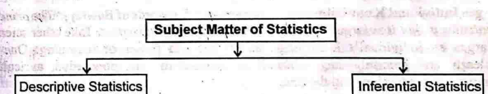
(II) Subject Matter of Statistics

In order to facilitate its study, subject matter of statistics is divided into two parts namely:

(1) Descriptive Statistics

(2) Inferential Statistics

A brief description of above branches of statistics is made as under:



(1) Descriptive Statistics: As the name suggests, the descriptive statistics merely describe the data and consists of the methods and techniques used in the collection, organisation, presentation and analysis of data in order to describe the various features and characteristics of such data. These methods can either be graphical or computational. These data can be presented in the form of chart or table in order to show trends, proportions, maximum and minimum values, etc. In addition to the organisation of data, descriptive statistics is concerned with the analysis of data so that the data can be easily understood. Measures of central tendency, dispersion, skewness and Kurtosis summarises and describes the univariate data and correlation and regression help in the establishing of the relationship in bivariate data. In descriptive statistics, nothing can be inferred from the data nor can decision be made or conclusion drawn.

(2) Inferential Statistics: It deals with methods which help in estimating the characteristics of a population or making decisions concerning a population on the basis of the sample results. Sample and population are the two relative terms. A population is treated as a universe and a sample is fraction or segment of the universe. The measured characteristics of the sample are called sample statistics, while the measured characteristics of the population are called population parameters. In inferential statistics, we study two major activities: (i) estimation of unknown parameter of a population on the basis of sample statistics and (ii) testing whether the sample data have sufficient evidence to support or reject a hypothesis about the population parameter. To be brief, inferential statistics help the decision maker to draw conclusion about the characteristics of a large population on the basis of sample results.

FUNCTIONS OF STATISTICS

Main functions of statistics are as follows:

(1) Expression of Facts in Numbers: One of the principal function of statistics is to express facts relating to different phenomena in numbers. Statement is vested with certainty when facts are expressed in numbers. For example, the statement that per capita income of India is increasing is not so precise. But if this statement is expressed in numbers as: India's per capita income which was Rs. 245 in 1950-51 rose to Rs. 9000 in 1995-96, then it becomes easy to understand and interpret with certainty.

(2) Simple Presentation: Another function of statistics is to present complex data in a simple form, so that it becomes easy to comprehend. Statistics renders complex data very simple by expressing it in terms of aggregate, average, percentage, graphs and diagrams. For example, data concerning changes in prices of main commodities between 1951 and 1995 may be so voluminous and cumbersome that it would be difficult to understand them or to draw any conclusion about them. But when presented in the form of index numbers these become simple to understand.

(3) Enlarges Individual Knowledge and Experience: In the words of **Bowley**, *"The principal function of statistics is that it enlarges individual's knowledge and expertise"*. Like other sciences, statistics enlarges an individual's knowledge, experience and power of reasoning. One can understand clearly and precisely such concepts as national income, population, agricultural production, industrial production, and the like.

(4) It Compares Facts: Another function of statistics is to compare the data relating to facts. Data have no meaning unless these are compared and inter-related. For example, if it is stated that the per capita consumption of sugar in India is 8 kg per annum, then some people may conclude that it is very low rate of consumption while the others may conclude that it is very high. But when it is compared with the consumption of sugar in other countries, say America where it is 38 kgs and Russia where it is 46 kgs, then one can draw a more meaningful conclusion.

(5) It Facilitates Policy Formulation: To facilitate formulation of policy is another function of statistics. Precise nature of each problem can be ascertained from the analysis and interpretation of data. As a result of it, some policy may be formulated. For example, it was with the help of data that **Engel** formulated a law concerning family budget. Monetary and Fiscal policies of country are formulated on the basis of relevant data.

(6) It Helps Other Sciences in Testing their Laws: Statistics also helps in testing the laws of other sciences. Many laws of economics, namely, law of demand, law of supply, Keynes' theory of employment have been verified with the help of statistics. On the contrary, classical theory of employment, quantity theory of money, etc. were not amenable to statistical verification and so were subjected to criticism.

(7) It Establishes Relationship between Facts: Statistics also establishes relationship between two or more than two facts. Tools of correlation of statistics tell if two facts have any relation between them or not and what kind of relation it has?

(8) It Helps in Forecasting: Statistics helps in forecasting changes in future with regard to a problem. For example, forecast can be made with regard to changes, in future, in food production, supply of power, growth of population, etc. as a result of five year plans in India, with the help of statistics. Extrapolation technique of statistics helps in making forecast on the basis of present facts.

(9) It Enables Realisation of Magnitude: Statistics enables realisation of magnitude of a problem. For example, from the statement that India's population has been increasing rapidly, one cannot fully realize the gravity of the situation. But, if it is stated numerically that population of India increases at the rate of 1.30 crore annually which is equal to the entire population of Australia, the magnitude of the population problem can be realized properly and easily.

(10) Presentation of Data in Condensed Form: Primary data are very much complex and haphazard. Such complex data make it impossible to draw any conclusion. Thus, it becomes imperative to present them in a concise form so that conclusions could be drawn. Statistics present complex data into a condensed form.

USES AND IMPORTANCE OF STATISTICS

In the words of **H.G. Wells** *"Statistical thinking will be one day as necessary for efficient citizenship as the ability to read and write"*. Usefulness and importance of statistics can be measured by the functions performed by it. In ancient times, statistics was used mainly as an aid to run administration. With the passage of time, utility of statistics increased manifold. Today, statistics has become an *"arithmetic of human welfare"*. In every walk of life, economic, social or political, statistics has assumed great importance. That is why **Croxton and Cowden** have expressed its importance in these words, *"Today, there is hardly a phase of human activity which does not find statistical devices at least occasionally useful"*. In the words of **Tippet**, *"Statistics affects everybody and touches life at many points"*. Usefulness of statistics becomes evident from the following:

(1) Importance for Administrators or Importance for Administration: Statistics has been in use in running the administration since ancient times. In modern times, usefulness of statistics has increased all the more due to increased welfare and other activities of the state. Every administrator has to depend on statistics for the sake of efficient administration. Statistics are the eyes of government administration. (*i*) Finance Minister makes use of data relating to revenue and expenditure while preparing budget. (*ii*) Finance Minister takes decision regarding imposition of new taxes or enhancement or curtailment of public expenditure on the basis of the data relating to production, import, export, national income, etc. (*iii*) Each government formulates its policies concerning family planning, nationalisation, establishment of new industries, twenty-point programmes, etc. on the basis of the data. (*iv*) Government makes assessment and measures efficiency of its different departments like health, education, industry, agriculture, etc. on the basis of numerical data. (*v*) Statistics prove helpful in taking decisions with regard to the defence of the country, internal law and order and crime situation, police and armed forces, etc. (*vi*) Reports of different commissions and committees appointed by the government are substantiated by the statistical data. (*vii*) All the policies and programmes chalked out by the governments of under-developed countries with the objective of economic development, economic planning, removal of income of inequalities, generation of employment opportunities and price stability are invariably based on statistical data.

(2) Importance for Businessman or Industrialist or Agriculturist or in the Field of Business, Industry or Agriculture: Knowledge of statistics is of great importance to every businessman, industrialist and agriculturist. In the words of **A.L. Boddington**, *"The successful businessman is one whose estimate most closely approaches accuracy"*. A successful businessman or agriculturist estimates demand for and supply of the commodity on the basis of relevant data. While making his estimates regarding demand and supply, he has to take into consideration data relating to seasonal changes, trade cycles, tastes of the consumers, customs, etc. It is necessary for the businessmen to know the nature and place of demand of the goods that they are dealing in and what is the future policy of the government and the possibility of changes in the price. All this can be known with the help of statistics alone. With a view to increasing his sales, a businessman conducts market survey and makes plans for advertisement on the basis of statistics. Thus, states **Ya-Lun-Chou**, *"It is not an exaggeration to say that today nearly every decision in business is made with the aid of statistical data and statistical method"*. Statistics is of great significance to every industrialist and farmer. They have to take decision about the volume of production by anticipating demand on the basis of past data and present trend. Besides, policies regarding purchase of raw materials, sale of finished products, publicity, transport, labour, mobilization of financial resources, price determination, etc. are checked out on the basis of appropriate data. All important facts discussed above are taken into account at the time of setting up of new industries.

(3) Importance in Economics: Statistics is the basis of Economics. In the words of **Ya-Lun-Chou** *"Economists depend upon statistics to measure economic aggregates such as gross national output, consumption, saving, investment, expenditure and changes in the value of money. They also use statistical methods to verify economic theory and to test hypothesis"*.

The importance of statistics in Economics is clear from the following:

(*i*) **Consumption:** How a consumer can get maximum satisfaction is determined on the basis of data pertaining to income and expenditure. Law of demand depends on data concerning price and quantity.

(*ii*) **Production:** A producer aims at earning maximum profit on the basis of data relating to cost and revenue. Efficiency of land, labour, capital and enterprise is studied with the help of statistical data. Government of a country estimates national income on the basis of data relating to production. Likewise, relative contribution of each sector to national income is also studied on the basis of statistical data.

(*iii*) **Exchange:** Price of a commodity is determined on the basis of data relating to its buyers, sellers, their demand and supply, cost and profit, etc.

(*iv*) **Distribution:** In the determination of factor-income, *viz*. rent, wage, interest and profit, statistical data play a significant role.

Study of modern economic problems like revenue, economic planning, national income, employment, inflation, etc. is very much dependent on statistical data. Indeed, economics as a science could develop so much only because of the increasing use of data.

Highlighting the significance of statistics, **Marshall** had to concede, *"Statistics are the straw out of which I, like every other economist, have to make bricks"*. It is because of the blend of Statistics that Economics is generating its new off- shoots such as Econometrics. In the words of **Tippet**, *"A day might come when the departments of economics in the universities will go out of the control of economic theoreticians and come under the control of statistical workshops, in the same manner as the department of physics and chemistry have come under the control of experimental laboratories"*. Statistics prove very useful in the formulation of economic policies. Statistical data is required to formulate and evaluate economic policies like monetary policy, fiscal policy, price control policy, export-import policy, etc. Analysis and interpretation of data enable us to know the precise nature of every problem. Indeed, in almost every field of economics, statistics plays an important role.

(4) Importance for Politicians or in political fields: Statistics has its importance for the politicians as well. In the modern democratic era, politicians play an important role in designing the economic, social and educational policies of the country. It is very essential that politicians be fully aware of the statistical data pertaining to per capita income, unemployment, import and export, black money, public debt, etc., of the country. The opposition parties, on the basis of these statistics, can make constructive criticism of the government and compel it for the revision of its programmes and policies. Policies of the party in power can give wide publicity to the achievements of their government on the basis of statistical data. A popular politician is one who admires or criticises the government on the strength of statistical data.

(5) Importance for Social Reformer or in the Social Field: Statistics is used in solving the social problems also. To a social reformer, statistics relating to such social evils as dowry, alcoholism, gambling, divorce, etc., are of great significance. With the help of the concerned data, they come to know of the gravity of these evils. They can suggest remedies against these evils only when they are equipped with relevant data. Civic problems like shortage of power and water supply can be solved if the seriousness of these problems is underlined with the help of numerical data. Whether the citizens of a country are getting adequate supply of necessaries of life like, food, cloth and shelter, etc., or not, is also known with the help of statistics. If not, then reasons for the same must be traced. It must be ascertained if the income of the people is very low or that they are spending it on liquor or other harmful drugs. One can know answer to all these questions from the data concerned. It is from the statistics relating to standard of living and level of education of the people belonging to scheduled castes and backward classes that the social reformers come to know of their difficulties. In short, no society can formulate plans and projects of its development in the absence of statistics.

(6) Importance in the Field of Science and Research: Statistics has great significance in the field of physical and natural sciences as well as research. Statistics is used both in propounding and verifying scientific laws. Thus, statistics is often used to formulate standards of body temperature, pulse rate, weight, blood pressure, etc., of an healthy person. In every science, research is conducted with the help of statistics and results of the research are also expressed in terms of statistics. Success of modern computers depends on the conclusions drawn on the basis of statistics.

(7) Importance for Banking: To every banker and banking industry, statistics relating to demand deposits, time deposits, credit, etc., are of great significance. It is on the basis of data relating to demand and time deposits that the banking system in a country determines its credit policy. Progress of banking industry in a country is measured in terms of statistics concerning total deposits, loans to primary sector, branches of banks, cash reserves, etc. Banks determine their credit policy on the basis of Theory of Probability.

(8) Importance for Insurance Companies: Insurance companies also use statistical information. These companies determine the rate of insurance premium on the basis of statistics relating to average expectancy of life in the country. Expectancy of life is calculated on the basis of Life Tables. These tables depend on the Theory of Probability. These tables show that chances of remaining alive at younger age are more and so the rate of premium is also low. Life-Expectancy reduces with the age; accordingly premium increases with the age. All this is the play of statistical methods and statistical facts and figures.

(9) Importance in the Field of Education: Progress in the field of education is measured in terms of the literacy rate of population, number of schools, colleges and universities in the country and the number of students studying therein. Shortcomings of education system are known from the data relating to examination results of the students. Data concerning male and female education, adult education, etc. is necessary to formulate any suitable education policy. Statistics regarding teacher-pupil ratio, number of students in each class, number of books issued by the library, etc. are of great significance for introducing education reforms.

(10) Importance for Economic Planning: According to **Tippet**, *"Planning is the order of the day and without statistics planning is inconceivable"*. Statistics is of prime importance in economic planning. Priorities of planning are determined on the basis of the statistics relating to resource base of the country and the short-term and long-term needs of the country. Again, success or failure of planning is measured in terms of statistical facts and figures. According to **Planning Commission**, *"Planning for the economic development of the country depends on the maximum use of statistics"*.

LIMITATIONS OF STATISTICS

In modern times, Statistics has come to occupy an important place. However, it has certain limitations. While making use of statistical methods, these limitations are kept in view. In the words of **Newshome**, *"Statistics must be regarded as an instrument of research of great value but barring severe limitations which are not possible to overcome"*. The main limitations of statistics are as follows:

(1) Study of Numerical Facts only: Statistics studies only such facts as can be expressed in numerical terms. It does not study qualitative phenomena, like honesty, friendship, wisdom, health, patriotism, justice, etc.

(2) Study of Aggregates only: Statistics studies only the aggregates of quantitative facts. It does not study any particular unit. For example, if the income of Ram is Rs. 2000 per month, it has no relevance in statistics. But if the income of Ram is Rs. 2000 p.m., that of Sohan is Rs. 3000 p.m., and that of Shyam is Rs. 4000 p.m., then these aggregates will form part of study of statistics. Their average income will work out to be Rs. 3000. This average income will lead to the conclusion that all the three persons belong to middle class. Such a conclusion would not have been possible from the study of Ram's income alone.

(3) Not the only Method: Statistical method is not the only method to study. Many a time this method does not suggest the best solution of each problem. The conclusions drawn on the basis of statistics should be verified with the help of the conclusions drawn with the help of qualitative methods.

(4) Homogeneity of Data: Quantitative data must be uniform and homogeneous. To compare the data, it is essential that whatever statistics are collected, the same must be uniform in quality. Data of different qualities and kinds cannot be compared. For example, production of foodgrains cannot be compared with the production of cloth. It is because cloth is measured in metres and foodgrains in tonnes. However, it is possible to compare their value instead of comparing the volume of their production.

(5) Results are true only on an Average: Laws of statistics are true only on an average. They express tendencies. Unlike the laws of physical science or chemistry, they are not absolutely true. They are not valid always and under all conditions. For instance, if it is said that per capita income in India is Rs. 6000 per annum, it does not mean that the income of each and every Indian is Rs. 6000 per annum. Some may have more and some may have less than it. It is true only on an average.

(6) Without Reference Results may Prove Wrong: In order to understand the conclusions very well, it is necessary that the circumstances and conditions under which these conclusions have been drawn are also studied, otherwise they may prove wrong. For example, in the business of cloth and paper, profits earned during three years may be Rs. 1000, Rs. 2000 and Rs. 3000 respectively. Thus the average profit in both the businesses comes to Rs. 2000 per annum. It may lead to the conclusion that both the businesses have similar economic position, but it is not true. If studied in proper perspective, we will find that whereas cloth-business is making progress, paper-business is on the decline.

(7) Can be used only by Experts: Statistics can be used only by those persons who have special knowledge of statistical methods. Those who are ignorant about these methods cannot make use of it. It can, therefore, be said that data in the hands of an unqualified person is like a medicine in the hands of a quack who may *abuse* it out of ignorance leading to dangerous results. In the words of **Yule and Kendall**, *"Statistical Methods are most dangerous tools in the hands of an inexpert"*.

(8) Misuse of Statistics is Possible: Misuse of statistics is possible. It may prove true what actually is not true. It is usually said, *"Statistics are like clay of which you can make a god or devil, as you please"*. Misuse of statistics is, therefore, its greatest misuse.

(9) Only Means and not a Solution: Some scholars are of the opinion that statistics are only a means in the solution of any problem. It is not a solution to the problem. To check the misuse of statistics, conclusions should be drawn impartially and without any selfish interest. Otherwise, statistics may not become a proper means for the solution of any problem.

In short, while making use of Statistics, its limitations as discussed above, must always be kept in mind.

DISTRUST OF STATISTICS

Inspite of the usefulness of statistics and the confidence of the people in its efficacy, some people have misgivings about it and they distrust it. Those who distrust statistics make the following observations about it:

(1) In the words of **Disraeli**, *"There are three kinds of lies- lies, damned lies and statistics"*.

(2) Statistics is a rainbow of lies.

(3) Statistics are tissues of falsehood.

(4) Statistics can prove anything.

(5) Statistics cannot prove anything.

(6) Statistics are like clay of which you can make god or devil, as you please.

It is evident from the above observations that statistics are nothing but bundle of lies and so are not trustworthy. The main cause of mistrust is that most of the people believe statistics readily. Thus, to take undue advantage of their *credulity*, some selfish people make misuse of the statistical data. They can present the statistics in such a distorted way as to prove right what is wrong and wrong what is right. For instance, the government claimed that in 1994-95, per capita income in India increased by about 5 percent and as such economic planning was a success. On the other hand, the opposition party claimed that is 1994-95, per capita income increased by 1.5 per cent only and as such economic planning was a failure. Statistics presented by the government as well as opposition party are correct, the only snag is that government statistics are calculated at current prices while the statistics presented by the opposition party are calculated at 1980-81 constant prices. Main causes of the mistrust of statistics are as under:

(1) Different kinds of statistics are obtained in respect of a given problem.

(2) Statistics can be altered to prove predetermined conclusions.

(3) Authentic statistics can also be presented in such a manner as to confuse the reader.

(4) When statistics are collected in a partial manner, the results are mostly wrong. Consequently, people lose faith in them.

It may be noted that if statistics are presented in a wrongful manner, the fault does not lie with the statistics. The fault lies with those people who collect wrong statistics or those who draw wrong conclusions. Statistics, as such, do not prove anything. They are simply tools in the hands of the statisticians. If the statistician misuses the data, then the blame lies squarely on the statistician and not on the data. A competent doctor can cure the malady by making good use of the medicine but the same medicine in the hands of an incompetent doctor becomes poison. The fault in this case is not of the medicine but of the unqualified doctor. In the same way, statistics are never faulty. It is pertinently said, *"Figures would not lie, but liar figure"*.

In fact, statistics should not be relied upon blindly nor distrusted outright. *"Statistics should not be used as blind man uses a lamp post for support rather than for illumination, whereas its real purpose is to serve as illumination and not as a support"*.

In making use of statistics one should be cautious and vigilant. In the words of **King**, *"The science of statistics is the most useful servant, but only of great value to those who understand its proper use"*.

In short, it is the duty of the students of economics to make use of the knowledge of statistics to seek the truth.

Remedies to Remove Distrust

Following measures may be taken to remove distrust of statistics:

(i) **Consideration of Statistical Limitations:** While making use of statistics, limitations of statistics must be taken care of, for instance, statistics should be homogeneous.

(ii) **No Bias:** Researcher should be impartial. He should make use only of proper data and draw conclusions without any bias or prejudice.

(iii) **Application by Experts:** Statistics should be used only by the experts. If they use it carefully and scientifically, the possibilities of errors will be little.

QUESTIONS

1. Define statistics and discuss its functions and limitations.
2. What is statistics? Explain the importance of statisticsin business world with suitable examples.
3. Explain the functions, importance and limitations of statistics.
4. Discuss the distrust of statistics.
5. Explain the utility of statistics as a managerial tool. Also discuss its limitations.
6. Define Statistics as a subject. Also bring out its scope.
7. Differentiate between descriptive statistics and inferential statistics.
8. "Statistics are numerical statements of facts in any department of inquiry and placed in relation to each other." Comment and discuss the characteristics of Statistics.

2 Collection of Data

INTRODUCTION

Data collection, is in fact, the most important aspect of a statistical survey. The term data as used in statistics means quantitative data. In other words, in data, we include that information which is capable of numerical expression. Qualitative aspects like intelligence, honesty, good or bad has no significance in statistics until and unless these are assigned some figures. Qualitative aspects when expressed numerically can be studied in statistics.

PRIMARY AND SECONDARY DATA

Statistical data are mainly of two types: (*i*) Primary Data, and (*ii*) Secondary Data.

(1) Primary Data: Data collected by the investigator for his own purpose, for the first time, from beginning to end, is called primary data. It is collected from the source of origin. In the words of **Wessel** *"Data originally collected in the process of investigation are known as primary data"*. Primary data are original. The concerned investigator is the first person to collect this information. The primary data are therefore, a first-hand information. To illustrate, you may be interested in studying the socio-economic status of those students studying in BBA-I class who secured first division in their XII examination. You collect information regarding their pocket allowance, their family income, educational status, their family members and the like. All this information would be termed as primary information or primary data, since you happen to be the first person to collect this information from the source of its origin.

(2) Secondary Data: In the words of **M.M. Blair** *"Secondary data are those which are already in existence, and which have been collected, for some other purpose than the answering of the question in hand"*. According to **Wessel**, *"Data collected by other persons are called secondary data"*. These data are, therefore, called second-hand data. Obviously, since these data have already been collected by somebody else, these are available in the form of published or unpublished reports. For example, data relating to Indian Railways which are annually published by the Railway Board would be secondary data for any researcher.

DISTINCTION BETWEEN PRIMARY AND SECONDARY DATA

The following are the points of difference between primary and secondary data:

(1) Difference in Originality: Primary data are original because these are collected by the investigator from the source of their origin. On the other hand, secondary data are already in existence and, therefore, are not original. Primary data are used as raw material while secondary data are finished products.

(2) Difference in the Suitability of Objectives: Primary data are always related to a specific objective of the investigator. These data, therefore, do not need any adjustment for the concerned study. On the other hand, secondary data have already been collected for some other purpose. Therefore, this data need to be adjusted to suit the objective of study in hand.

(3) Difference in Cost of Collection: Primary data are costlier in terms of time, money and efforts involved than the secondary data. This is because primary data are collected for the first time from their source of origin. Secondary data are simply collected from the published or unpublished reports. Accordingly, these are much less expensive.

Of course, it may be noted that there are no fundamental differences between primary data and secondary data. Data are data, whether primary or secondary. These are classified as primary or secondary just on the basis of their collection: first-hand or second-hand. Thus, a particular set of data when collected by the investigator for a specific purpose from the source of origin would be primary data. And the same set of data, when used by some other investigator for his own purpose would be known as secondary data. Thus, **Secrist** has rightly pointed out, *"This distinction between primary and secondary data is one of the degree. Data which are primary in the hands of one party may be secondary in the hands of other"*.

METHODS OF COLLECTING PRIMARY DATA

The primary data may be collected by using any of the following methods:

(1) Direct Personal Investigation.

(2) Indirect Oral Investigation.

(3) Information from Correspondents.

(4) Mailed Questionnaire Method.

(5) Schedules sent through Enumerators.

These methods are discussed below:

(1) Direct Personal Investigation

In this method, data are collected personally by the investigator. There is a face-to-face contact with the persons from whom the information is to be obtained. Data are collected by asking questions relating to the enquiry to the informants. Suppose, if an investigator wants to collect data about the involvement in politics of the students of Kurukshetra University, Kurukshetra, he would go to the campus and contact each student and obtain the required information.

Suitability

This method is suitable particularly when:

(1) the field of investigation is limited;

(2) a greater degree of originality of the data is required;

(3) information is to be kept secret; and

(4) investigation needs lot of expertise, care and devotion.

▶ **Merits**

(1) Originality: Data have a high degree of originality according to this method.

(2) Accuracy: Data are fairly accurate when personally collected.

(3) Reliable: Because the information is collected by the investigator himself, reliability of the data is not doubted.

(4) Other Information: When in direct contact with the informants, the investigator may obtain any other related information as well.

(5) Uniformity: There is a fair degree of uniformity in the data collected by the investigator himself from the informants. Comparison becomes easy because of uniformity of data.

(6) Flexible: This method is fairly flexible because the investigator can always make necessary adjustments in his set of questions.

▶ **Demerits**

(1) Not Proper for Wide Areas: Direct personal investigation becomes very difficult when the area of the study is very wide.

(2) Personal Bias: This method is highly prone to the personal bias of the investigator. As a result, the data may lose their credibility.

(3) Costly: This method is very expensive in terms of the time, money and efforts evolved.

(4) Wrong Conclusions: In this method, area of investigation is generally small. The results are, therefore, less representative. This may lead to wrong conclusions.

(2) Indirect Oral Investigation

In the method, the investigator obtains the information not from those persons for whom the information is needed. Information is collected orally from other persons who are expected to possess the necessary information. These other persons are known as witnesses. Indirect oral investigation is usually adopted in those cases where information through direct sources is not possible or is less reliable. For example, if a case of murder is to be investigated, it would be quite impossible to know the facts by contacting the persons directly who are involved in it. In such case, information is to be obtained from third persons such as friends, neighbours, witnesses, etc. Similarly, if a fire has broken out at a certain place, the cause of the fire may be traced by contacting persons living in the neighbourhood of that area.

▶ **Suitability**

This method is suitable particularly when:

(1) the field of investigation is large.

(2) it is not possible to have direct contact with the concerned informants.

(3) the concerned informants are not capable of giving information because of their ignorance.

(4) Enquiry committees and commissions appointed by the Government generally adopted this method.

▶ **Merits**

(1) Wider Area: This method can be applied even when the field of investigation is very wide.

(2) Less Costly: This is relatively a less costly method.

(3) Expert Opinion: Using this method an investigator can seek opinion of the experts and thereby make his information more reliable.

(4) Free from Bias: This method is relatively free from the personal bias of the investigator.

(5) Simple: This is relatively a simple method of data collection.

▶ **Demerits**

(1) Less Accurate: The data collected by this method are relatively less accurate. This is because the information is obtained from persons other than the concerned informants.

(2) Biased: There is possibility of personal bias of the witness giving information.

(3) Wrong Conclusions: This method may lead to doubtful conclusions due to ignorance and carelessness of the witness.

(3) Information from Local Sources or Correspondents

In this method, the investigator appoints local agents or correspondents in different places to collect information. These correspondents collect the information in their own way and send the same to the central office where the data are processed. Newspaper agencies generally adopt this method. This method is also adopted by various government departments where regular information is to be collected from a wide area. For example, in the construction of wholesale price indices, regular information is obtained from correspondents appointed in different areas.

▶ **Suitability**

This method is suitable particularly when:

(1) accuracy of the data is only modestly needed.

(2) regular and continuous informations are needed.

(3) the area of investigation is large.

(4) the information is to be used by journals, magazines, radio, TV, etc.

▶ **Merits**

(1) Economical: This method is quite economical in terms of time, money or efforts involved.

(2) Wider Coverage: Investigator can cover wider area.

(3) Continuity: The correspondents keep on supplying almost regular information.

(4) Suitable for Special Purpose: This method is particularly advantageous for some special purpose investigations, *e.g.*, price quotations from the different grain markets for the construction of Index Number of Agricultural Prices.

▶ **Demerits**

(1) Less Originality: In this method, there is less originality. Investigation depends more on estimation rather than actual enumeration.

(2) Lack of Uniformity: There is lack of uniformity of data. This is because data is collected by a number of correspondents.

(3) Personal Bias: This method suffers from the personal bias of the correspondents.

(4) Less accurate: The data collected by this method are not very accurate.

(5) Delay in Collection: Generally, there is delay in the collection of information through this method.

(4) Mailed Questionnaire Method

In this method, a list of questions (known as questionnaire) relating to the survey is prepared and sent to the informants by post. The questionnaire contains questions and provides space for answers. A covering letter is addressed to the informant explaining the object of survey and making a request to fill up the questionnaire and send it back within a specified time. It is also assured that the information would be kept secret. The informants write the answers against the questions and return the completed questionnaire to the investigator.

▶ **Suitability**

This method is most suited when:

(1) the area of the study is very wide and

(2) when the informants are educated.

▶ **Merits**

(1) Economical: This method is economical in terms of time, money and efforts involved.

(2) Originality: This method is original and, therefore, fairly reliable. This is because the information is supplied by the concerned persons themselves.

(3) Wider area: This method can cover wider areas.

▶ **Demerits**

(1) Lack of Interest: Generally, the informants do not take interest in questionnaires and fail to return the questionnaires. Those who return, often send incomplete answers.

(2) Lack of Flexibility: This method lacks flexibility in the sense that when questions are not properly replied, these cannot be changed to obtain the required information.

(3) Limited Use: This method has limited use in that questionnaires are answered only by the educated informants. Thus, this method cannot be used when the informants are uneducated.

(4) Biased: If the informants are biased, the informations will also be biased.

(5) Less Accuracy: The conclusions based on such investigation have only limited accuracy. This is because some questions may be difficult and accurate answers may not be possible.

(5) Schedules Filled Through Enumerators

In this method, a questionnaire is prepared as per the purpose of enquiry. The enumerator himself approaches the informant with the questionnaire. **The questionnaires which are filled by the enumerators themselves by putting questions are called schedules.** Thus, under this method, the enumerator himself fills the schedules after making enquiries from the informants. Enumerators

are those persons who help the investigators in collecting the data. The enumerators are given training to fill the schedules and put the questions intelligently in the interest of accuracy of information.

▶ **Suitability**

This method is mostly used when

(1) field of investigation is large.

(2) the investigation needs specialised and skilled investigators.

(3) the investigators are well versed in the local language and cultural norms of the informants.

▶ **Merits**

(1) Wide Coverage: This method is capable of wider coverage in terms of the area involved. Even illiterates will also provide information.

(2) Accuracy: There is a fair degree of accuracy in the results. This is because investigations are done by specialized enumerators.

(3) Personal Contact: Unlike in the case of mailing questionnaires, there is personal contact with the informants in this method. Accordingly, accurate and right answers are obtained.

(4) Impartiality: This method is impartial. This is because the enumerators themselves do not need the required information; so they are impartial to the nature of information they obtain.

(5) Complete: Schedules have the merits of completeness, because these are filled in by the enumerators themselves.

▶ **Demerits**

(1) Expensive: This is a very expensive method of investigation because of the involvement of trained investigators.

(2) Difficulties regarding Enumerators: Competent enumerators may not be available. Accuracy of the information accordingly suffers.

(3) Time Consuming: Enumerators may need specialised training for particular investigators. The process of investigation thus becomes time consuming.

(4) Not Suitable for Private Investigation: Since this method is very expensive, it is generally not suitable for private investigations. This method is generally used by the Government institutions.

(5) Inaccurate Data: If the enumerators are biased, the data will not be accurate.

ESSENTIALS/QUALITIES OF A GOOD QUESTIONNAIRE

In the context of collection of primary data, questionnaire has special significance. A questionnaire is a list of questions relating to the field of enquiry and provides space for answers. It may be defined as an instrument of collecting primary data from a large number of persons. The success of the investigation largely depends upon the proper drafting of the questionnaire. Following are some of the notable essentials or qualities of a good questionnaire:

(1) Limited number of Questions: The number of questions should be as limited as possible. Questions should be only relating to the purpose of enquiry.

(2) Simplicity: Language of the questions should be simple and clear. Questions should be short and not long or complex. Mathematical questions be avoided.

(3) Proper Order of the Questions: Questions must be placed in a proper order.

(4) No Undesirable Questions: Undesirable questions or personal questions in particular must be avoided. The questions should not offend the informants. Questions likely to offend the personal, social and religious feelings of the informants be avoided.

(5) Less Chance of Partially: Questions should be such as can be answered impartially. No controversial questions should be asked.

(6) Calculation: Questions involving calculations be avoided. Investigator himself should do the calculation job.

(7) Pre-Testing: Before giving the questionnaire a final shape, it should be subjected to pre-test. To achieve this objective, some questions be asked from the informants on trial basis. If their answers involve some difficulty, the same be changed accordingly. Such testing is technically called pilot survey.

(8) Instructions: Clear instructions for filling the questionnaire form be issued.

(9) Cross Verification: Such questions should also be asked as may help cross-verifications.

(10) Request for Return: Request should be made to return the questionnaire duly filled in. The informant be assured that the information conveyed by him will be treated as confidential.

METHODS OF COLLECTING SECONDARY DATA/SOURCES OF SECONDARY DATA

The secondary data can be collected from the following two sources:

(I) Published Sources

(II) Unpublished Sources.

(I) Published Sources

Some of the published sources of secondary data are:

(1) Government Publications: Ministries of the Central and State Governments in India publish a variety of statistics as their routine activity. As these are published by the Government, data are fairly reliable. Some of the notable Government publications on Statistics are: *Statistical Abstract of India, Annual Survey of Industries, Agricultural Statistics of India, Report on Currency and Finance, Labour Gazette, Reserve Bank of India Bulletin.*

(2) Semi-Government Publications: Semi-Government Bodies (such as Municipalities and Metropolitan Councils) publish data relating to education, health, births, and deaths. These data are also fairly reliable and useful.

(3) Reports of Committees and Commissions: Committees and Commissions appointed by the Government also furnish lot of statistical information in their reports. Finance Commission, Monopolies Commission, Planning Commission are some of the notable commissions in India which supply detailed statistical information in their reports.

(4) Publications of Trade Associations: Some of the big trade associations, through their statistical and research divisions, collect and publish data on various aspects of trading activity. For example, Sugar Mills Association published information regarding sugar mills in India.

(5) Publications of Research Institutions: Various universities and research institutions publish information regarding their research activities. In India, for example, **Indian Statistical Institute, National Council of Applied Economic Research** publish a variety of statistical data as a regular feature.

(6) Journals and Papers: Many newspapers such as Economic Times as well as Magazines such as **'Commerce', 'Facts for You'** also supply a large variety of statistical information.

(7) Publications of Research Scholars: Individual research scholars also sometimes publish their research work containing some useful statistical information.

(8) International Publications: International organisations such as U.N.O., I.M.F., I.L.O. and foreign governments, etc., also publish lot of statistical information. These are used as secondary data.

(II) Unpublished Sources

There are some unpublished sources as well. These data are collected by the government organisations and others, generally for their self use or office record. These data are not published. These unpublished numerical informations are, however, used as secondary data.

PRECAUTIONS IN THE USE OF SECONDARY DATA

We know, secondary data are collected by others to suit their specific requirements. Therefore, one needs to be very careful while using these data. **Connor** has rightly stated, *"Statistics especially other people's Statistics are full of pitfalls for the users"*. Some of the notable questions to be borne in mind while laying hands at the secondary data are:

(a) Whether the data are reliable?

(b) Whether the data are suitable for the purpose of enquiry?

(c) Whether the data are adequate?

In order to assess the reliability, suitability and adequacy of the data, following points must be kept in mind:

(1) Ability of the Collecting Organisation: One should check the ability of the organisation which initially collected the data. The data should be used only if collected by able, experienced and impartial investigators.

(2) Objective and Scope: One should note the objective of collecting data as well as the scope of investigation. Data should be used only if the objective and scope of the study undertaken earlier match with the objective and scope of the present study.

(3) Method of Collection: The method of collection of data by the original investigator should also be noted. The method adopted must match the nature of investigation.

(4) Time and Conditions of Collection: One should also make sure regarding the period of investigation as well as the conditions of investigations. For example, data collected during war times may not be suitable to generalize certain facts during peace times.

(5) Definition of the Unit: One should also make sure that the units of measurement used in the initial collection of data is the same as adopted in the present study. If the unit of measurement differs, data must be readusted before use.

(6) Accuracy: Accuracy of the data should also be checked. If the available data do not conform to the high degree of accuracy.

In short, as stated by **Bowley,** *"It is never safe to take published statistics at their face value without knowing their meaning and limitations"*. In using secondary data, the above precautions must be taken.

ROUNDING OFF DATA (OR APPROXIMATION)

Most of the statistical data are approximate figures. The process of approximation (or rounding off data) makes it simple to understand the significance of data. It enables grasp of figures easy and clear and facilitates calculation. If it is stated that Indian population is 68,38,05,051 it will be difficult to remember this figure. On the other hand, if it is said that Indian population is about 68.4 crores, it will become simple and will easily be remembered. Thereby, there will be no loss of its importance. The extent of approximation or rounding off data depends upon the degree of accuracy desired. When approximations are made, the figures should be rounded in such a way to indicate precision about facts.

Rules of Approximation (or Rounding off Data)

Before rounding off the figures, it should be decided to what extent they are to be approximated, *i.e.,* upto three, two or one decimal point or unit, 10, 100, 1,000, 1,00,000 or 1,00,00,000. The various rules of approximation are:

Rule I :

By discarding the digits entirely. According to this method, the digits, after the point to which the figures are to be approximated, are left out entirely. For example, if the figure 43,82,35,282.3759 is to be rounded up according to this method, it will be approximated like this:

upto three decimal points	43,82,35,282.375
" two " "	43,82,35,282.37
" one " "	43,82,35,282.3
" unit	43,82,35,282
" tens	43,82,35,280
" hundreds	43,82,35,200
" thousands	43,82,35,000
" ten thousands	43,82,30,000
" lakhs	43,82,00,000
" ten lakhs	43,80,00,000
" crores	43,00,00,000

Under this method of approximation, the rounded up figures will always be less than the actual figure. The lower is the digit to be removed, the lesser will be the error.

Rule II :

By raising the figure to the next higher figure. According to this method, the last digit is raised by one, eliminating the rest of the digits. For example, if the figure 43,82,35,282.3759 is to be rounded up according to this method, the rounded figure will be:

Upto three decimal points	43,82,35,282.376
" two " "	43,82,35,282.38
" one " "	43,82,35,282.4
" unit	43,82,35,283
" tens	43,82,35,290
" hundreds	43,82,35,000
" thousands	43,82,36,000
" ten thousands	43,82,40,000
" lakhs	43,83,00,000
" ten lakhs	43,90,00,000
" crores	44,00,00,000

Rule III :

Approximation to the nearest whole number. In this method, the value is unchanged when the remainder to be dropped is less than one-half. It is raised to the next higher digit, if the remainder exceeds one-half. For example, 43,82,35,282.3759 will be rounded as

upto three decimal points	43,82,35,282.376
" two " "	43,82,35,282.38
" one " "	43,82,35,282.4
" unit	43,82,35,282
" tens	43,82,35,280
" hundreds	43,82,35,300
" thousands	43,82,35,000
" ten thousands	43,82,40,000
" lakhs	43,82,00,000
" ten lakhs	43,80,00,000
" crores	44,00,00,000

This method of approximation is regarded as the best, because it is more scientific and reasonable. In such approximation, the approximated value may be more or less than the actual value. The errors being positive as well as negative, they are of compensating nature.

Approximation of percentages. Percentages can also be rounded up like:

	Ist Method	IInd Method	IIIrd Method
45.765%	45.7%	45.8%	45.8%
41.325%	41.3%	41.4%	41.3%

Precautions: The following points should be borne in mind while approximate the figures (or rounding off data):

(1) As approximation changes the origin figures, they should be rounded up only when the importance of the figures does not change.

(2) Figures should be approximated to the lowest point.

(3) The fact of approximation should be mentioned.

(4) Approximation should be done at the last stage.

QUESTIONS

1. Distinguish between primary data and secondary data.
2. Explain the various methods that are used in the collection of primary data pointing out their merits and demerits.
3. What do you mean by secondary data? Mention some of its sources. Explain briefly the precautions to be taken while making use of the secondary data.
4. Describe the different methods of collecting primary data indicating the merits and demerits of each of them.
5. "It is never to take published statistics at their face value without knowing their meaning and limitations"—*Bowley*. Elucidate.
6. What is a questionnaire? What are the essential characteristics of a good questionnaire?
7. Distinguish between 'primary data' and 'secondary data'. Are the data collected for the census in India and published in the census reports primary or secondary data?
8. Distinguish between primary and secondary data. Explain briefly the various methods of collecting primary data.
9. State the basic rules of approximating the data (or rounding off data) and their utility in statistics.
10. State the advantages of rounding off data in statistics.

Classification of Data: Frequency Distribution

3

INTRODUCTION

After the data have been collected, the next step is to present the data in some orderly and logical form so that their essential features may become explicit. The need for proper presentation of data arises because the mass of collected data in their raw form is often so volumnous, unintelligible and uninteresting that it starts at the face of the reader. The unorganised and shapeless data can neither be easily competent nor interpreted. Therefore, after the collection of data, it is imperative that data are classified and presented in such a way as to bring out points of similarities and dissimilarities in the data.

CLASSIFICATION OF DATA

Classification is the process of arranging the data into different groups or classes according to some common characteristics. In the words of **Connor,** "*Classification is the process of arranging things (either actually or notionally) in groups according to their resemblances and affinities*". According to **Spurr and Smith,** "*Classification is the grouping of related facts into classes*". The process of classification can be compared to the process of sorting out letters and packets in a post office. Just as, in sorting out operation, all collected letters and packets are separated on the basis of a common characteristic, *i.e.*, their destinations, similarly in the process of classification data are classified into various homogenous groups or classes on the basis of similarities and resemblances, *e.g.*, persons living in a certain locality may be classified on the basis of income groups to which they belong, agricultural holdings may be classified on the basis of their sizes and so on. Thus, in the process of classification data are classified into various homogenous groups or classes on the basis of similarities and resemblances.

OBJECTIVES OF CLASSIFICATION

The main objects of classification are as follows:

(1) To condense the mass of data in such a way that their similarities and dissimilarities become very clear.

(2) To facilitate comparisons, *i.e.*, to make the data comparable.

(3) To point out the most important features of the data at a glance.

(4) To present the data in a brief form.

(5) To enable statistical treatment of the data collected.

(6) To make data attractive and effective.

METHODS OF CLASSIFICATION

Broadly, the data can be classified on the following four basis in accordance with their characteristics:

(1) Geographical Classification

(2) Chronological Classification

(3) Qualitative Classification

(4) Quantitative Classification.

(1) Geographical Classification: In geographical classification, data are classified on the basis of geographical or locational differences between the various items. For example, we may present the number of firms producing bicycles stateswise as follows.

No. of Firms Producing Bicycles in 2001

State	No. of firms
Punjab	30
Haryana	20
U.P.	25

(2) Chronological Classification: When data are classified on the basis of time, it is known as chronological classification. For example, we may present population figures on the basis of time as given in the following manner:

Population of India (1951 to 1991)

Year	Population (in crores)
1951	36.1
1961	43.9
1971	54.8
1981	68.4
1991	84.4

(3) Qualitative Classification: In this type of classification, data are classified on the basis of some attribute or quality such as sex, literacy, religion, etc. This classification may be of two types: (*i*) **Simple classification:** When only one attribute is studied, *e.g.*, classification of population according to sex—male or female, this type of classification is called simple classification; (*ii*) **Manifold Classification:** When more than one attribute is studied, it is called manifold classification, *e.g.*, population may be classified as rural and urban. These may be further classified as male or female and still further as educated or uneducated.

(4) Quantitative or Numerical Classification: When data are classified on the basis of some characteristics which is capable of direct quantitative measurement such as height, weight, income, marks, etc., it is called quantitative classification. This type of classification is also called as numerical classification or grouped classification. For instance, students of a college may be classified according to weight as shown in the following table:

Weight (in lbs)	No. of students
70—80	40
80—90	50
90—100	150
100—110	250
110—120	200

In this type of classification, there are two elements: one variable (*i.e.*, weight in our example) and other frequency (*i.e.*, number of students here). **Quantitative classification is the most popular method of classifying the data.** Before we take up a detailed study of quantitative classification, it is necessary to understand the term **'Variable'**.

Variable: The characteristic, which is capable of direct quantitative measurement is called a variable or variate. Height, weight, production, consumption, marks, etc., are all variables. A variable may be either discrete or continuous.

(1) Discrete variable: A discrete variable is that one which takes only isolated or discontinuous values. There are jumps in case of a discrete variable, *e.g.*, number of goals scored in a match is a discrete variable as there can be only 1, 2, 3, etc., goals scored in a match. Intermediate values between 1 and 2 or between 2 and 3 are not possible as there cannot be 1.4 or 2.5 goals in a match. Similarly, the number of children in a family, number of students in a class, fans produced in a factory in a particular year, etc., are examples of discrete variables.

(2) Continuous variable: A continuous variable is one which can take any value in a specified interval. Temperature recorded of patients in a hospital, heights of all BBA students of Kurukshetra University, wages of all workers in a factory are examples of continuous variables.

WAYS TO CLASSIFY NUMERICAL DATA OR RAW DATA

Numerical data or raw data can be classified in any of the two ways:

(I) Ordered Array or Individual Series

(II) Frequency Distribution:

(*a*) Discrete Frequency Distribution or Discrete Series

(*b*) Continuous Frequency Distribution or Continuous Series

(I) Ordered Array or Individual Series

An ordered array or individual series is an orderly arrangement of data according to the ascending or descending order of magnitude. So, in order to prepare an array, the only thing to be done is to arrange the data or various values of variable in ascending or descending order of magnitude. An array may be useful if the data are small, but if the variable takes a large number of values, an array becomes unwieldy.

Example 1. Following data relate to the pocket expenses (rupees) of 10 students of B.Com. II class. Array them in ascending and descending order:

50, 20, 30, 15, 45, 40, 35, 25, 20, 43

Solution:

(a)	Pocket Expenses (rupees) of 10 students (In Ascending Order)		(b)	Pocket Expenses (rupees) of 10 students (In Descending Order)	
	15	35		50	30
	20	40		45	25
	20	43		43	20
	25	45		40	20
	30	50		35	15

(II) Frequency Distribution

The frequency distribution is a statistical table which shows the values of the variable arranged in order of magnitude, either individually or in groups, and also the corresponding frequencies side by side. There are two types of frequency distributions:

(*a*) **Discrete frequency distribution**

(*b*) **Grouped frequency distribution.**

(*a*) **Discrete Frequency Distribution:** It is a statistical table which shows the values of variable individually and also the corresponding frequencies side by side. The construction of discrete frequency distribution is very simple. In its construction, we count the frequencies of the various items. To find the frequency of a particular item, we make use of **tally bars**. Each tally bar indicates the presence of one value of the item. Tally bars are used in the form of **'Four and Cross Method'**. If the value of the item is repeated five times, a cross is put on four lines (~~||||~~).

Example 2. Twenty students of B.Com. II class secured the following marks in Economics out of 50 marks:

11	12	14	11	16	11	17	16	17	14
17	18	20	14	20	17	20	17	14	20

Present the data in a discrete frequency distribution.

Solution: **Formation of a Discrete Frequency Distribution**

Marks	Tally bars	Frequency
11					3
12			1
14						4
16				2
17	~~||||~~	5
18			1
20						4
	Total	20

(*b*) **Grouped Frequency Distribution:** It is a statistical table which shows the values of the variable in groups and also the corresponding frequencies side by side. An example of a grouped frequency distribution is given below:

Daily wages (Rs.)	No. of workers
40—50	7
50—60	12
60—70	8
70—80	6
80—90	2
Total	35

Useful terms associated with Grouped Frequency Distribution

For a detailed study of grouped frequency distribution, it is necessary to define and understand the following terms:

(*a*) **Class interval, or Class:** It is a group of numbers in which items are placed such as 10—20, 20—30, etc.

(*b*) **Class frequency:** The number of observations falling within a class is called its class frequency. It is denoted by 'f'.

(*c*) **Class limits:** Each class is located between two numbers. These two numbers constitute class limits. The lowest value of a class is its lower limit and the higher value is termed as upper limit. For example, in the class 10—20, the lower limit is 10 and the upper limit is 20.

(*d*) **Class mark (or mid-value):** It is the average value of the upper limit (l_2) and the lower limit (l_1). Symbolically,

$$M.V.(m) = \frac{l_1 + l_2}{2}$$

(*e*) **Width or Magnitude of the class:** The width or size or magnitude of a class is the difference between its lower and upper class limits. Symbolically,

$$i = l_2 - l_1$$

where, i = is the size of the class interval.

GENERAL RULES FOR CONSTRUCTING A GROUPED FREQUENCY DISTRIBUTION OR PROBLEMS IN THE CONSTRUCTION OF FREQUENCY DISTRIBUTION

The following rules are to be observed while forming a grouped distribution or continuous frequency distribution:

(1) Selection of Number of Classes

There are no hard and fast rules about the selection of number of classes. It depends on a number of factors such as (*i*) the number of items to be classified, (*ii*) the magnitude of the class interval, (*iii*) the accuracy desired, (*iv*) the ease of calculation for further processing of data and (*v*) size of class interval. However, the number of classes should be neither too large nor too small. It may be recommended that the number of classes should not be less than 5 or 6 and should not be greater than 15 or 20. However there is no rigidity about it.

Prof. **H.A. Sturge** have given a formula by which the number of class interval can be ascertained.

The formula is:

$$\mathbf{k = 1 + 3.322 \log N}$$

(Here, k = number of class intervals; N = Total number of observations).

Example 3. If the total number of observations is 1000, the number of class intervals can be determined by the formula:

$$\begin{aligned} k &= 1 + 3.322 \log N \\ &= 1 + 3.322 \log 1000 \\ &= 1 + 3.322 \times 3 \\ &= 1 + 9.966 \\ &= 10.966 = 11 \end{aligned}$$

Thus, the number of class intervals would be 11, after round up the fraction.

(2) Size (or Width) of Class Intervals

The choice of class interval depends on the number of classes for a given distribution and size of the data. As far as possible the class intervals should be of equal size. **Prof. Sturge** have given the following formula for determining the size of class intervals:

$$\textbf{Size of Class Interval: } i = \frac{\textbf{Largest Value} - \textbf{Smallest Value}}{\mathbf{1 + 3.322 \log N}}$$

(Here, N = Total Frequency, i = Size of Class Interval)

Example 4. If the salary of 1,000 employees in a public sector undertaking varied between Rs. 3,000 and 14,000, the size of class intervals according to **Sturge's formula** would be:

$$i = \frac{\textit{Largest Value} - \textit{Smallest Value}}{1 + 3.322 \log N}$$

$$i = \frac{\text{Rs.}14{,}000 - \text{Rs.}3{,}000}{1 + 3.322 \log 1000}$$

$$i = \frac{\text{Rs.}11{,}000}{1 + 3.322 \times 3} = \frac{\text{Rs.}11{,}000}{1 + 9.966}$$

$$i = \frac{\text{Rs.}11{,}000}{10.966} = \frac{\text{Rs.}11{,}000}{11}$$

$$i = 1000$$

Therefore, the size of class interval (i) would be Rs. 1000.

The following points must be kept in mind while making a choice of class intervals.

(*i*) As far as possible, class intervals should be such as the class limits are multiplies of 5, *e.g.*, 0—5, 5—10, 10—15, 15—20, etc. However, any number can be taken as class interval.

(*ii*) As far as possible, the class interval should be uniform through out the distribution.

(*iii*) As far as possible, every class interval should have a convenient mid point.

(3) Selection of Class Limits

Class limits should be selected in such a way that (a) the mid values of the classes coincide or come very close to the point of concentration in the data (b) the overlapping of classes is avoided (c) the class limits must be stated precisely enough so that there will be no confusion as to what they include.

(4) Kinds of Continuous Series

There is another important problem relevant to constructing a frequency distribution. These relate to the kinds of grouped or continuous series to be formed.

The following are the important kinds of continuous series:

(*a*) **Exclusive series**

(*b*) **Inclusive series**

(*c*) **Open ended series**

(*d*) **Mid-value series**

(*e*) **Cumulative frequency series.**

(*a*) **Exclusive Series:** Exclusive series is that series in which every class interval excludes items corresponding to its upper limit. In this series, the upper limit of one class interval is the lower limit of the next class interval.

For example, in a class interval of 10—15, only such items would be included the value of which is 10 or more than 10 but less than 15. Any item of the value of 15 would be included in the next class interval, *viz.*, 15—20. The following table shows the exclusive series:

Exclusive Series

Marks	Frequency
10—15	4
15—20	5
20—25	8
25—30	5
30—35	4
Total	26

It is clear from this table that the upper limit of a class interval repeats itself as the lower limit of the next class interval. Also, it may be noted that all values corresponding to, say, 15, have been incorporated not in the class interval of 10—15, but in the class interval of 15—20.

(b) Inclusive Series: An inclusive series is that series which includes all items upto its upper limits, in such series, the upper limit of class interval does not repeat itself as a lower limit of the next class interval. Thus, there is a gap between the upper limit of a class interval and the lower limit of the next class interval. The gap ranges between 0.1 to 1.0. For example, 10—14, 15—19, 20—24, etc., represent an inclusive series. Thus, all the items ranging between 10—14 are included in that class interval. Likewise, all the items ranging between 15—19 would be included in that class interval. In short, while in the exclusive series there is an overlapping of the class limits (upper class limit of one series being the lower class limit of the next class interval), there is no such overlapping in the inclusive series. Following table shows an inclusive series.

Inclusive Series

Marks	Frequency
10—14	4
15—19	5
20—24	8
25—29	5
30—34	4
Total	26

Conversion of Inclusive Series into Exclusive Series: Inclusive series is used when there is some definite difference among the values of various items in the population. In the above table if a student has obtained 14.5 or 19.5 marks, these can be expressed only if the inclusive series is converted into an exclusive series. Following steps are involved in the conversion of an inclusive series into an exclusive series:

(i) First, we find the difference between the upper limit of a class interval and the lower limit of the next class interval.

(ii) In the case of first class interval half of the difference is subtracted from the lower class and half is added to the upper class.

(iii) In case of subsequent class intervals half of that difference is added to the upper limit of a class interval and remaining half to the lower limit of the next class interval.

Using these steps, the above inclusive series can be converted into an exclusive series as follows:

Conversion of an Inclusive Series into an Exclusive Series

Marks	Frequency
9.5—14.5	4
14.5—19.5	5
19.5—24.5	8
24.5—29.5	5
29.5—34.5	4
Total	26

(c) Open Ended Series: In some series, the lower class limit of the first class interval and the upper limit of the last class interval are missing. Instead, **less than or below** is specified in place of the lower class limit of the first class interval and **more than or above** is specified in place of the upper class limit of the last class interval. Such series are called 'Open Ended Series'. Thus an open end series is that series in which lower limit of the first class interval and the upper limit of last class interval is missing. The following table shows such a series:

Open Ended Series

Marks	Frequency
Less than 5	1
5—10	3
10—15	4
15—20	6
20 and above	1

In order to determine the limit of the open end class interval, the general practice is to give same magnitude to these class intervals, as is of the other class intervals in the series. However, this practice is adopted when the known magnitudes of different class intervals in the series are equal to each other.

(d) Frequency Series containing Mid-values: Frequency series containing mid-values is that series in which we have only mid-values of the class intervals and the corresponding frequencies. For example:

Mid-value:	5	15	25	35	45
Frequency:	6	5	11	9	8

Such series may be converted into simple frequency series using the following method:

(i) First, difference between mid-values is determined; and

(ii) Second, the difference so obtained is reduced to half which when deducted from the mid-value gives lower limit of the class interval and when added to the mid-value gives the corresponding upper limit.

Thus,

$$l_1 = m - \frac{i}{2}$$

$$l_2 = m + \frac{i}{2}$$

(Where, m = mid-value; i = difference between mid-values; l_1 = lower limit and l_2 = upper limit)

In the above noted frequency distribution with mid-values, the difference between mid-values $i = 15 - 5$. Half of it is 5. Deducting 5 from each mid-value we get lower limits and adding 5 to each mid-value we get the corresponding upper limits.

The following example shows the frequency distribution with mid-values:

Preparation of a Frequency Distribution with Class Interval from Mid-values

Mid-value	Frequency	Class	Technique
5	6	0—10	$l_1 = 5 - \frac{10}{2} = 0,\ l_2 = 5 + \frac{10}{2} = 10$
15	5	10—20	$l_1 = 15 - \frac{10}{2} = 10,\ l_2 = 15 + \frac{10}{2} = 20$
25	11	20—30	$l_1 = 25 - \frac{10}{2} = 20,\ l_2 = 25 + \frac{10}{2} = 30$
35	9	30—40	$l_1 = 35 - \frac{10}{2} = 30,\ l_2 = 35 + \frac{10}{2} = 40$
45	8	40—50	$l_1 = 45 - \frac{10}{2} = 40,\ l_2 = 45 + \frac{10}{2} = 50$

(v) Cumulative Frequency Series: Cumulative frequency series is that series in which the frequencies are added corresponding to each class interval in the distribution. The frequencies then become cumulative frequencies. Cumulative frequency corresponding to the first 'class interval' would obviously be the same as the frequency itself. But for the second 'class interval' the cumulative frequency would be the sum total of the frequencies corresponding to both the second as well as first 'class intervals. The following table gives an example how a cumulative frequencies are obtained:

Cumulative Frequency Series

Marks	Frequency (No. of students)	Cumulative frequencies
5—10	3	3
10—15	8	8 + 3 = 11
15—20	9	11 + 9 = 20
20—25	4	20 + 4 = 24
25—30	4	24 + 4 = 28

Cumulative frequency series may be presented in two ways.

(*i*) Cumulative frequencies may be expressed on the basis of upper limits of the class intervals, *e.g.*, less than 10, less than 15, less than 20, when the class intervals are 5—10, 10—15 and 15—20.

(*ii*) Cumulative frequencies may be expressed on the basis of lower limits of the class intervals, *e.g.*, more than 5, more than 10, more than 15, when the class intervals are 5—10, 10—15 and 15—20.

Thus, when a frequency distribution is to be converted into a cumulative frequency distribution, the cumulative frequencies would correspond to either the lower class limits or the upper class limits of the class intervals in a distribution. Accordingly, the class intervals would get converted into 'less than' or 'more than' values. If it is of the less than type, it will represent the total frequency of all values less than and equal to the class value to which it is related. If it is a 'more than' type it will represent the total frequency of all values more than and equal to the class values to which it is related. Following is an example how a simple frequency distribution is converted into a cumulative frequency distribution.

Cumulative Frequency Series

Less than method		More than method	
Marks	No. of students	Marks	No. of students
Less than 10	3	More than 5	25 + 3 = 28
Less than 15	3 + 8 = 11	More than 10	17 + 8 = 25
Less than 20	11 + 9 = 20	More than 15	8 + 9 = 17
Less than 25	20 + 4 = 24	More than 20	4 + 4 = 8
Less than 30	24 + 4 = 28	More than 25	4

Conversion of Cumulative Frequency Series into Simple Frequency Series

Cumulative frequency series may be converted into simple frequency series. Following examples explain this process:

Example 5. Convert the following cumulative frequency distribution into a simple frequency distribution.

Marks	No. of students
Less than 10	4
Less than 20	20
Less than 30	40
Less than 40	48
Less than 50	50

Solution:

Conversion of a Cumulative Frequency Distribution into a Simple Frequency Distribution

Marks	No. of students
0—10	4
10—20	20 − 4 = 16
20—30	40 − 20 = 20
30—40	48 − 40 = 8
40—50	50 − 48 = 2

Example 6. Convert the following cumulative frequency distribution into a simple frequency distribution.

Marks	No. of students
More than 0	55
More than 5	51
More than 10	43
More than 15	28
More than 20	16
More than 25	6
More than 30	0

Solution:

Conversion of Cumulative Frequency Distribution into Simple Frequency Distribution

Marks	No. of students (f)
0—5	55 − 51 = 4
5—10	51 − 43 = 8
10—15	43 − 28 = 15
15—20	28 − 16 = 12
20—25	16 − 6 = 10
25—30	6 − 0 = 6
30—35	0

Construction of Frequency Distribution

The technique of constructing frequency distribution is illustrated as follows:

Example 7. Given below are the marks of 20 students of a class. Make a discrete frequency distribution.

10	12	18	14	13	10	12	15	17	19
18	16	14	15	17	11	20	13	12	14

Solution:

Construction of a Discrete Frequency Distribution

Marks	Tally bars	No. of students (f)
10	||	2
11	|	1
12	|||	3
13	||	2
14	|||	3
15	||	2
16	|	1
17	||	2
18	||	2
19	|	1
20	|	1
	Total	20

Example 8. Following is the record of marks obtained by 75 students in an examination. Form a frequency distribution on exclusive basis:

84	19	58	44	87	58	43	40	73	43	56	55	40	91	35
18	59	27	92	13	45	61	39	78	23	11	71	62	22	41
63	47	39	19	22	35	30	80	37	80	52	73	65	50	43
40	27	84	53	19	35	72	44	19	51	67	58	76	38	16
37	74	45	50	53	70	36	33	63	67	85	45	55	41	49

Solution: Total number of students whose marks are given as 75. Lowest marks obtained by a student is 11 and highest marks obtained is 92. Range is 92 − 11, *i.e.*, 81. So if we take class interval of 10, 9 classes will be formed. So we take the lowest class interval as 10 to 20.

Frequency Distribution of Marks Obtained by 75 Students

Marks	Tally bars	Frequency
10—20	卌 ||	7
20—30	||||	5
30—40	卌 卌 |	11
40—50	卌 卌 卌	15
50—60	卌 卌 |||	13
60—70	卌 ||	7
70—80	卌 |||	8
80—90	卌 |	6
90—100	|||	3
	Total	75

Example 9. For the following raw data prepare a frequency distribution with a class interval of 5 on inclusive basis:

Marks in English									
12	36	40	16	10	10	19	20	28	30
19	27	15	21	33	45	7	19	20	26
26	37	6	5	20	30	37	17	21	20

Solution: The lowest value is 5 and highest is 45. We take class interval of 5 on inclusive basis. The various classes will be 5—9, 10—14 and so on upto 45—49.

Formation of a Frequency Distribution

Marks	Tally bars	No. of students (f)
5—9	III	3
10—14	III	3
15—19	~~IIII~~ I	6
20—24	~~IIII~~ I	6
25—29	IIII	4
30—34	III	3
35—39	III	3
40—44	I	1
45—49	I	1
	Total	30

Example 10. The following are the marks obtained by 30 students of B.Com.II class in statistics:

Marks out of 50					
15	10	8	7	6	11
20	12	14	16	18	13
0	5	4	7	9	17
8	16	18	19	20	0
24	28	26	25	29	4

Prepare a frequency distribution by taking a class interval of 5 on exclusive basis.

Solution: The lowest value is 0 and highest is 29. We have to take a class interval of 5. The various classes will be 0—5, 5—10, and so on upto 25—30.

Frequency Distribution

Marks	Tally bars	Frequency
0—5	IIII	4
5—10	~~IIII~~ II	7
10—15	~~IIII~~	5
15—20	~~IIII~~ II	7
20—25	III	3
25—30	IIII	4
	Total	30

Example 11. The following is a record of weight of 70 students (in lbs). Tabulate the data in the form of frequency distribution taking the lowest class as (60—69).

61	73	93	107	112	76	78	69	96	72
80	88	96	109	103	84	84	106	91	75
91	92	102	91	101	90	77	105	90	86
113	101	114	72	77	118	95	63	99	82
100	106	87	89	92	107	111	76	83	86
106	107	62	94	73	108	115	85	98	93
109	97	74	98	67	82	104	88	88	92

Solution:

Preparation of Frequency Distribution

Weight in lbs	Tally bars	Frequency
60—69	~~IIII~~	5
70—79	~~IIII~~ ~~IIII~~ I	11
80—89	~~IIII~~ ~~IIII~~ IIII	14
90—99	~~IIII~~ ~~IIII~~ ~~IIII~~ III	18
100—109	~~IIII~~ ~~IIII~~ ~~IIII~~ I	16
110—119	~~IIII~~ I	6
	Total	70

IMPORTANT TYPICAL EXAMPLES

Example 12. If the class mid-points in a frequency distribution of age of a group of persons are 25, 32, 39, 46, 53 and 60, find.

(*i*) the size of the class interval

(*ii*) the class boundaries

(*iii*) the class limits, assuming that the age quoted is the age completed last birthday.

Solution: (*i*) Size of class interval = Difference between the mid-value of any two consecutive classes.

$$= 32 - 25 = 39 - 32 = \ldots = 60 - 53 = 7.$$

(*ii*) Since, the size of the class is 7 and the mid values of classes are 25, 32, 39, 46, 53 and 60, the corresponding class boundaries for different classes are obtained by using the formula:

$$l_1 = m - i/2 \qquad l_2 = m + i/2$$

where, m = mid-value, i = size of class.

The class boundaries for the first class will be:

$$l_1 = 25 - \frac{7}{2} = 25 - 3.5 = 21.5,\ l_2 = 25 + \frac{7}{2} = 25 + 3.5 = 28.5$$

The class boundaries for the 2nd class will be:

$(32 - 3.5, 32 + 3.5)$, *i.e.*, $(28.5, 35.5)$ and so on.

Thus, the various classes with class boundaries are given as

Classes	Mid-value	Technique
21.5—28.5	25	$l_1 = 25 - \frac{7}{2} = 21.5,\ l_2 = 25 + \frac{7}{2} = 28.5$
28.5—35.5	32	$l_1 = 32 - \frac{7}{2} = 28.5,\ l_2 = 32 + \frac{7}{2} = 35.5$
35.5—42.5	39	$l_1 = 39 - \frac{7}{2} = 35.5,\ l_2 = 39 + \frac{7}{2} = 42.5$
42.5—49.5	46	$l_1 = 46 - \frac{7}{2} = 42.5,\ l_2 = 46 + \frac{7}{2} = 49.5$
49.5—56.5	53	$l_1 = 53 - \frac{7}{2} = 49.5,\ l_2 = 53 + \frac{7}{2} = 56.5$
56.5—63.5	60	$l_1 = 60 - \frac{7}{2} = 56.5,\ l_2 = 60 + \frac{7}{2} = 63.5$

(*iii*) Assuming that the age quoted (X) is the age completed on last birthday, then X will be a discrete variable which takes only integral values. Hence, the given distribution can be expressed in an inclusive type of classes with class interval magnitude 7, as given in the adjoining table:

Age (on last birthday)	Mid-value
22—28	25
29—35	32
36—42	39
43—49	46
50—56	53
57—63	60

EXERCISE 3.1

1. Arrange the following data in an ascending order:
   18, 30, 15, 20, 10, 25, 19, 28

2. Following are the marks obtained by 25 students in statistics. Construct a discrete frequency distribution:

5	6	8	10	11	13	6	8	5	13	8	10	5
11	6	8	5	13	11	8	5	8	5	8	6	

3. From the following data relating to wages of 20 workers, prepare frequency distribution with a class interval of 5 on exclusive and inclusive basis:

10	15	25	27	29	20	24	23	22	12
14	16	17	18	19	18	16	15	5	9

[**Hint:** See Example 16]

4. From the following data related to the weight of college students in kg, prepare a frequency distribution with a class interval of 10 on exclusive and inclusive basis:

40	70	63	53	85
92	72	65	53	79
49	42	43	47	50
52	50	48	65	42
69	60	54	82	55

5. The weights in grams of 50 apples picked from a box are as follows:

110	103	89	75	98	121
110	108	93	128	185	123
113	92	86	70	126	78
139	120	129	119	105	120
100	116	85	99	114	189
205	111	141	136	123	90
115	128	160	78	90	107
81	137	25	84	104	100
87	115				

Construct a frequency distribution with class intervals of 15 gms on exclusive and inclusive basis.

6. Convert the following simple frequency distribution into cumulative frequency distribution by using 'less than method' and 'more than method'.

Marks	No. of students
0—10	7
10—20	10
20—30	23
30—40	30
40—50	3

7. Convert the following into exclusive form:

Class	Frequency
15—19	2
20—24	7
25—29	20
30—34	11
35—39	5

8. Prepare a frequency distribution from the following information:

7 students get less than 10 marks
18 students get less than 20 marks
38 students get less than 30 marks
63 students get less than 40 marks
70 students get less than 50 marks

9. Construct the simple frequency distribution from the following data:

Mid-values	Frequency
5	2
15	8
25	15
35	12
45	7
55	6

10. Prepare a simple frequency distribution from the following data:

Marks	No. of students
More than 0	21
More than 10	19
More than 20	14
More than 30	7
More than 40	2

MISCELLANEOUS SOLVED EXAMPLES

Example 13. From the following data related to the weight of 25 college students in kg, construct a grouped frequency distribution with class interval 40-49 and so on.

40	49	69	72	50	63	43	54	43	65	85	50	55
42	52	70	42	60	65	48	53	47	82	79	42	

Solution:

Weights (in kg)	Tally bars	No. of students
40-49	~~IIII~~ IIII	9
50-59	~~IIII~~ I	6
60-69	~~IIII~~	5
70-79	III	3
80-89	II	2
	Total	25

Example 14. Using **Sturges' rule** $k = 1 + 3.322 \log N$, prepare group frequency distribution from the marks obtained by 50 students:

4	47	84	65	15	44	13	42	60	20
12	50	92	54	17	55	15	25	70	24
25	48	17	72	80	67	20	30	72	35
45	70	22	14	70	70	25	12	75	42
65	72	34	20	72	12	30	15	80	45

Solution: Here largest item is 92 and the smallest item is 4 and number of items are 50.

As per **Sturges' rule**

$$\text{Number of classes } k = 1 + 3322 \log N = 1 + 3 \cdot 322 \log 50$$
$$= 1 + 3.322\,(1.6990)$$
$$= 1 + 5.6 = 6.6$$
$$\text{Range} = \text{Largest} - \text{Smallest}$$
$$= 92 - 4 = 88$$
$$\text{Width of the classes} = \frac{Range}{No.\ of\ Class} = \frac{88}{6.6} = 13.3$$

In multiple of 5 width of the class intervals would be 15, so the starting class would be 0–15 and so on.

Marks	Tally bars	No. of students
0—15	𝍸 𝍷	6
15—30	𝍸 𝍸 𝍷𝍷𝍷	13
30—45	𝍸 𝍷𝍷	7
45—60	𝍸 𝍷𝍷	7
60—75	𝍸 𝍸 𝍷𝍷	12
75—90	𝍷𝍷𝍷𝍷	4
90—105	𝍷	1
	Total	50

Example 15. Classify the following data by taking class interval such that their mid-values are 17, 22, 27, 32 and so on:

30	42	30	54	40	48	14	17	51	42	25	41
30	27	42	36	28	28	37	54	44	31	36	40
36	22	30	31	19	48	16	42	32	21	22	40
33	41	21	16	17	36	37	41	46	47	52	53

Solution: If the mid-value of the first class is 17 and the subsequent mid-values are 22, 27, then the first class should be 15—19 as $\frac{15+19}{2} = 17$, and the second class would be 20—24 as $\frac{20+24}{2} = 22$ as so on. Therefore, the classified data is as shown below:

Class	Tally bars	Frequency
15—19	𝍸 𝍷	6
20—24	𝍷𝍷𝍷𝍷	4
25—29	𝍷𝍷𝍷𝍷	4
30—34	𝍸 𝍷𝍷𝍷	8
35—39	𝍸 𝍷	6
40—44	𝍸 𝍸 𝍷	11
45—49	𝍸	4
50—54	𝍸	5
	Total	48

Example 16. From the following data relating to wages of 20 workers, prepare frequency distribution with a class interval of 5 on exclusive and inclusive basis:

10	15	25	27	29	20	24	23	22	12
14	16	17	18	19	18	16	15	5	9

Solution: The lowest value is 5 and highest is 29. We have to take a class interval of 5. The classes will be (*i*) 5—10, 10—15, 15—20 ... 25—30 on exclusive basis and (*ii*) 5—9, 10—14, 15—19, ... 25—29 on inclusive basis:

(*i*) Formation of a Frequency Distribution on Exclusive Basis:

Wages	Tally bars	Frequency
5—10	𝍷𝍷	2
10—15	𝍷𝍷𝍷	3
15—20	𝍸 𝍷𝍷𝍷	8
20—25	𝍷𝍷𝍷𝍷	4
25—30	𝍷𝍷𝍷	3
	Total	20

(*ii*) Formation of a Frequency Distribution on Inclusive Basis:

Wages	Tally bars	Frequency
5—9	𝍷𝍷	2
10—14	𝍷𝍷𝍷	3
15—19	𝍸 𝍷𝍷𝍷	8
20—24	𝍷𝍷𝍷𝍷	4
25—29	𝍷𝍷𝍷	3
	Total	20

QUESTIONS

1. What is meant by classification of data? What are its various objectives? Also discuss various methods of classification.
2. What is frequency distribution? What are the problems in its construction?
3. What are the general rules of framing a frequency distribution with special reference to the choice of class interval and number of classes?
4. Explain the advantages of classification of data. Discuss the different methods of classification.
5. Explain giving examples the inclusive and exclusive form of class intervals.

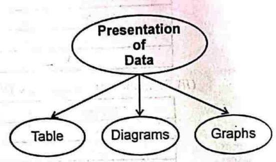
4 Presentation of Data

INTRODUCTION

The collection and classification of data lead to the problems of presentation of data. The presentation of data means exhibition of the data in such a clear and attractive manner that these are easily understood and conclusions are drawn thereof. There are many methods of presenting the data of which the following three are generally use:

(1) Tables, (2) Diagrams, and (3) Graphs.

Let us study them briefly.



TABULAR PRESENTATION OF DATA

After classification of the data, the same is presented in the form of tables. The main objective of the tables is to present the data in such a way as to attract the attentiop of an individual to important information. A table is that method of presentation of data which symmetrically organises the data rows and columns. In the words of **Neiswanger,** *"A statistical table is a systematic organisation of data in columns and rows."* Vertical disections in the formation of a table are known as column and horizontal disection (=) are known as rows. Tabulation is a process of presenting data in the form of table. According to **Prof. L.R. Connor,** *"Tabulation involves the orderely and systematic presentation of numerical data in a form designed to elucidate the problem under consideration."* In the words of **Prof. Blair,** *"Tabulation in its broadest sense is any orderly arrangement of data in columns and rows."*

OBJECTIVES OF TABULATION

(1) Simple: The principal objective of tabulation is to organise the data in such a manner that these become simple to understand.

(2) Brief: The tabulation presents a large volume of statistical data in a very brief form.

(3) Facilities Comparison: The tabulation facilitates comparision of data by presenting the data in different classes.

(4) Helpful in Presentation: Tabulation makes the data very brief. Tables are, therefore, very useful in graphic or diagrammatic presentation of the data.

(5) Helpful in Analysis: It is very easy to analyse the data from Tables. It is by organising the data in the form of table that one finds out mean and dispersion.

(6) Clarifies the Chief Characteristis of Data: The tabulation highlights the characteristics of data. Accordingly, it becomes easy to remember the statistical facts.

(7) Economy: Tabular presentation is a very economical mode of data presentation. It saves time as well as space. Important figures can be easily located in a Table.

DIFFERENCE BETWEEN CLASSIFICATION AND TABULATION

The main difference between classification and tabulation are as follows:

(1) Classification and tabulation have to be done in the sequence. First data are classified and then they are presented in tables.

(2) Classification forms the basis of tabulation.

(3) In classification, data are classified into different classes according to their similarities and dissimilarities. On the other hand, in tabulation, the classified data are placed in rows and columns. Thus, tabulation is a mechanical function of classification.

(4) Classification is a process of statistical analysis whereas tabulation is a process of presentation.

(5) Classification divides the data into classes and sub-classes, while tabulation presents the data under headings and sub-headings.

MAIN PARTS OR COMPONENTS OF A STATISTICAL TABLE

The main parts of a table are as follows:

(1) Table Number: First of all, a table must be numbered. Different tables must have different numbers, *e.g.,* 1, 2, 3..., etc. These numbers must be in the same order as the tables. Numbers facilitate location of the tables.

(2) Title of the Table: A table must have a title. Title must be written in bold letters. It should attract the attention of the readers. The title must be simple, clear and short. A good title must reveal: (*i*) the problem in hand, (*ii*) the time period of the study, (*iii*) the place of study; and (*iv*) the nature of classification of data. A good title is short but complete in all respects.

(3) Head Note: If the title of the table does not give complete information, it is supplemented with a head note. Head note completes the information in the title of the table. Thus, units of the data are generally expressed in the form of a head note below the title of the table.

(4) Stubs: Stubs are titles of the rows of a table. These titles indicate information contained in the rows of the table.

(5) Caption: Caption means title given to the columns of a table. A caption indicates information contained in the columns of the table. A caption may have sub-heads when information contained in the columns is divided in more than one class. For example, a caption of 'Students' may have boys and girls as sub-heads.

(6) Body or Field: Body of a table means sum total of the items in the table. Thus, body is the most important part of a table. It shows the whole of information contained in the table. Each item in the body is called 'cell'.

(7) Footnotes: Footnotes are given for clarification of the reader. These are generally given when information in the table is not self-explanatory.

(8) **Source:** When tables are based on secondary data, source of the data is to be given. Source of the data is specified below the footnote. It should give: name of the publication and publisher, year of publication, reference, page numbers, etc.

Format of a Table
Table Number,
Title,
Head Note,
Columns

	Stub Heads	Caption	Caption	Total
⟶	Stub-Entries	Cell	Cell	Cell
Rows	....	B O	D	Y
⟶	....			
	Total	....	....	....

Footnote:
Source:

Example 1. In 2001-2002, total production of foodgrains was 1,928 lakh tons of which production of rice, wheat and other crops was, 860, 708 and 360 lakh tons, respectively. Percentage share of rice, wheat and other crops in the total production of foodgrains was 44.60, 36.72 and 18.68 respectively. Present this information in the form of a Table indicating its various parts.

Solution: **Table 1: Title: Production of Foodgrains in India in 2001-2002**

	S. No.	Foodgrains	Production		
			Total Quantity	Percentage of Total Output	
⟶	1.	Rice	860	44.60	⟵
Stub Head	2.	Wheat	708	36.72	Column
Stubs	3.	Other crops	360	18.68	Head
		Total	1,928	100.00	⟵ Body

Footnote: In 'others', all remaining foodgrains are included.
Source: *Economic Survey, 2002.*

GENERAL RULES FOR THE CONSTRUCTION OF A TABLE OR ESSENTIALS OF A GOOD TABLE

Construction of a good table depends upon the objective of study. It also depends upon the intelligence of the statistician. There are no hard and fast rules for the construction of a table. However, certain guidelines may be noted in this context:

(1) Title: Title of a table must be provided at the top centre of the table.

(2) Special Emphasis: Some items in the table may need special emphasis. Such items should be placed in the head rows (top above) or head columns (extreme left). Moreover, such items should be presented in bold letters.

(3) Size of Table: To determine an ideal size of a Table, a rough draft or sketch must be drawn. Rough draft will give an idea as to how many rows and columns should be drawn for presentation of the data.

(4) Headings: Headings should generally be written in the singular form. For example, in the columns indicating goods, the word 'good' should be used.

(5) Abbreviations should not be Used: Use of abbreviations should be avoided in the headings or sub-headings of the table. Short forms of the words such as Govt., M.P. (monetary policy), etc. should not be used. Also such signs as "(ditto)" should not be used in the body of the table.

(6) Footnote: Footnote should be given only if needed. However, if footnote is to be given, it must bear some asterisk mark (*) corresponding to the concerned item.

(7) Units: Units used must be specified above the columns. If figures are very large, units may be noted in the short form as '000' hectare or '000' tons.

(8) Total: In the table, sub-totals of the items must be given at the end of each row. Grand total of the items must also be noted.

(9) Percentage and Ratio: Percentage figures should be provided in the table, if possible. This makes the data more meaningful.

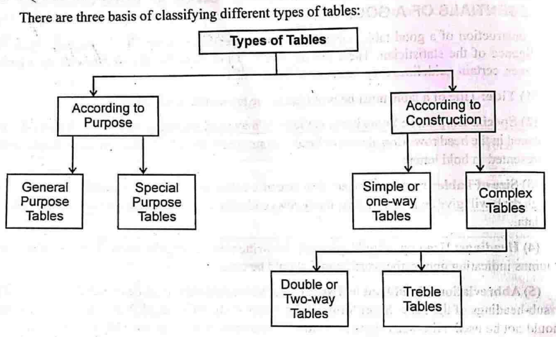
(10) Place of Approximation: If some approximate figures have been used in this table, the extent of approximation must be noted. This may be indicated at the top of the table as a part of head note or at the foot of the table as a footnote.

(11) Source of Data: Source of data must be noted at the foot of the table. It is generally noted next to the footnote.

(12) Simplicity, Economy and Attractiveness: A table must be simple, attractive and economical in space.

◘ TYPES OF TABLES

There are three basis of classifying different types of tables:



(1) According to Purpose

According to purpose, there are two kinds of tables:

(*i*) **General Purpose Table:** General purpose table is that table which is of general use. It does not serve any specific purpose. Such tables are just 'data bank' for the use of researchers for their various studies. These tables are generally attached to some official reports, *e.g.*, Census Reports of India.

(*ii*) **Special Purpose Table:** Special purpose table is that table which is prepared with some specific purpose. Generally these are small tables and limited to the problem in hand. In these tables, data are presented in the form of result of the analysis.

(2) According to Construction

According to construction, also, tables are of two kinds:

(*i*) **Simple Table:** A simple table is that which shows only one characteristics of the data. Table 2, for example, is simple table. It shows number of students in a college:

Table 2: No. of Students in a College

Class	No. of students
BBA (I)	200
B.A. (I)	100
B.A. (II)	80
B.A. (III)	60
Total	440

(*ii*) **Complex Table:** A complex or multiple table is one which shows more than one characteristics of the data. On the basis of the characteristics shown, these tables may be further classified as:

(*a*) **Two-way Table:** A two-way table is that which shows two characteristics of the data. For example, table 3, showing the number of students in different classes according to their sex is a two-way table:

Table 3: No. of Students in a College
(According to Sex and Class)

Class	No. of students		Total
	Boys	Girls	
BBA (I)	160	40	200
B.A. (I)	40	60	100
B.A. (II)	60	20	80
B.A. (III)	50	10	60
Total	310	130	440

(*b*) **Three-way Table:** A three-way table is that which shows three characteristics of the data. For example, table 4 shows number of students in a college according to class, sex and habitation.

Table 4: No. of Students in a College
(According to Class, Sex and Habitation)

Class	Boys			Girls			Total		
	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
BBA (I)	10	10	20	5	5	10	15	15	30
B.Com (I)	10	30	40	15	45	60	25	75	100
B.Com (II)	15	45	60	5	15	20	20	60	80
B.Com (II)	10	40	50	5	5	10	15	45	60
Total	45	125	170	30	70	100	75	195	270

Similarly, a manifold table showing more than three characteristics can be constructed. Such a type of table gives information about so many related questions. These tables are also called Higher Order Tables.

Example 2. Prepare a blank table to show the percentage of rural and urban population in India in 1971, 1981, 1991 and 2001.

Solution: **Percentage Distribution of Rural and Urban Population of India**

Census Year	Percentage of population		Total
	Rural	Urban	
1971			
1981			
1991			
2001			

Example 3. Point out the mistakes in the following table and rearrange it in the form of a good table.

Literate	Less than 20	20-30	30-40	40 and above
Male				
Female				

Solution: The following mistakes may be noted in the above table:

1. There is no table number and no title of the table.
2. The table is without any head note.
3. Caption and sub-entries have not been properly noted. Literacy and age be divided according to caption and sub-entries respectively.
4. Total of the rows and columns has not been provided.

Removing these mistakes, the Table may be presented in the following form:

Distribution of Population
(According to Age, Sex and Literacy)

Age (Years)	Boys			Girls		
	Male	Female	Total	Male	Female	Total
Less than 20 years						
20-30 years						
30-40 years						
40 and above						
Total						

Example 4. In 1995-96, total production of foodgrains was 1,720 lakh tons of which production of rice, wheat and other crops was, 795, 625 and 300 lakh tons, respectively. Percentage share of rice, wheat and other crops in the total production of foodgrains was 46.22, 36.34 and 17.44 respectively. Present this information in the form of a table.

Solution: **Production of Foodgrains in India in 1995-96** (lakh tonnes)

S.No.	Foodgrains	Production	
		Total Quantity	Percentage of Total Output
1.	Rice	795	46.22
2.	Wheat	625	36.34
3.	Others	300	17.44
	Total	1,720	100.00

Footnote: In 'others', all remaining foodgrains are included.
Source: *Economic Survey, 1997, P. S-16.*

Example 5. In a sample study about the coffee habits in two town, following data were observed:

Town X: 52% persons were males
25% were coffee drinkers, and
16% were male coffee drinkers

Town Y: 55% persons were males
28% were coffee drinkers, and
18% were male coffee drinkers.

Represent the above data in a tabular form.

Solution: **Percentage of Coffee Drinkers in Town X and Town Y** (in percentage)

Attribute	Town X			Town Y		
	Males	Females	Total	Males	Females	Total
Coffee Drinkers	16	9	25	18	10	28
Non-coffee Drinkers	36	39	75	37	35	72
Total	52	48	100	55	45	100

Example 6. In a trip organised by a college there were 80 persons each of whom paid Rs. 15.50 on an average. There were 60 students each of whom paid Rs. 16. Members of the teaching staff were charged at a higher rate. The number of servants was 6 (all males) and they were not charged anything. The number of ladies was 20% of the total of which one was a lady staff member.

Solution: No. of ladies $= \frac{20 \times 80}{100} = 16$

Trip Goers and their Contribution

Trip-Goers	Sex			Contribution Per Member	Total Contribution
	Male	Femlae	Total		
Students	45	15	60	Rs. 16	Rs. 960
Staff	13	1	14	Rs. 20	Rs. 280
Servants	6	Nil	6	Nil	
Total	64	16	80	Rs. 15.50	Rs. 1,240

EXERCISE 4.1

1. In 1998, the contribution of agriculture, to India's National Income, Industry and services was 48%, 21% and 31% respectively. In 1999 these shares were 32%, 27% and 41% respectively. This information is based on Economic Survey of 2000-2001. Present this information in the form of a table.

2. Point out the mistakes in the following table. Re-arrange it in the form of a good table.

No. of Students	Subject			
	Economics	English	Hindi	History
Boys:				
Girls:				

3. Following information relates to the marks secured by 50 boys and girls in their paper in Economics. Present the information in the form of a two way table.

Marks	0—10	10—20	20—30	30—40
Boys:	10	7	6	1
Girls:	5	5	12	4

4. Draw a blank table to show the distribution of population according to sex, age and literacy.
5. Prepare a blank table in which can be shown the prices per kilograms of rice and wheat for the years 1994 and 1995 for three important markets in Punjab.
6. Point out the mistakes made in the following table and re-arrange the following table with a view to make it more intelligible.

Sex	Hindus		Muslims		Sikhs	
	Literate	Illiterate	Literate	Illiterate	Literate	Illiterate
Male:						
Female:						

7. Prepare a blank table showing the particulars of students studying in B.Com classes of Himachal Pradesh University College according to sex and class.
8. Present the following information in a suitable tabular form:
   (i) In 1985, out of total 2000 workers in a factory, 1550 were members of a trade union. The number of women workers employees was 250, out of which 200 did not belong to any trade union.
   (ii) In 1990, the number of union worker was 1725 of which 1600 were men. The number of non-union workers was 380, among which 155 were women.

DIAGRAMMATIC PRESENTATION OF DATA

Data may be presented in a simple and attractive manner in the form of diagrams. Diagrammatic presentation is the technique of presenting data in the form of Bar diagrams, Rectangles, Pie-diagrams, Pictographs and Cartographs.

Utility of Diagrammatic Representation

(1) Make Data Simple and Understandable: The most complex statistical data is made simple with the help of diagrams. One can understand the features of data merely by having a look at the diagrams.

(2) Remembrance for Long Period: In the form of diagrams, data are easily remembered for a long period. These are not easily forgotten.

(3) No Need of Training or Special Knowledge: One needs no training or special knowledge in reading the diagrams. Diagrams are easily understood even by a layman.

(4) Attractive and Effective Means of Presentation: Diagrams are very attractive and effective means of presenting data. It is rightly said that, a picture is worth of a thousand words.

(5) Saving of Time and Labour: Diagrammatic presentation of data saves lot of time and labour compared to other techniques of data presentation.

(6) Facilities Comparison: Diagrams facilitate comparison of data. Thus, data on investment in Private and Public sectors, when presented in the form of diagrams, can be easily compared. One can easily note the difference between the two.

(7) Informative and Entertaining: Besides being informative, diagrammatic presentation is an entertaining means of data presentation.

(8) Helpful in Predictions: Diagrammatic presentation helps in predicting the behaviour of variables.

(9) Helpful in Transmission of Information: Diagrams are very helpful in transmission of statistical information.

General rules for Constructing Diagrams

Some of the general rules for constructing diagrams are as follows:

(1) Attractive and Effective: Diagrams must be attractive and effective in communicating the information.

(2) Proper Size: Diagrams must suit the size of the paper. It should be neither too big nor too small.

(3) Proper Heading: Diagrams must bear proper headings. A heading must be simple, short and informative.

(4) Proper Scale: Before making a diagram, its scale should be properly determined and the same be mentioned on the diagram.

(5) Use of Signs and Colour: Diagrams must carry some signs on the nature and classification of information. Colours may be used to indicate different aspects of a diagram. These signs and colours must be clarified.

(6) Less Use of Words or Numerical: In diagrammatic presentation of data one should make use of minimum number of words and numericals.

(7) Drawing the Border: Diagrams must be bordered with bold lines to make them attractive.

(8) Simple: Simplicity is the principal feature of a diagram. It should not be ignored.

(9) From Left to Right or Bottom to Top: The construction of diagrams should flow from left to right or from bottom to the top.

(10) Statistics: One should indicate the statistics or data used in the construction of diagrams.

Limitations of Diagrammatic Presentation

The following are the main limitations of diagrammatic presentation of data:

(1) Estimate: Diagrammatic presentation of data shows only an estimate of the actual behaviour of the variables. In other words, diagrams show only an aggregate behaviour of the variables.

(2) Limited Use: Only a limited set of data can be presented in the form of a diagram. In fact, diagrams are generally used only when comparisons are involved.

(3) More Time: Diagrammatic presentation of data is a time consuming process. It involves too much verification of the data.

(4) Misuse: Diagrams may be misused for false projection of the statistical facts, especially in case of advertisement.

(5) Analysis: It is not very easy to arrive at final conclusions after seeing the diagrams. Generally, a diagram offers preliminary conclusions.

TYPES OF DIAGRAMS

Several types of diagrams are used to present statistic data. The following main types are discussed below:

(1) Bar Diagrams

(2) Rectangular Diagrams

(3) Pie Diagrams

(4) Pictograms and Cartograms.

(1) Bar Diagrams

Bar diagrams are the most common type of diagrams used in practice. In these diagrams, only the length of the bars are taken into account. The width of the bar is adjusted in accordance with the space available and the number of the bars to be used. The gap between one bar and another is also kept constant. In the construction of bar diagrams, either vertical or horizontal bars are used. But vertically bars are generally preferred.

Types of Bar Diagrams

Bar diagrams are of the following types:

(i) Simple bar diagrams

(ii) Multiple bar diagrams

(iii) Sub-divided bar diagrams

(iv) Percentage bar diagrams

(v) Deviation bar diagrams.

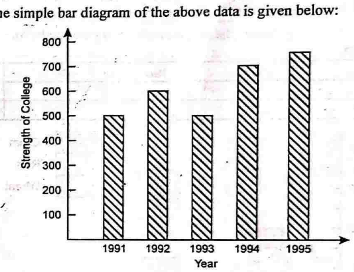
(i) Simple bar diagrams: A simple bar diagram is used to represent the values of a single variable with respect to time or geographical location etc. For example, the data of sales, production, population, etc., for various years may be shown by means of simple bar diagrams. Different values of the variable are shown by bars of proportionate lengths. The bars may be shaded or coloured to make the diagram more attractive.

Example 7. The strength of a college from 1991 to 1995 are given below:

Year:	1991	1992	1993	1994	1995
Strength of College:	500	600	500	700	750

Represent the data by a simple bar diagram.

Solution: The simple bar diagram of the above data is given below:

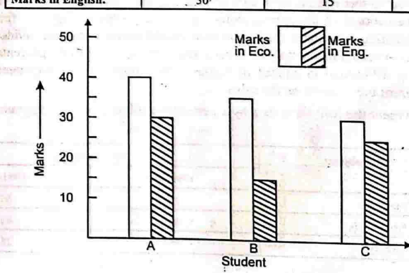


(ii) Multiple bar diagrams: The multiple bar diagrams are used to show two or more related variables with respect to time and location. In other words, they represent one or more than one type of data at a time. For every set of values, the separate bars are drawn but they are kept adjacent to each other. These bars are shaded or coloured differently to distinguish between variables.

Example 8. Draw a multiple bar diagram to show the following data:

Student:	A	B	C
Marks in Economics:	40	35	30
Marks in English:	30	15	25

Solution:

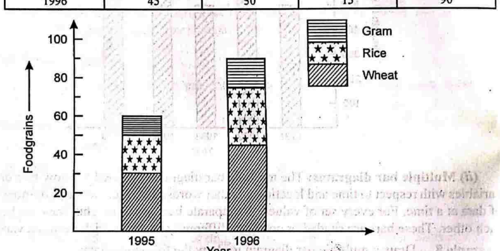


(iii) **Sub-divided bar diagrams (or Component bar diagrams):** Sub-divided bar diagrams are used to present the constituent (or component) parts of the aggregates with respect to time or location. In drawing such diagrams, first bars are drawn for the total sum of the values of different variables and then these bars are divided to show the various components of the variables. Different parts of the bar must be distinguished by difference in shades or colours. The order of the components of the various aggregates on different bars should be the same.

Example 9. Present the following data on the production of foodgrains in the form of a sub-divided bar diagram:

Year	Wheat	Rice	Gram	Total
1995	30	20	10	60
1996	45	30	15	90

Solution:



(iv) **Sub-divided bar diagrams drawn on percentage basis:** The sub-divided bar diagrams drawn on percentage basis are those diagrams which are used to present parts of the value of the aggregate with respect to some characteristic on percentage basis. Here, the values of the components are expressed on percentage of the corresponding aggregate. In these diagrams, first the bars of equal lengths representing aggregates (100%) are drawn and then these bars are divided to show the various components in accordance with their percentage. Different parts of the percentage bars must be distinguished by difference in shades or colours. The order of the components of various aggregates on different bars should be the same.

Example 10. Represent the following data by a percentage sub-divided bar diagram:

(in '000)

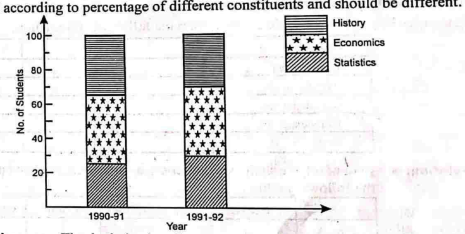
Subjects	No. of sudents	
	1990-91	1991-92
Statistics	25	30
Economics	40	42
History	35	28

Solution: First we prepare a percentage table.

Percentage Table

Subject	1990-91		1991-92	
	No. of students ('000)	%	No. of students ('000)	%
Statistics	25	25	30	30
Economics	40	40	42	42
History	35	35	28	28
Total	100	100	100	100

Now, we draw bars of equal length (100%) corresponding to each year. Then each bar is divided according to percentage of different constituents and should be different.

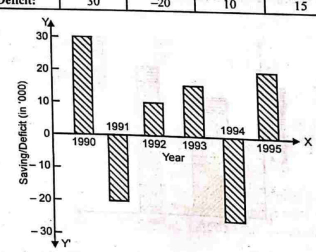


(v) **Deviation bar diagrams:** The deviation bar diagrams are used to compare the net deviation of related variables with respect to time and location. Bars representing positive and negative deviations are drawn above and below the base line. Such type of diagrams represent the deviations in magnitude as well as in direction.

Example 11. Represent the following data by a deviation bar diagrams:

Years:	1990	1991	1992	1993	1994	1995
Saving/Deficit:	30	–20	10	15	–25	20

Solution:



(2) Rectangular Diagrams

In case of bar diagrams, only length of the bar is taken into account as the comparison is one dimensional. On the other hand, in case of rectangular diagrams, both the lengths and breadth are taken into account and data are represented by the area. Here, comparison is two dimensionals. These diagrams may be constructed in either of the two ways: (*i*) By representing the figures as they are given, and (*ii*) By converting the figures into percentages and then sub-dividing the length into various components.

The following examples would illustrate both of the methods of constructing rectangular diagrams:

▶ Sub-divided Rectangular Diagrams

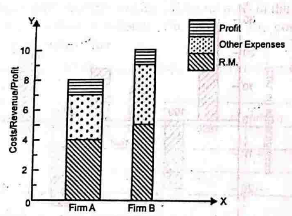
Example 12. Draw a suitable diagram from the following information:

	Firm A	Firm B
Price per unit	Rs. 8	Rs. 10
Quantity sold	1,000	600
Raw materials	4,000	3,000
Other expenses	3,000	2,400
Profit	1,000	600

Solution: First of all, we shall calculate cost (raw materials, etc.) and profit per unit as given in the following table:

	Raw materials	Other expenses	Profit	Selling price	Quantity sold
Firm A	4	3	1	8	1000
Firm B	5	4	1	10	600

An appropriate diagram for representing the data would be the rectangular diagram whose widths are in the ratio of quantities sold, *i.e.*, 1,000 : 600, *i.e.*, 10 : 6. Selling price would be represented by the corresponding heights (or lengths) of the rectangles with various cost (raw materials, others, etc.) and profit represented by various divisions of the rectangules as shown in the following diagram:



Example 13. Draw a suitable diagram to represent the following information:

	Selling price per unit (in Rs.)	Quantity sold	Total Cost (in Rs.)			
			Wages	Materials	Misc.	Total
Factory X	400	20	3,200	2,400	1,600	7,200
Factory Y	600	30	6,000	6,000	9,000	21,000

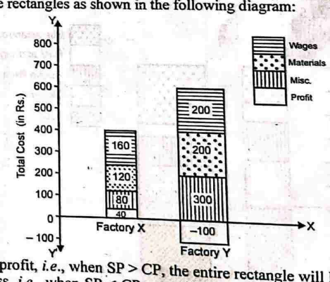
Show also the profit or loss as the case may be.

Solution: First of all we shall calculate the cost (wages, materials, misc.) and profit per unit as given in the following table:

Total Cost	Factory X (20 units)		Factory Y (30 units)	
	Total (Rs.)	Per Unit (Rs.)	Total (Rs.)	Per Unit (Rs.)
Wages	3200	160	6000	200
Materials	2400	120	6000	200
Misc.	1600	80	9000	300
Profit/Loss	800 (8,000 – 7,200)	40	–3000 (18,000 – 21,000)	–100

Note: Negative profit is regarded as loss.

An appropriate diagram for representing this data would be the rectangles whose widths are in the ratio of the quantities sold, *i.e.*, 20 : 30, *i.e.*, 2 : 3. Selling prices would be represented by the corresponding heights of the rectangles with various costs (wages, materials, misc.) and profit or loss represented by the various divisions of the rectangles as shown in the following diagram:



Note: In case of profit, *i.e.*, when SP > CP, the entire rectangle will lie above the X-axis. But in case of loss, *i.e.*, when SP < CP, we will have rectangle with a portion lying below the X-axis which will reflect the loss incurred, *i.e.*, can not recovered through sales.

▶ Percentage Sub-divided Rectangular Diagrams

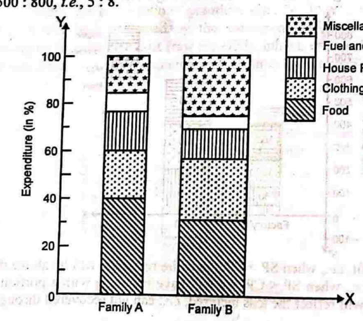
Example 14. Represent the following data through a suitable diagram:

Items of Expenditure	Family A (Income Rs. 500)	Family B (Income Rs. 800)
Food	200	250
Clothing	100	200
House Rent	80	100
Fuel and Lighting	40	50
Miscellaneous	80	200
Total	Rs. 500	Rs. 800

Solution: Since the total incomes of the two families are different, an appropriate diagram for the above data will be rectangular diagram on percentage basis. We first find the percentage of expenditure on each items for each family.

Item	Family A (Income Rs. 500)			Family B (Income Rs. 800)		
	Expenditure (Rs.)	%	Cumulative %	Expenditure (Rs.)	%	Cumulative %
Food	200	40	40	250	31.2	31.2
Clothing	100	20	60	200	25.0	56.2
House Rent	80	16	76	100	12.5	68.6
Fuel and Lighting	40	8	84	50	6.3	75.0
Miscellaneous	80	16	100	200	25.0	100
Total	500	100		800	100	

The width of the rectangles will be taken in the ratio of total income of families, *i.e.*, 500 : 800, *i.e.*, 5 : 8.



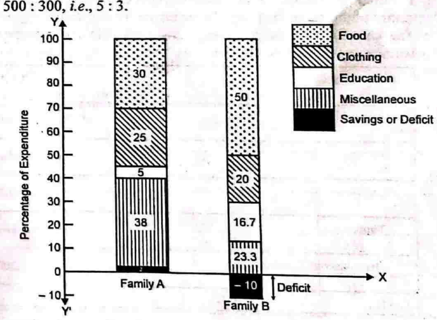
Example 15. Present the following information through a suitable diagram:

Items of Expenditure	Family A (Income Rs. 500)	Family B (Income Rs. 300)
Food	150	150
Clothing	125	60
Education	25	50
Miscellaneous	190	70
Savings or Deficit	+10	–30

Solution: Since the total incomes of the two families are different, an appropriate diagram for the above data will be rectangular diagram on percentage basis. We first find the percentage of expenditure on each item for each family.

Items of Expenditure	Family A			Family B		
	Expenditure (Rs.)	%	Cumulative (%)	Expenditure (Rs.)	%	Cumulative (%)
Food	150	30	30	150	50	50
Clothing	125	25	55	60	20	70
Education	25	5	60	50	16.7	86.7
Miscellaneous	190	38	98	70	23.3	110.0
Savings or Deficit	+10	2	100	–30	–10	100
Total	500			300		

The width of the rectangles will be taken in the ratio of the total income of families, *i.e.*, 500 : 300, *i.e.*, 5 : 3.



Note: In case total expenditure = total income, *i.e.*, when TE = TI, the entire rectangle will lie above the X-axis. But in case of deficit, *i.e.*, when TE > TI, we will have rectangles with a portion lying below the X-axis which will reflect the deficit.

(3) Pie Diagrams

A pie diagram is a circle sub-divided into component sectors. Just as a sub-divided bar or rectangle represents the whole data sub-divided into various components, similarly, a sub-divided circle represents the whole data sub-divided in to various component parts. For example, a pie diagram may show the distribution of money spent by the government on various heads of expenditure. The whole circle represents the total expenditure and various sectors of the circle show the percentage of total expenditure spent on various heads.

The following example would illustrate the construction of a pie diagram.

Example 16. Present the following data in the form of a pie diagram relating to cost of construction of a house in a city:

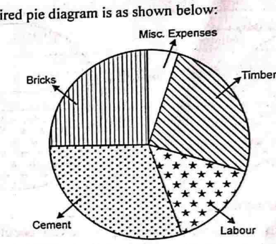
Items	Cost (Rs. '000)
Bricks	50
Cement	60
Labour	30
Timber	50
Miscellaneous Expenses	10
Total	200

Solution: First draw a circle of any radius. The circle, however, should neither be too big nor too small. The whole circle represents 100% of the total expenditure. The angle at the centre of the circle is of 360°. We calculate the degrees of angles of different items. The following table gives the degrees of angles of the various items representing the various components of the data:

Items	Cost (Rs. '000)	Degree of the angle
Bricks	50	$360° \times \frac{50}{200} = 90°$
Cement	60	$360° \times \frac{60}{200} = 108°$
Labour	30	$360° \times \frac{30}{200} = 54°$
Timber	50	$360° \times \frac{50}{200} = 90°$
Miscellaneous Expenses	10	$360° \times \frac{10}{200} = 18°$
Total	200	360°

Now, divide the circle into five sectors according to the degrees of angles at the centre, as calculated in the above table. It is preferably to arrange the angles in the descending order and place the longest at the top.

The required pie diagram is as shown below:



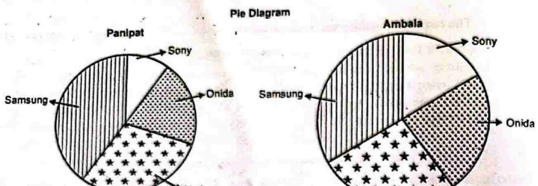
Example 17. Following are the data about the market share of 4 brands of T.V. sets sold in Panipat and Ambala. Present the data by a pie diagram.

Brands of sets	Units sold in Panipat	Units sold in Ambala
Samsung	480	625
Akai	360	500
Onida	240	438
Sony	120	312

Solution: Total sets sold in Panipat and Ambala are 1,200 and 1,875 respectively. Data are to be represented by two circles whose radii are in the ratio of square roots of total T.V. sets sold in each city in the ratio of $\sqrt{1,200} : \sqrt{1,875}$ or 1 : 1.18. The calculations regarding construction of pie diagram are as follows:

Brands of sets	Panipat		Ambala	
	Sets sold	Degree	Sets sold	Degree
Samsung	480	$\frac{480}{1,200} \times 360° = 144°$	625	$\frac{625}{1,875} \times 360° = 120°$
Akai	360	$\frac{360}{1,200} \times 360° = 108°$	500	$\frac{500}{1,875} \times 360° = 96°$
Onida	240	$\frac{240}{1,200} \times 360° = 72°$	438	$\frac{438}{1,875} \times 360° = 84°$
Sony	120	$\frac{120}{1,200} \times 360° = 36°$	312	$\frac{312}{1,875} \times 360° = 60°$
Total	1,200	360°	1,875	360°
Square root of total	36.64		43.30	

Pie Diagram



(4) Pictograms and Cartograms

(*i*) **Pictograms:** Pictograms also known as pictographs, are very popularly used in presenting statistical data. Pictographs show data in the form of pictures. For examples, data on scooters would be represented by pictures of scooters, data on milk diary by picture of cows, data on population would be represented by pictures of men, data on milk would be represented by the bottles of milk and the like.

(*ii*) **Cartograms:** Cartograms also known as cartographs are used to give quantitative information on a geographical basis. Cartographs show data in the form of maps. For example, rainfall in different parts of the country, size of population in different regions; sugar mills located in different areas and the like. Can be represented by maps.

EXERCISE 4.2

1. Present the following data diagrammatically by using simple bar diagram:

Years:	1990	1991	1992	1993
No. of students:	400	500	700	400

2. Represent the following data by multiple bar diagram:

Years:	1991-92	1992-93	1993-94	1994-95
Imports (Rs. crore):	600	700	800	900
Exports (Rs. crore):	500	600	700	800

3. Represent the following data by sub-divided bar diagram:

Production of Electricity from Different Sources in India

('000 million Kwts)

Year	Hydro-Electricity	Thermal Electricity	Total Production
1992-93	46	64	110
1993-94	49	72	121
1994-95	48	82	130
1995-96	51	89	140

4. Represent the following data by a percentage sub-divided bar diagram:

Faculty	Number of students	
	1990–1991	2000–2001
Arts	300	720
Commerce	480	800
Science	420	480

5. Represent the following data diagrammatically:

Particulars	Firm 'A'	Firm 'B'	Firm 'C'
	Rs.	Rs.	Rs.
Wood	4	4	5
Labour	3	3.5	4
Polishing	2	2.5	2
Cost	9	10	11
Selling Price	10	10	10
Profit or Loss	+ 1	0	− 1

[**Hint:** See Example 34]

6. Represent the following data diagrammetically:

Items of Expenditure	Family A (Income Rs. 600)	Family B (Income Rs. 1000)
Food	200	400
Rent	100	200
Clothing	150	200
Fuel	100	100
Miscellaneous	50	50

[**Hint:** Saving of B is Rs. 50]

7. Represent the following data by a rectangular diagram:

	Commodities	
	A	B
Price per unit of commodity (Rs.)	10	12
Quantity sold	20	24
Cost of raw materials used (Rs.)	100	120
Other costs (Rs.)	60	96
Profit (Rs.)	40	72

8. Represent the following data by means of a pie diagram:

	Cost (Rs.)
Cost of Labour	10
Cost of Material	25
Cost of Electricity	5
Cost of Transportation	15
Overhead	35
Total	90

GRAPHIC PRESENTATION OF DATA

Graphic presentation is another method of data presentation. In this method, statistical data is presented on the graph paper. Graphic presentation is the technique of presenting the data in the form of curves or lines on the graph paper.

Utility of Graphic Presentation

The main advantage of graphic presentation of data are as follows:

(1) Presentation of Time Series and Frequency Distribution: Graphic Presentation is a very effective technique of data presentation in case of time series data and frequency distributions.

(2) Location of Averages: Using graphic technique, we can easily locate the value of certain averages, such as mode and median. It is not possible with the help of diagrams.

(3) Easy Estimation: Graphic presentation facilities interpolation and extrapolation of the data in a more convenient and precise manner. For example, given the population data for the years 1951 and 1971, one can easily make an estimate of the population in the years 1981 or 1991.

(4) Study of Correlation: Graphic technique helps in studying correlation between different variables, such as price and demand, cost and output, and the like.

(5) Comparison of Multiple-Dam: Data of different dimensions can be easily compared with the help of graphic presentation.

In short, **Habbard** rightly observed, wherever keeping of record of the data, drawing of conclusions, describing of facts are necessary, these graphs provide such an important means whose power we have started experiencing and using.

Limitations of Graphic Presentation

The following are some of the important limitations of graphic presentation of statistical data:

(1) Less Significant: Graphs are not of equal significance to all the people. It is generally difficult for a layman to interpret graphs. Accordingly these are of little value to them.

(2) Only a Measure of Tendency: Graphs show only tendency of the data. Actual values are not always clear from the graphs.

(3) Lack of Precise Value: Since graphs are based on brief information, these do not show precise values.

(4) Wrong Conclusions: Graphs may sometimes suggest wrong conclusions. In fact even a small change in the scale of the graph causes a lot of difference in the structure of the graph. This may lead to wrong concusions.

General Rules for Constructing a Graph

The following points must be kept in mind while constructing a graph:

(1) Heading: Every graph must have a suitable and precise heading. Heading must be self explanatory about the nature of information in the graph.

(2) Choice of Scale: One should fix an appropriate scale on which data should be presented. An appropriate scale is that scale by which the entire data is easily represented by the graph. The graph should be on the middle of the graph paper to make it attractive.

(3) Proportion of Axis: As far as possible, length of X-axis in the graph should be one and a half 1-1/2 times the length of Y-axis.

(4) Method of Plotting the Points: Economic and business statistics are generally positive. These are to be presented in the first quadrant of the graph. Accordingly the point of origin is fixed to the left and lower portion of the graph. On the X-axis, the points are plotted from left to right and on the Y-axis, the points are plotted upward from bottom to top.

(5) Lines of Different Types: If more than one line or curve are to be drawn in the same graph, these lines should be differentiated from each other in the form of broken lines (_ _ _), dotted lines (..........), bold lines (____), etc.

(6) Table of Data: It would be useful to give the table of data along with the graph of data. This helps in the verification of the curve.

(7) Use of false Base line: If the values in a series are very large and the difference between the smallest value and zero is high, then a false base line is drawn.

(8) To draw a line or curve: We mark different points on the graph paper corresponding to different values of a series. These points are joined to make line or curve. The joining line must be uniform throughout its length. It should not be of different thickness of its different points.

TYPES OF GRAPHS

Graphs can be broadly classified into the following two main types:

(A) Time Series Graphs

(B) Frequency Distribution Graphs.

(A) Time Series Graphs

When we observe the values of a variable at different points of time, the series is called time series. A time series is, thus, a chronological arrangement of statistical data. The time series can be represented geometrically with the help of a graph. In time series graph, the time, *i.e.*, years, months, weeks, etc., are taken on the X-axis and the corresponding values of the variables are shown on the Y-axis.

There are many types of time series graphs which are described as follows:

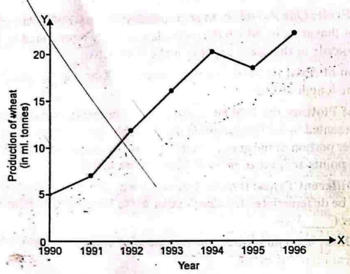
(1) Time Series Graph of One Variable

Time series graphs can be constructed for one variable in which the points are plotted and then joined to form a curve.

Example 18. Represent the following data graphically:

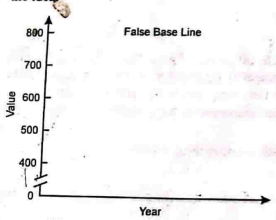
Years:	1990	1991	1992	1993	1994	1995	1996
Production of wheat: (in ml. tonnes)	5	8	13	16	20	17	22

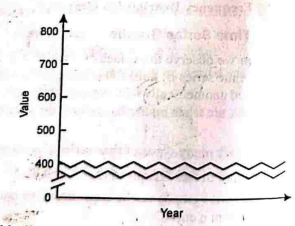
Solution:



False Base Line

A basic rule of the construction of graphs is that the scale on the Y-axis must begin from zero. Such a rule can be followed in all such cases where (*i*) values of the dependent variable vary greatly, and (*ii*) these values lie between zero and the highest value in the scale (or the graph). In many cases this is not so, *i.e.*, (a) the values of the Y or dependent variables lie within a narrow range and (b) the least value is far away from zero. In such cases, false base line technique is used. The technique of false base line is very simple. The vertical scale is broken very close to the base line or near 'zero' which is shown by leaving the space blank or by using a zig-zig line. The following graphs illustrate the idea:

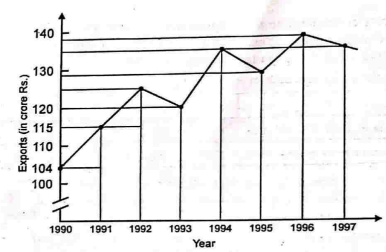




Example 19. Represent the following data graphically:

Years:	1990	1991	1992	1993	1994	1995	1996	1997
Exports (in crore Rs.):	104	115	125	120	135	128	138	135

Solution:



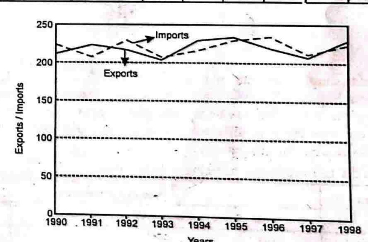
(2) Time Series Graphs of Two or More Variables

Two or more variables can also be shown in a graph provided the units of measurement is the same. If two or more variables are represented on the same graph, it would be better if lines (curves) of different types are drawn. For example, one line is dotted, another thin and a third one thick, etc. can be drawn to distinguish between the variables.

Example 20. Represent the following data graphically:

Years:	1990	1991	1992	1993	1994	1995	1996	1997	1998
Exports (in crore Rs.):	212	220	215	205	225	228	216	208	225
Imports (in crore Rs.):	220	210	225	210	215	225	228	212	220

Solution:



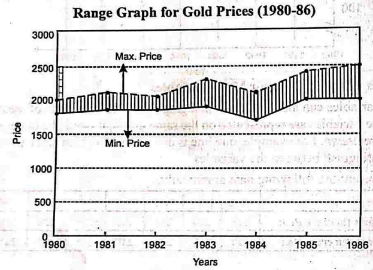
(3) Range Graph

Range graphs or charts are used to depict the maximum and minimum values of variables with respect to time. For example, we may use range graph to show the maximum and minimum prices of commodities, shares, gold, etc. In range graph, time is measured along the X-axis and maximum and minimum value on the Y-axis. Points are plotted corresponding to the maximum and minimum values of the variables corresponding to the given time periods. The points corresponding to the maximum values and minimum values are separately joined by lines (or curves). The space between the two curves is either coloured or shaded to show the range.

Example 21. Show the following data by a suitable graph:

Years:	1980	1981	1982	1983	1984	1985	1986
Minimum price of Gold (Rs.):	1800	1850	1850	1900	1700	2000	2000
Maximum price of Gold (Rs.):	2000	2100	2050	2300	2100	2400	2500

Solution: **Range Graph for Gold Prices (1980-86)**



▶ (4) Band Graph

Band graphs are used to show the constituent parts of the value of a variable with respect to time. For constructing band graph, time is measured along the X-axis. Graphs of different constituent parts are drawn one above the other. This may be facilitated by computing the cumulative totals for each time period. The portions between different graphs are coloured using different colours. Different shades may also be used for this purpose.

Example 22. Represent the following data by using a Band Graph:

Years	Number of students			
	Arts	Commerce	Science	Total
1992	300	100	50	450
1993	400	150	100	650
1994	600	100	150	850
1995	1000	200	150	1350
1996	1000	200	200	1400

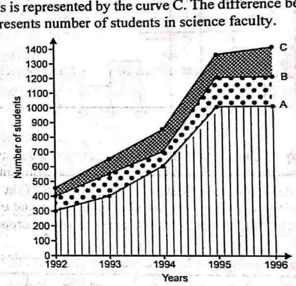
Solution: The procedure of constructing such a graph is as follows:

(*i*) Take the years on the X-axis and the variable on the Y-axis.

(*ii*) Plot the various points for different years for number of students in arts faculty and join them by straight lines. This is represented by curve A.

(*iii*) Add the figures of Arts' students for various years to the figures of commerce students and plot the points and join them by straight line. This is represented by curve B. The difference between the two curves, *i.e.*, B and A, gives number of students in commerce faculty.

(*iv*) Add the figures of Arts and Commerce to Science students and plot the points. This is represented by the curve C. The difference between curve C and curve B represents number of students in science faculty.



(B) Frequency Distribution Graphs

A frequency distribution can be presented graphically in any of the following ways:

(1) Histogram

(2) Frequency Polygon

(3) Frequency Curve

(4) Cumulative Frequency Curve or Ogive

▶ (1) Histogram

The histogram is the most popular and widely used method of presenting a frequency distribution graphically. Histogram is a set of adjoining rectangles whose areas are proportional to class frequencies. For constructing historgram, class intervals are taken on the X-axis and frequencies on Y-axis. Histograms are mainly of two types:

(*a*) **Histogram for equal class intervals.**

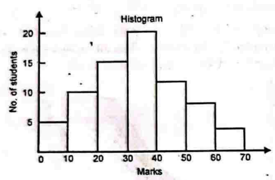
(*b*) **Histogram for unequal class intervals.**

(*a*) **Histogram for Equal Class Intervals:** When class intervals are equal, we take class frequency on the Y-axis, the class intervals on the X-axis and construct adjacent rectangles. The height of the rectangles will be proportional to the frequencies.

Example 23. Represent the following data by means of a histogram:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of students:	5	10	15	20	12	8	4

Solution:



(*b*) **Histogram for Unequal Class Intervals:** When the class intervals are unequal, frequencies are first adjusted before constructing a histogram. For making the adjustment, we first determine the adjustment factor by using the formula: **Adjustment Factor = Size of the class interval/Lowest class interval.** Then we divide the frequencies by adjustment factor and obtain the new adjusted frequency or frequency density. The height of the rectangles would be proportional to the adjusted frequencies.

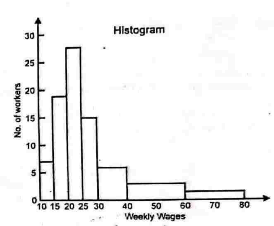
Example 24. Represent the following data by means of histogram:

Weekly wages:	10—15	15—20	20—25	25—30	30—40	40—60	60—80
No. of workers:	7	19	27	15	12	12	8

Solution: Since the class intervals are unequal, frequencies are to be adjusted.

Computations of New Frequency or Frequency Density

Weekly wages	No. of workers (f)	Adjustment factor	New frequencies
10—15	7	5/5 = 1	7
15—20	19	5/5 = 1	19
20—25	27	5/5 = 1	27
25—30	15	5/5 = 1	15
30—40	12	10/5 = 2	12 ÷ 2 = 6
40—60	12	20/5 = 4	12 ÷ 4 = 3
60—80	8	20/5 = 4	8 ÷ 4 = 2



▶ (2) Frequency Polygon

A frequency polygon is a graph of frequency distribution. There are two ways in which a frequency polygon may be constructed:

(*a*) **By Histogram**

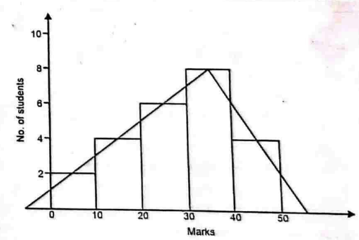
(*b*) **Without Histogram.**

(*a*) **By Histogram:** In this method, a histogram of the frequency distribution is first drawn. Then the upper mid-points of the adjacent rectangles of a histogram are joined. The figure so obtained is the frequency polygon.

Example 25. Draw a frequency polygon for the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	2	4	6	8	4

Solution:



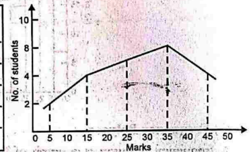
(b) Without Histogram: Frequency polygon can also be constructed without the help of histogram. Under this method, mid-points of various class intervals are taken and the corresponding frequencies to each mid-points are plotted. Then we join all these points by straight lines. The figure so obtained would be the same as obtained by the previous method.

Example 26. Construct a frequency polygon of the following data without constructing a histogram:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	2	4	6	8	4

Solution: To construct a frequency polygon without histogram, firstly we find the mid-points of the class intervals:

Marks	Mid-values (m)	No. of students (f)
0—10	5	2
10—20	15	4
20—30	25	6
30—40	35	8
40—50	45	4



▶ (3) Frequency Curve

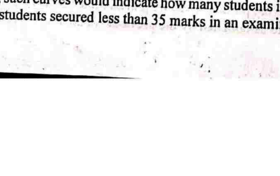
Frequency curve is the smoothed form of a frequency polygon. That is why it is known as **smoothed frequency curve**. A frequency curve can be drawn through different points of the polygon. The curve is drawn freehand in such a way that the area included under it is approximately the same as that of the polygon. With a freehand, the angularities of the frequency polygon are eliminated.

For drawing a frequency curve, it is necessary to first draw the frequency polygon and then smooth it out in such a way that all sudden turns are eliminated.

Example 27. Draw a histogram, a frequency polygon and frequency curve of the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	5	12	15	22	14	4

Solution:



▶ (4) Cumulative Frequency Curve or Ogive

Cumulative frequency curve also known as ogive is the graphical representation of a cumulative frequency distribution. An ogive or cumulative frequency curve is the curve which is constructed by plotting the cumulative frequencies in the form of a smooth curve. From such curves, we come to know about the frequencies corresponding to certain lower limits or upper limits in the distribution of data. For example, such curves would indicate how many students in a class secured more than 40 marks or how many students secured less than 35 marks in an examination.

Construction of a Cumulative Frequency Curve or Ogive

There are two methods of constructing an ogive, viz.,

(*a*) **'Less than' method.**

(*b*) **'More than' method.**

(*a*) **'Less than' Method:** In this method, we start with the upper limits of the classes and go on adding the frequencies. When these frequencies are plotted, we get a rising curve. The resultant curve is called 'Less than' ogive. (See Example 28)

(*b*) **'More than' Method:** In this method, we start with the lower limits of the classes and add frequencies from the bottom. When these frequencies are plotted, we get a declining curve. The resultant curve is called a 'More than' ogive. (See Example 28)

Utility of Ogives

(1) To determine the number or proportion of cases below or above a given value.

(2) To compare two or more frequency distributions.

(3) Ogives are also used to determine certain values graphically such as median, quartiles, deciles, etc.

The following examples would illustrate the construction of ogives:

Example 28. Draw less than and more than ogives from the data given below:

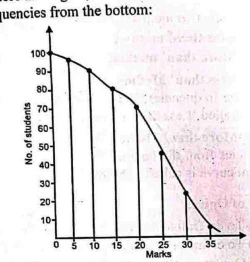
Marks:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
No. of students:	4	6	10	10	25	22	18	5

Solution: **Less than ogive:** In order to draw less than ogive, we start with the upper limits of the classes and go on adding the frequencies from the top.

Marks	*c.f.*
Less than 5	4
Less than 10	10
Less than 15	20
Less than 20	30
Less than 25	55
Less than 30	77
Less than 35	95
Less than 40	100

More than ogive: In order to draw more than ogive, we start with the lower limits of the classes and go on adding the frequencies from the bottom:

Marks	c.f.
More than 0	100
More than 5	96
More than 10	90
More than 15	80
More than 20	70
More than 25	45
More than 30	23
More than 35	5

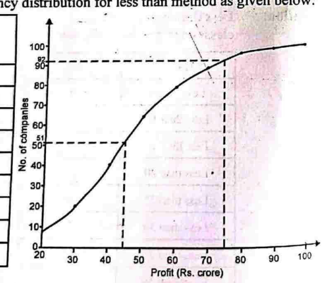


Example 29. Draw a less than ogive and determine the number of companies getting profits between Rs. 45 crores and Rs. 75 crores.

Profits (Rs. crore):	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90	90—100
No. of companies:	8	12	20	24	15	10	7	3	1

Solution: Firstly, we arrange the frequency distribution for less than method as given below:

Profits (Rs. crore)	No. of Companies (c.f.)
Less than 20	8
Less than 30	20
Less than 40	40
Less than 50	64
Less than 60	79
Less than 70	89
Less than 80	96
Less than 90	99
Less than 100	100



(b) It is clear from the graph that the number of companies getting profits less than Rs.75 crore is 92 and the number of companies getting profits less than Rs. 45 crores is 51. Hence the number of companies getting profits between Rs. 45 crores and Rs. 75 crores is $92 - 51 = 41$.

EXERCISE 4.3

1. Represent the following data graphically:

Years:	1980	1981	1982	1983	1984	1985
Wheat Production (in '000 tonnes):	5	12	8	15	20	17

2. Represent the following data graphically:

Years:	1980-81	1981-82	1982-83	1983-84	1984-85	1985-86	1986-87
Exports (Rs. in lakh):	51	56	58	65	60	66	61
Imports (Rs. in lakh):	53	54	60	70	65	73	70

3. Following are the maximum and minimum market values of the shares of a company. Represent the data by using a range chart.

Years:	July '86	Aug. '86	Sept. '86	Oct. '86	Nov. '86	Dec. '86
Minimum (in Rs.):	45	41	47	42	50	55
Maximum (in Rs.):	50	60	67	68	68	70

4. Represent the following data by using band chart:

Year	Production (in '000 tonnes)		
	Wheat	Rice	Pulses
1996	20	8	4
1997	22	10	3
1998	25	7	5
1999	23	11	7
2000	30	13	4
2001	25	12	5
2002	31	15	7

5. Construct a histogram from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of students:	5	12	20	35	24	12	4

6. Represent the following data by means of a histogram:

Daily wages (Rs.):	10—20	20—30	30—40	40—60	60—100
No. of workers:	5	20	30	40	20

7. Construct a histogram with the help of data given below:

Marks:	10—19	20—29	30—39	40—49	50—59	60—69
No. of students:	4	25	45	60	35	10

8. Draw a frequency polygon from the following data by using (a) histogram (b) without using histogram:

Daily wages (in Rs.):	10—15	15—20	20—25	25—30	30—35
No. of workers:	40	70	60	80	60

9. From the following data, draw: (*i*) Histogram, (*ii*) Frequency polygon and (*iii*) Frequency curve.

Wage groups (Rs.):	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90
No. of workers:	2	4	11	15	25	18	15	4	2

10. From the following data, draw: (*i*) Histogram, (*ii*) Frequency polygon and (*iii*) Frequency curve.

Wage (in Rs.):	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of persons:	2	7	15	25	18	15	8

11. Draw 'less than' as well as 'more than' ogives for the following data:

Weight (in kg)	30—34	35—39	40—44	45—49	50—54	55—59	60—64
Frequency:	3	5	12	18	14	6	2

(**Hint:** Convert into Exclusive Groups]

12. Prepare a less than cumulative frequency curve for the following data:

Weekly wages:	0—20	20—40	40—60	60—80	80—100
No. of workers:	40	51	64	38	7

Find the number of workers earning between Rs. 55 and Rs. 65 per week.

MISCELLANEOUS SOLVED EXAMPLES

Example 30. In a sample study about coffee habits in two towns, the following information is given:

Town A: Females were 40%, total coffee drinkers were 45% and male non-coffee drinkers were 20%.

Town B: Males were 55%, male non-coffee drinkers were 30% and female coffee drinkers were 15%.

Present the above data in a tabular form.

Solution: **Percentage of coffee drinkers in Towns A and B**

	Town A			Town B		
	Male	Female	Total	Male	Female	Total
Coffee Drinkers	40	5	45	25	15	40
Non-coffee Drinkers	20	35	55	30	30	60
Total	60	40	100	55	45	100

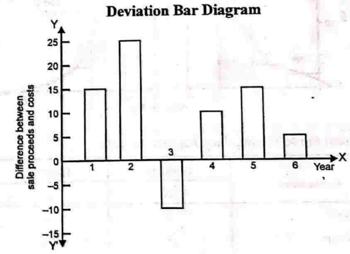
Example 31. Present the following data by a deviation bar diagram, showing the difference between proceeds and costs of a firm.

Year	Sale proceeds (Rs. in lakh)	Costs (Rs. in lakh)
1	115	100
2	140	115
3	145	155
4	150	140
5	160	145
6	170	165

Solution: In order to present the data by deviation bar diagram, we first calculate the deviations between sale proceeds and cost as follows:

Year	Sale proceeds	Costs	Difference between sale proceeds and costs
1	115	100	15
2	140	115	25
3	145	155	–10
4	150	140	10
5	160	145	15
6	170	165	5

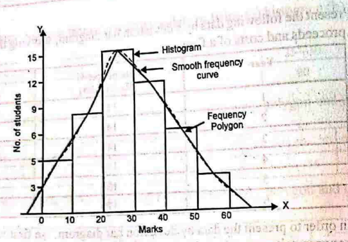
Deviation Bar Diagram



Example 32. Prepare histogram, frequency polygon and smooth frequency curve from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	5	8	15	11	6	4

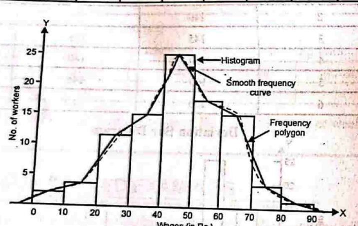
Solution:



Example 33. From the following data, construct (1) Frequency Histogram (2) Frequency Polygon and (3) Frequency Curve.

Wages (in Rs.):	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90
No. of workers:	2	4	11	15	25	18	15	4	1

Solution:



Example 34. Represent the following data diagrammatically:

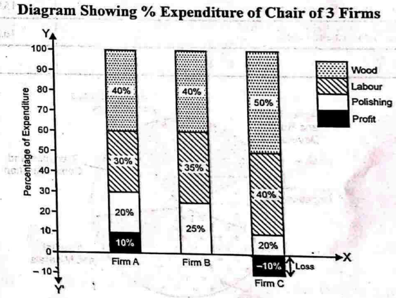
Particulars	Firm 'A' Rs.	Firm 'B' Rs.	Firm 'C' Rs.
Wood	4	4	5
Labour	3	3.5	4
Polishing	2	2.5	2
Cost	9	10	11
Selling Price	10	10	10
Profit or Loss	+ 1	0	– 1

Solution:

Selling Price of Chair on Percentage Basis

Particulars	Firm 'A'		Firm 'B'		Firm 'C'	
	Actual	%	Actual	%	Actual	%
Wood	4	40	4.0	40	5	50
Labour	3	30	3.5	35	4	40
Polishing	2	20	2.5	25	2	20
Cost	9	90	10.0	100	11	110
Selling Price	10	100	10.0	100	10	100
Profit or Loss	+ 1	+ 10%	—	—	– 1	– 10%

Diagram Showing % Expenditure of Chair of 3 Firms



Example 35. Draw a pie diagram for the following data of Sixth Five Year Plan Public Sector outlays:

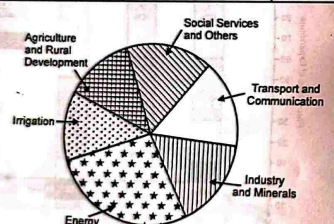
Agricultural and Rural Development	12.9%
Irrigation	12.5%
Energy	27.2%
Industry and Minerals	15.4%
Transport and Communication	15.9%
Social Services and Others	16.1%

Solution: The angle at the centre is given by

$$\frac{\text{Percentage outlay}}{100} \times 360 = \text{Percentage outlay} \times 3.6^\circ$$

Computations for Pie Diagram

Sector	Percentage outlays	Angle of the centre
(1)	(2)	(3) = (2) × 3.6°
Agriculture and Rural Development	12.9	12.9 × 3.6 = 46°
Irrigation	12.5	12.5 × 3.6 = 45°
Energy	27.2	27.2 × 3.6 = 98°
Industry and Minerals	15.4	15.4 × 3.6 = 56°
Transport and Communication	15.9	15.9 × 3.6 = 57°
Social Services and others	16.1	16.1 × 3.6 = 58°
Total	100.0	360°



Example 36. The following table shows the area in millions of sq. km of oceans of the world:

Ocean	Area (million sq. km)
Pacific	70.8
Atlantic	41.2
Indian	28.5
Antarctic	7.6
Arctic	4.8

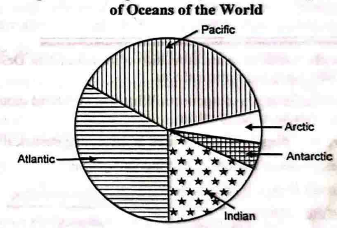
Draw a pie diagram to represent the data.

Solution:

Calculations for Pie Diagram

Ocean	Area	Degrees
Pacific	70.8	$\frac{70.8}{152.9} \times 360 = 166.7^\circ$
Atlantic	41.2	$\frac{41.2}{152.9} \times 360 = 97.0^\circ$
Indian	28.5	$\frac{28.5}{152.9} \times 360 = 67.1^\circ$
Antarctic	7.6	$\frac{7.6}{152.9} \times 360 = 17.9^\circ$
Arctic	4.8	$\frac{4.8}{152.9} \times 360 = 11.3^\circ$
Total	152.9	360°

Pie Diagram Showing the Area (in Millions of Square Kms) of Oceans of the World



Example 37. Draw the 'less than' and 'more than' ogive on the same graph paper from the data given below:

Weekly Wages (Rs.):	0—20	20—40	40—60	60—80	80—100
No. of Workers:	10	20	40	20	10

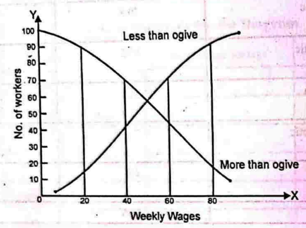
Solution: (*i*) Less than method

Weekly wages (Rs.)	c.f.
Less than 20	10
Less than 40	30
Less than 60	70
Less than 80	90
Less than 100	100

(*ii*) More than method

Weekly wages (Rs.)	c.f.
More than 0	100
More than 20	90
More than 40	70
More than 60	30
More than 80	10

Both less than and more than ogives based on the above data are presented in the following graph:



QUESTIONS

1. Explain the purpose of classification and tabulation of statistical data. Describe the rules that serve as a guide in tabulating the statistical data.
2. What is a statistical table? State its components. What considerations should govern framing a table? Also discuss various types of tables.
3. Explain different types of diagrams that are used in presenting statistical data.
4. What are the different types of frequency distributions graphs?
5. What is an ogive curve? How is it constructed?
6. Write short notes on the following:
   (*i*) Histogram; (*ii*) Ogive;
   (*iii*) Frequency Polygon; (*iv*) Frequency Curve;
   (*v*) Pie Diagram.
7. Discuss the utility and limitations of graphic method of presentation of statistical data.
8. Explain the advantages and limitations of diagrammetic and graphic presentations.
9. What is frequency distribution? Describe various ways by which a frequency distribution can be represented graphically.

5 Measures of Central Tendency

INTRODUCTION

A large number of big figures is confusing to mind. It is also difficult to analyse them. In order to reduce the complexity of data and to make them comparable, it is essential that the various figures which are being compared are reduced to one single figure each. If, for example, a comparison is made between the marks obtained by 100 students of B.Com. II class of a college and the marks obtained by 100 students belonging to B.Com. II class of another college, it would be impossible to arrive at any conclusion, if the two series relating to these marks are directly compared. On the other hand, if each of these series is represented by one single figure, comparison and understanding would become very easy. The single figure which represents the whole set of data is called an average. Averages are also called **measures of central tendency** or **measures of location.**

MEANING OF AVERAGE OR CENTRAL TENDENCY

An average is a single value which represents the whole set of figures and all other individual items concentrate around it. In other words, an average is single value within the range of the data that is used to represent all of the values in the series. Such an average is somewhere within the range of the data, it is therefore called measure of central tendency.

DEFINITIONS

Some important definitions of average are given below:

1. "*An average is a single figure that represents the whole group.*" —**Clark**
2. "*An average is a single value selected from a group of values to represent them in some ways.*" —**A.E. Waugh**
3. "*An average is a typical value that represents all the individual values in a series.*" —**Croxten and Cowden**

The above definitions make it clear that an average is a single value that represents a group of values.

PURPOSE AND FUNCTIONS OF AVERAGE

Some of the important functions and purposes of averages are as under:

(*i*) **Brief Description:** The main purpose of average is to present a simple and systematic description of the raw data. The raw data may be complex and unorganised. An average reduces a mass of data into a single typical figure. It enables one to draw a general conclusion about the characteristics of the phenomena under study.

(ii) **Helpful in Comparison:** Averages help in making comparison of different sets of data. For example, a comparison of the per capita income of India and USA shows that per capita income of India is much less in comparison to the per capita income of USA. It leads us to the conclusion that India is a poor country in comparison to USA.

(iii) **Helpful in the Formulation of Policies:** Averages help in the formulation of policies. To illustrate, when the Government finds that the average per capita income in India is very low, it can formulate suitable policies to increase it.

(iv) **Basis of Statistical Analysis:** Averages constitute the basis of statistical analysis. For example, if one knows the average marks secured by the students of a class in their different subjects, one can easily analyse the general interest of the students in different subjects.

(v) **Representation of the Universe:** Averages represent the universe or the mass of statistical data. Accordingly, averages help in knowing the characteristics of the universe as a whole. For example, by calculating the per capita income of a nation, we get one single value that gives an idea of the economic condition of that nation.

CHARACTERISTICS/PROPERTIES OF A GOOD AVERAGE

A good average should possess the following properties:

(1) It should be easy to understand.
(2) It should be simple to compute.
(3) It should be uniquely defined.
(4) It should be based on all observations.
(5) It should not be unduly affected by extreme values.
(6) It should be capable of further algebraic treatment.

TYPES OF AVERAGES

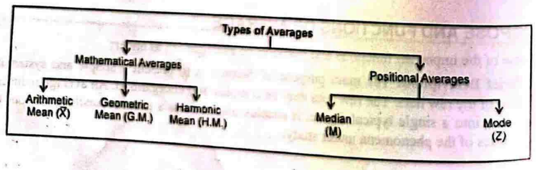
There are different kinds of averages. The following are the important types of averages which are commonly used in business and industry:

1. Mathematical Averages

(i) Arithmetic Mean
(ii) Geometric Mean
(iii) Harmonic Mean

2. Positional Averages

(i) Median
(ii) Mode



(1) ARITHMETIC MEAN

Arithmetic mean is the most popular and widely used measure of central tendency. Generally, when we talk of 'average', it signifies arithmetic mean. It is generally known as 'Mean'. Arithmetic mean is defined as the value which is obtained by adding all the items of a series and dividing this total by the number of items. Arithmetic mean may be of two types:

(a) Simple Arithmetic Mean.

(b) Weighted Arithmetic Mean.

Simple Arithmetic Mean

Individual Series

In case of individual series, arithmetic mean can be computed by applying any of the two methods:

(i) Direct Method

When direct method is used, the following formula is used:

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

Here, $\bar{X}$ = Arithmetic mean; ΣX = Sum of the values of the item of a series.
N = Number of observations.

Steps for Calculation

(i) Add together all the values of the variable X and obtain the total, *i.e.*, ΣX

(ii) Divide this total by the number of observations, *i.e.*, N. The result will give the value of arithmetic mean.

Example 1. The pocket allowances (in Rs.) of ten students are given below:

15, 20, 30, 22, 25, 18, 40, 50, 55 and 65

Calculate the arithmetic mean of pocket allowance.

Solution: Let pocket allowance be denoted by X

Students	Pocket Allowance (X)
1	15
2	20
3	30
4	22
5	25
6	18
7	40
8	50
9	55
10	65
$N = 10$	$\Sigma X = 340$

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

$$\therefore \quad \overline{X} = \frac{340}{10} = 34$$

Thus, the average pocket allowance is Rs. 34.

▶ ***(ii)* Short-cut Method**

When the number of observations are large, the arithmetic mean can be calculated by using short-cut method or assumed mean method. When deviations are taken from an assumed mean, the following formula is used:

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

Where, d = Deviations of the items from the assumed mean, *i.e.*, $X - A$

A = Assumed Mean.

Steps for Calculation

(*i*) Any one of the items in the series is taken as assumed mean A.

(*ii*) Take the deviations of the items from the assumed mean, *i.e.*, $X - A$ and denote these deviations by 'd'.

(*iii*) Obtain the sum of these deviations, *i.e.*, Σd.

(*iv*) Substitute the values of A, Σd and N in the above formula. The result will give the value of arithmetic mean.

Example 2. The pocket allowances (in Rs.) of ten students are given below:

15, 20, 30, 22, 25, 18, 40, 50, 55 and 65

Calculate the arithmetic mean by taking 40 as assumed mean.

Solution:

Calculation of Arithmetic Mean

Students	Pocket Allowances (X)	$A = 40$ $d = X - 40$
1	15	$15 - 40 = -25$
2	20	$20 - 40 = -20$
3	30	$30 - 40 = -10$
4	22	$22 - 40 = -18$
5	25	$25 - 40 = -15$
6	18	$18 - 40 = -22$
7	40 = A	$40 - 40 = 0$
8	50	$50 - 40 = +10$
9	55	$55 - 40 = +15$
10	65	$65 - 40 = +25$
$N = 10$		$\Sigma d = -60$

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

Substituting to values, we get

$$\overline{X} = 40 + \frac{(-60)}{10} = 40 - 6 = 34$$

Hence, the average pocket allowance is Rs. 34.

Discrete Series

For calculating arithmetic mean in discrete series, the following two methods may be used:

(*i*) Direct method

(*ii*) Short-cut method.

▶ **(*i*) Direct Method**

When direct method is used, the following formula is used:

$$\overline{X} = \frac{\Sigma fX}{N}$$

Where, f = frequency, X = values of the variable, N = total number of observations, *i.e.*, Σf.

Steps for Calculation

(*i*) Multiply the frequency of each item with the values of variable and obtain total ΣfX.

(*ii*) Find the sum of frequencies, *i.e.*, Σf or N.

(*iii*) Divide the total obtained (ΣfX) by the number of observations N or Σf. The result would be the required arithmetic mean.

Example 3. Calculate the arithmetic mean from the following data:

Wages (Rs.):	10	20	30	40	50
No. of workers:	4	5	3	2	5

Solution: Denoting wages by X and number of workers by f.

Calculation of Arithmetic Mean

Wages (X)	No. of workers (f)	fX
10	4	40
20	5	100
30	3	90
40	2	80
50	5	250
	N or $\Sigma f = 19$	$\Sigma fX = 560$

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{560}{19} = 29.47$$

Thus, the mean wage is Rs. 29.47.

▶ ***(ii)* Short-cut Method**

When this method is used, the formula for calculating arithmetic mean is:

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

Where, A = Assumed mean; $d = X - A$; N = Total number of observations, *i.e.,* Σf

Steps for Calculation

(*i*) Any one of the item in the series is taken as assumed mean A.

(*ii*) Take the deviations of the items from the assumed mean, *i.e.,* $X - A$ and denote these deviations by d.

(*iii*) Multiply these deviations with the respective frequency and obtain the total Σfd.

(*iv*) Divide the total obtained (Σfd) by the total frequency Σf or total number of observations N.

Example 4. From the following data of the wages obtained by 19 workers of a factory, calculate the arithmetic mean by using short-cut method:

Wages (Rs.):	10	20	30	40	50
No. of workers:	4	5	3	2	5

Solution:

Calculation of Arithmetic Mean

Wages (Rs.) (X)	No. of workers (f)	$A = 30$ $d = X - 30$	fd
10	4	$10 - 30 = -20$	-80
20	5	$20 - 30 = -10$	-50
30 A	3	$30 - 30 = 0$	0
40	2	$40 - 30 = +10$	$+20$
50	5	$50 - 30 = +20$	$+100$
	Σf or $N = 19$		$\Sigma fd = -10$

$$\bar{X} = A + \frac{\Sigma fd}{N} = 30 - \frac{10}{19} = 30 - 0.53 = 29.47$$

Thus, the mean wage is 29.47.

Continuous Series

In continuous series, arithmetic mean can be calculated by using any one of the method:

(*i*) **Direct Method**

(*ii*) **Short-cut Method**

(*iii*) **Step deviation Method.**

▶ ***(i)* Direct Method:**

When direct method is used, we apply the following formula to calculate arithmetic mean:

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

Where, m = mid-point of various classes, f = frequency of each class, N = total frequency.

Steps for Calculation

(*i*) Obtain the mid-value of each class and denote it by m.

(*ii*) Multiply each mid-value by the corresponding frequency and obtain the total Σfm.

(*iii*) Divide the total obtained (Σfm) by the sum of frequencies, *i.e.,* N.

Example 5. Calculate the arithmetic mean from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	20	24	40	36	20

Solution:

Calculation of Arithmetic Mean

Marks	f	Mid-value (m)	fm
0—10	20	$\frac{0+10}{2} = 5$	100
10—20	24	$\frac{10+20}{2} = 15$	360
20—30	40	$\frac{20+30}{2} = 25$	1,000
30—40	36	$\frac{30+40}{2} = 35$	1,260
40—50	20	$\frac{40+50}{2} = 45$	900
	Σf or $N = 140$		$\Sigma fm = 3620$

$$\bar{X} = \frac{\Sigma fm}{N} = \frac{3620}{140} = 25.85$$

▶ ***(ii)* Short-cut Method**

When this method is used, arithmetic mean is computed by applying the following formula:

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

Where, A = assumed mean; d = deviations of mid-value from assumed mean, *i.e.,* $m - A$;
N = total number of observations, *i.e.,* Σf.

Steps for Calculation

(*i*) Find the mid-value of each class and denote it by m.

(*ii*) Take any mid-value as assumed mean A.

(*iii*) Take deviations of the mid-value (m) from the assumed mean $m - A$ and denote it by d.

(*iv*) Multiply the respective frequencies of each class by these deviations and obtain the total Σfd.

(*v*) Divide the total obtained (Σfd) by the total frequency Σf or total number of observations.

Example 6. Calculate the arithmetic mean by the short-cut method from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	20	24	40	36	20

Solution:

Calculation of Arithmetic Mean

Marks	No. of students (f)	Mid-value (m)	$A = 25$ $d = m - A$	fd
0—10	20	5	−20	− 400
10—20	24	15	−10	− 240
20—30	40	25 = A	0	0
30—40	36	35	+10	+360
40—50	20	45	+20	+400
	N = 140			$\Sigma fd = 120$

$$\overline{X} = A + \frac{\Sigma fd}{N} = 25 + \frac{120}{140} = 25 + 0.85 = 25.85$$

Thus, the mean marks = 25.85

(iii) Step Deviation Method

In case of continuous series with class intervals of equal magnitude, the arithmetic mean is computed by applying the following formula:

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

Where, $d' = \frac{m - A}{i}$; m = mid-value of the class; i = common magnitude of the class intervals; A = assumed mean.

Note: Step deviation method is most commonly used in case of continuous series.

Steps for Calculation

(i) Find the mid-value of each class and denote it by m.

(ii) Take any mid-value as assumed mean A.

(iii) Take deviations of the mid-value (m) from the assumed mean $m - A$ and denote it by d.

(iv) Compute step deviations d'. These are obtained by dividing the deviations by the magnitude of class intervals, *i.e.*, $d' = d / i$.

(v) Multiply the respective frequencies of each class by these deviations (d') and obtain the total $\Sigma fd'$.

(vi) Divide the total obtained ($\Sigma fd'$) by the total frequency Σf or N and then multiply by 'i' in the formula for getting arithmetic mean.

Example 7. From the following data, compute arithmetic mean by step deviation method:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	20	24	40	36	20

Solution:

Calculation of Arithmetic Mean

Marks	Mid-value (m)	f	A = 25 $d = m - 25$	$d' = d/10$	fd'
0—10	5	20	−20	−2	−40
10—20	15	24	−10	−1	−24
20—30	25 = A	40	0	0	0
30—40	35	36	+10	+1	+36
40—50	45	20	+20	+2	+40
		N = 140			$\Sigma fd' = 12$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 25 + \frac{12}{140} \times 10$$

$$= 25 + 0.85 = 25.85$$

Inclusive Class Intervals

While calculating mean in a continuous series with inclusive class intervals, it is not necessary to convert the series into exclusive class intervals by adjusting class limits because the mid-values remain the same whether or not the adjustment is made:

Example 8. Calculate the arithmetic mean from the following data:

Size:	20—29	30—39	40—49	50—59	60—69
Frequency:	10	8	6	4	2

Solution:

Calculation of Arithmetic Mean

Size	Mid-value (m)	f	A = 44.5 $d = m - A$	$d' = d/10$	fd'
20—29	24.5	10	− 20	−2	−20
29—39	34.5	8	− 10	−1	−8
40—49	44.5 = A	6	0	0	0
50—59	54.5	4	+10	+1	+4
60—69	64.5	2	+20	+2	+4
		N = 30			$\Sigma fd' = -20$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 44.5 + \frac{-20}{30} \times 10$$

$$= 44.5 - \frac{20}{3} = 44.5 - 6.67 = 37.83$$

Cumulative Frequency Series

While calculating mean in a cumulative frequency series, it is necessary to convert the series into a simple frequency series, and only after that the arithmetic mean is calculated.

Example 9. Calculate the arithmetic mean from the following data:

Marks	No. of students
Less than 10	5
Less than 20	17
Less than 30	31
Less than 40	41
Less than 50	49

Solution: Since, cumulative frequencies are given, first we find the simple frequencies:

Calculation of Arithmetic Mean

Marks	No. of students (f)	Mid-value (m)	$A = 25$, $d = m - 25$	$d' = d/10$	fd'
0—10	5	5	–20	–2	–10
10—20	$17 - 5 = 12$	15	–10	–1	–12
20—30	$31 - 17 = 14$	25 = A	0	0	0
30—40	$41 - 31 = 10$	35	+10	+1	+10
40—50	$49 - 41 = 8$	45	+20	+2	+16
	$N = 49$				$\Sigma fd' = +4$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 25 + \frac{4}{49} \times 10 = 25 + 0.81 = 25.81$$

Example 10. Calculate the arithmetic mean from the following data:

Marks	No. of students
More than 0	30
More than 2	28
More than 4	24
More than 6	18
More than 8	10

Solution: Since, cumulative frequencies are given, first we find the simple frequencies.

Calculation of Arithmetic Mean

Marks	No. of students (f)	Mid-value (m)	$A = 5$, $d = m - 5$	$d' = d/2$	fd'
0—2	$30 - 28 = 2$	1	–4	–2	–4
2—4	$28 - 24 = 4$	3	–2	–1	–4
4—6	$24 - 18 = 6$	5 = A	0	0	0
6—8	$18 - 10 = 8$	7	+2	+1	+8
8—10	10	9	+4	+2	+20
	$N = 30$				$\Sigma fd' = 20$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 5 + \frac{20}{30} \times 2$$

$$= 5 + 1.33 = 6.33$$

Missing Frequency

Example 11. From the following data, calculate the missing value when its mean is 115.86:

Wages (Rs.):	110	112	113	117	—	125	128	130
No. of workers:	25	17	13	15	14	8	6	2

Solution: Let the missing item be denoted by a

Calculation of Missing Value

Wages (Rs.) (X)	No. of workers (f)	fX
110	25	2750
112	17	1904
113	13	1469
117	15	1755
a	14	$14a$
125	8	1000
128	6	768
130	2	260
	$N = 100$	$\Sigma fX = 9906 + 14a$

$$\overline{X} = \frac{\Sigma fX}{N}$$

or $115.86 = \dfrac{9906 + 14a}{100}$

or $115.86 \times 100 = 9906 + 14a$

or $14a = 1680$

or $a = \dfrac{1680}{14}$

$a = 120$

Hence, the missing value is 120.

Example 12. Find the missing frequency from the following data for which the mean is 52:

Marks:	10—20	20—30	30—40	40—50	50—60	60—70	70—80
No. of students:	5	3	4	—	2	6	13

Solution: Let the missing frequency be denoted by f

Calculation of Missing Frequency

Marks	No. of students (f)	Mid-value (m)	fm
10—20	5	15	75
20—30	3	25	75
30—40	4	35	140
40—50	f	45	$45f$
50—60	2	55	110
60—70	6	65	390
70—80	13	75	975
	$N = 33 + f$		$\Sigma fm = 1765 + 45f$

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

$\Rightarrow$ $$52 = \frac{1765 + 45f}{33 + f}$$

or $52(33 + f) = 1765 + 45f$

or $1716 + 52f = 1765 + 45f$

or $7f = 49$

$$f = \frac{49}{7} = 7$$

Thus, the missing frequency is 7.

IMPORTANT TYPICAL EXAMPLES

Example 13. The following are the monthly salaries in rupees of 20 employees of a firm:

130	62	145	118	125	76	151	142	110	98
65	116	100	103	71	85	80	122	132	95

The firm gives bonus of Rs. 10, 15, 20, 25 and 30 for individuals in the respective salary groups exceeding Rs. 60 but not exceeding Rs. 80, exceeding Rs. 80 but not exceeding Rs.100 and so on upto exceeding Rs. 140 but not exceeding Rs. 160. Find the average bonus paid per employee.

Solution:

Monthly salary	Tally Bars	No. of workers (f)	Bonus (X)	fX
Exceeding 60 but not exceeding 80	$\cancel{||||}$	5	10	50
Exceeding 80 but not exceeding 100	||||	4	15	60
Exceeding 100 but not exceeding 120	||||	4	20	80
Exceeding 120 but not exceeding 140	||||	4	25	100
Exceeding 140 but not exceeding 160	|||	3	30	90
		Σf or $N = 20$		$\Sigma fX = 380$

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{380}{20} = \text{Rs. } 19$$

$\therefore$ Average bonus = Rs. 19.

Example 14. The sum of the deviations of a certain number of items measured from 2.5 is 50 and from 3.5 is – 50. Find N and $\bar{X}$.

Solution: Now, when assumed mean (A) is 2.5, $\Sigma d = 50$

$\therefore$ $$\bar{X} = A + \frac{\Sigma d}{N} = 2.5 + \frac{50}{N} \quad ...(i)$$

When assumed mean (A) is 3.5, $\Sigma d = -50$

$\therefore$ $$\bar{X} = A + \frac{\Sigma d}{N} = 3.5 - \frac{50}{N} \quad ...(ii)$$

From (i) and (ii),

$$2.5 + \frac{50}{N} = 3.5 - \frac{50}{N}$$

$$\frac{100}{N} = 1 \Rightarrow N = 100$$

From (i), $$\bar{X} = 2.5 + \frac{50}{100} = \frac{250 + 50}{100}$$

$$= \frac{300}{100} = 3$$

$\therefore$ $\bar{X} = 3, N = 100$

EXERCISE 5.1

1. Following are the marks obtained by 8 students. Calculate the arithmetic mean:

Marks:	15	18	16	45	32	40	30	28

[Ans. $\bar{X} = 28$ marks]

2. Following are the marks obtained by 25 students in Economics. Find out the mean marks by using direct and short-cut method:

Marks:	10	20	30	40	50	60
No. of students:	5	2	3	8	4	3

[Ans. $\bar{X} = 35.2$]

3. Calculate the arithmetic mean from the following data:

Class interval:	20—25	25—30	30—35	35—40	40—45	45—50	50—55
Frequency:	10	12	8	20	11	4	5

[Ans. $\bar{X} = 35.5$]

4. Calculate the average marks from the following data by using Short-cut method:

Marks:	6—10	11—15	16—20	21—25	26—30
No. of students:	20	30	50	40	10

[Ans. $\overline{X} = 17.67$]

5. Find out the arithmetic mean from the following data:

Marks (less than):	5	10	15	20	25	30
No. of students:	10	23	30	54	69	80

[Ans. $\overline{X} = 15.87$]

6. Find out the arithmetic mean from the following data:

Bonus (more than):	0	4	8	12
No. of workers:	15	11	3	1

[Ans. $\overline{X} = 6$]

7. Calculate the arithmetic mean from the following:

X:	–40 to –30	–30 to –20	–20 to –10	–10 to 0	0 to 10	10 to 20	20 to 30
f:	10	28	30	42	65	180	10

[Ans. $\overline{X} = 4.287$]

8. Find out the missing frequencies of the following series if the A.M. is 35 and total number of items is 100:

Class interval:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
Frequency:	5	10	?	4	20	3	?

[Ans. $f_1 = 41$ app.; $f_2 = 17$ app.]

9. The following are the monthly salaries in rupees of 30 employees in a firm:

140	139	120	114	100	88	62	77	99	103
108	129	144	148	134	63	69	148	132	118
142	116	123	104	95	80	85	106	123	133

The firm gave bonus of Rs. 10, 15, 20, 25, 30 and 35 for individuals in the respective salary groups: exceeding 60 but not exceeding 75, exceeding 75 but not exceeding 90 and so on upto exceeding 135 but not exceeding 150. Find out the average bonus paid.

[Ans. Average bonus = Rs. 24.5]

10. The sales of a balloon seller on seven days of a week are given as below:

Days:	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Sales (in Rs.):	100	150	125	140	160	200	250

If the profit is 20% of sales, find his average profit per day.

[Hint: Calculate profit per day $= \frac{20}{100} \times$ Sales]

[Ans. Rs. 32.14]

11. The sum of the deviations of a certain number of items measured from 4 is 72 and the sum of the deviations of the items from 7 is – 3. Find the number of items and their mean.

[Ans. $N = 25, \overline{X} = 6.88$]

COMBINED ARITHMETIC MEAN

If we have the arithmetic mean and the number of items of two or more than two related sub-groups, we can calculate the combined arithmetic mean of the whole group by applying the following formula:

$$\overline{X}_{12} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2}{N_1 + N_2}$$

Where, $\overline{X}_{12}$ = Combined arithmetic mean of the two groups.

$\overline{X}_1$ = A.M. of the first group; $\overline{X}_2$ = A.M. of the second group.

The above formula can be extended to calculate the arithmetic mean of three or more groups. For example, combined arithmetic mean of three groups is given by:

$$\overline{X}_{123} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2 + N_3 \overline{X}_3}{N_1 + N_2 + N_3}$$

The following examples illustrate the application of the above formula:

Example 15. The mean height of 25 male workers in a factory is 61 cms. and the mean height of 35 female workers in the same factory is 58 cm. Find the combined mean height of 60 workers in the factory.

Solution: Given: $N_1 = 25$, $\overline{X}_1 = 61$, $N_2 = 35$, $\overline{X}_2 = 58$

$$\overline{X}_{12} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2}{N_1 + N_2}$$

$$= \frac{(25 \times 61) + (35 \times 58)}{25 + 35} = \frac{1525 + 2030}{60} = \frac{3555}{60} = 59.25$$

$\therefore \quad \overline{X}_{12} = 59.25$

Thus, the combined mean height of 60 workers is 59.25 cm.

Example 16. The mean wage of 100 workers in a factory, running two shifts of 60 and 40 workers respectively is Rs. 38. The mean wage of 60 labourers working in the morning shift is Rs. 40. Find the mean wage of 40 workers working in the evening shift.

Solution: Given: $N = 100$, $\overline{X}_{12} = 38$, $N_1 = 60$, $\overline{X}_1 = 40$; $N_2 = 40$, $\overline{X}_2 = ?$

Using the formula of combined arithmetic mean, we have:

$$\overline{X}_{12} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2}{N_1 + N_2}$$

$$38 = \frac{60 \times 40 + 40 \overline{X}_2}{100}$$

or $\quad 3800 = 2400 + 40 \overline{X}_2$

or $\quad 1400 = 40 \overline{X}_2$

$\therefore \quad \overline{X}_2 = \frac{1400}{40} = 35$

$\therefore \quad \overline{X}_2 = 35$

IMPORTANT TYPICAL EXAMPLES

Example 17. The mean monthly salary paid to all employees in a certain company was Rs.600. The mean monthly salaries paid to male and female employees were Rs. 620 and Rs. 520 respectively. Find the percentage of male to female employees in the company.

Solution: Given: $\overline{X}_{12} = 600,\ \overline{X}_1 = 620,\ \overline{X}_2 = 520$

Let N_1= Male Employees and N_2= Female Employees.

Using the formula of combined arithmetic mean, we have

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

$$600 = \frac{N_1(620) + N_2(520)}{N_1 + N_2}$$

or $600(N_1 + N_2) = 620\,N_1 + 520\,N_2$

or $600N_1 + 600N_2 = 620\,N_1 + 520\,N_2$

$\therefore$ $80N_2 = 20\,N_1$

or $\dfrac{N_1}{N_2} = \dfrac{80}{20} = \dfrac{4}{1}$

or $N_1 : N_2 = 4 : 1$

Hence, the percentage of male employees $= \dfrac{4}{4+1} \times 100 = 80\%$ and the percentage of female employees $= \dfrac{1}{4+1} \times 100 = 20\%$

Aliter: The problem can also be solved by using the following method:

Let x be the percentage of male in the combined group. Therefore, the percentage of women $= 100 - x$.

We are given that $\overline{X}_1(\text{men}) = \text{Rs. } 620$ and $\overline{X}_2(\text{women}) = \text{Rs. } 520$

Also $\overline{X}_{12}(\text{combined mean}) = 600$

$$\therefore \quad 600 = \frac{620x + 520(100 - x)}{100} = \frac{620x + 52000 - 520x}{100}$$

$\Rightarrow 60{,}000 - 52{,}000 = 100x$

$\Rightarrow \quad x = \dfrac{8000}{100} = 80\%$

Thus, there are 80% men and 20% women in this group.

Example 18. A bookseller has 150 books of Economics and Accountancy. The average price of these books is Rs. 40 per book. Average price of books on Economics is Rs. 43 and that of Accountancy is Rs. 35. Find the number of books on Economics with the seller.

Solution: Let, N_1 = Economics, N_2 = Accountancy

$N_1 + N_2 = 150$

$\Rightarrow \quad N_2 = 150 - N_1$

$\overline{X}_{12} = 40, \overline{X}_1 = 43,\ \overline{X}_2 = 35$

$$\therefore \quad \overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

$$40 = \frac{43\,N_1 + 35(150 - N_1)}{150}$$

$$40 = \frac{43\,N_1 + 5250 - 35\,N_1}{150}$$

$$40 = \frac{8\,N_1 + 5250}{150}$$

$6000 = 8N_1 + 5250$

$\therefore \quad 8N_1 = 750$

$\Rightarrow \quad N_1 = \dfrac{750}{8} = 93.75 \cong 94$

Thus, books on Economics with the seller are 94.

EXERCISE 5.2

1. If the average marks obtained by the students of sections A and B of B.Com. class in a college are 40 and 30 respectively whereas the number of students in sections A and B are 60 and 40 respectively. Find out the combined mean marks. **[Ans.** $\overline{X}_{12}$ = 36 marks]

2. The mean monthly salary paid to 77 employees in a company was Rs. 78. The mean salary of 32 of them was Rs. 45 and of the other 25 was Rs. 82. What was the mean salary of the remaining? **[Ans.** $\overline{X}_3 = 125.8$]

3. The average rainfall for a week excluding Sunday was 10 cm. Due to heavy rainfall on Sunday, the average for the week rose to 15 cm. How much rainfall was on Sunday?
[Hint: $\overline{X}_{12} = 15,\ \overline{X}_1 = 10, N_1 = 6, N_2 = 1, \overline{X}_2 = ?$] **[Ans.** 45 cm]

4. The mean annual salary paid to all employees of a company was Rs. 5000. The mean annual salaries paid to male and female employees were Rs. 5200 and Rs. 4200 respectively. Determine the percentage of male and female employed by the company. **[Ans.** Male = 80%; Female = 20%]

5. The average marks of 39 students of a class is 50. The marks obtained by 40th student are 39 more than the average marks of all the 40 students. Find the mean marks of all the 40 students.
[Hint: $\overline{X} + 39 + 39 \times 50 = 40\overline{X}$] **[Ans.** $\overline{X} = 51$]

6. The mean of 99 items is 55. The value of 100th item is 99 more than the mean of 100 items. What is the value of 100th item? **[Ans.** $X_{100} = 155$]

7. 50 students took a test. The result of those who passed the test is given below:

Marks:	4	5	6	7	8	9
Students:	8	10	9	6	4	3

If the average marks for all the 50 students were 5.16, find out the average marks of the students who failed. [Ans. $\bar{X}_F = 2.12$]

CORRECTING INCORRECT VALUES OF MEAN

Sometimes, due to mistake in copying, certain items may be wrongly taken in calculating the arithmetic mean. The problem now arises is to find out correct mean. In this case, without calculating the arithmetic mean from the beginning, we can directly calculate the arithmetic mean. The process is very simple. Wrong values are deducted and correct values are added to ΣX. Then correct ΣX is divided by the number of observations. The result, thus, obtained will give us the correct mean.

The following examples would illustrate the procedure:

Example 19. The mean of 100 items is 80. By mistake one item is misread as 92 instead of 29. Find the correct mean.

Solution: We are given, $N = 100$, $\bar{X} = 80$

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

$\therefore$ ΣX (Incorrected) $= 100 \times 80 = 8000$

Corrected $\Sigma X = 8000 +$ Correct item $-$ Incorrect item

$= 8000 + 29 - 92$

$= 7937$

$\therefore$ Corrected Mean $= \frac{\text{Corrected } \Sigma X}{N} = \frac{7937}{100} = 79.37$

Example 20. The mean of 5 observations is 7. Later on, it was found that two observations 4 and 8 were wrongly taken instead of 5 and 9. Find the correct mean.

Solution: We are given, $N = 5$, $\bar{X} = 7$

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

$\therefore$ ΣX (Incorrected) $= 5 \times 7 = 35$

Corrected $\Sigma X = 35 +$ Correct item $-$ Incorrect item

$= 35 + 5 + 9 - 4 - 8$

$= 37$

$\therefore$ Corrected $\bar{X} = \frac{\text{Corrected } \Sigma X}{N} = \frac{37}{5} = 7.4$

EXERCISE 5.3

1. The mean marks of 100 students were found to be 40. Later on it was discovered that a score of 53 was misread as 83. Find the correct mean. [Ans. $\bar{X} = 39.7$]
2. Mean of 100 observations is found to be 40. If at the time of computation two items are wrongly taken as 30 and 27 instead of 3 and 72. Find the correct mean. [Ans. $\bar{X} = 40.18$]
3. The mean of 20 observations is 6.21. Later on it was discovered that two items +5 and +3 were taken as –5 and –3. Find the correct mean. [Ans. $\bar{X} = 7.01$]
4. The average daily price of share of a company from Monday to Friday was Rs. 130. If the highest and lowest price during the week were Rs. 200 and Rs. 100 respectively, find the average daily price when the highest and lowest price are not included. [Ans. $\bar{X} = 116.67$]

MATHEMATICAL PROPERTIES OF ARITHMETIC MEAN

The following are some of the important mathematical properties of the arithmetic mean:

(1) The sum of the deviations of the items from arithmetic mean is always zero. Symbolically,

$$\Sigma(X - \bar{X}) = 0$$

This property is verified from the following example:

X	$\bar{X} = 30$ $X - \bar{X}$
10	$10 - 30 = -20$
20	$20 - 30 = -10$
30	$30 - 30 = 0$
40	$40 - 30 = +10$
50	$50 - 30 = +20$
$\Sigma X = 150$ $N = 5$ $\bar{X} = \frac{150}{5} = 30$	$\Sigma(X - \bar{X}) = 0$

It is clear that when the deviations from the actual mean, *i.e.*, 30 is taken, its sum comes to be zero.

(2) The sum of the squared deviations of the items from arithmetic mean is minimum. That is, it is less than the sum of the squared deviations of the items from any other value. This property is clear from the following example:

X	$\bar{X}=5$ $X-\bar{X}$	$(X-\bar{X})^2$	$A=4$ $X-4$	$(X-4)^2$
3	−2	4	−1	1
4	−1	1	0	0
5	0	0	+1	1
6	+1	1	+2	4
7	+2	4	+3	9
$\Sigma X=25$ $N=5$ $\bar{X}=5$		$\Sigma(X-\bar{X})^2=10$		$\Sigma(X-4)^2=15$

It is clear that the sum of squared deviations taken from the arithmetic mean is 10 whereas the sum of the squared deviations taken from the assumed mean 4 is 15. Therefore,

$$\Sigma(X-\bar{X})^2 < \Sigma(X-A)^2$$

(3) If we have the arithmetic mean and number of items of two groups, then the combined arithmetic mean of these groups can be calculated by using the following formula:

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

The above formula can be extended to calculate the combined arithmetic mean of three or more groups.

(4) If each item of a series is increased, decreased, multiplied or divided by some constant, then A.M. is also increases, decreases, multiplied or divided by the same constant. The following example would clarify this property:

X	$X+2$	$X-2$	$X\times 2$	$X\div 2$
10	12	8	20	5
20	22	18	40	10
30	32	28	60	15
40	42	38	80	20
50	52	48	100	25
$\Sigma X=150$ $N=5$ $\bar{X}=30$	$\Sigma X=160$ $N=5$ $\bar{X}=32$	$\Sigma X=140$ $N=5$ $\bar{X}=28$	$\Sigma X=300$ $N=5$ $\bar{X}=60$	$\Sigma X=75$ $N=5$ $\bar{X}=15$

The above example shows that if 2 is added or subtracted to different items of a series or the items of a series are multiplied or divided by 2, the A.M. will also be affected accordingly.

(5) The product of the arithmetic mean and number of items on which mean is based is equal to the sum of all given items. That is,

As $\bar{X} = \frac{\Sigma X}{N} \Rightarrow \Sigma X = N \cdot \bar{X}$

(6) If each item of the original series is replaced by the actual mean, then the sum of these substitutions will be equal to the sum of the individual items.

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

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

Some Examples Based on Properties of Arithmetic Mean

Example 21. The arithmetic mean $(\bar{X})$ of a series is 15. What is the new mean $\bar{X}$ if each item is increased by 5 and then divided by 3?

Solution: When each item is increased by 5, new $\bar{X} = 15 + 5 = 20$

When each item is divided by 3, new $\bar{X} = \frac{15}{3} = 5$

Example 22. If the average salary of a firm is 400 and the number of workers is 60, find the total salary bill of the firm.

Solution: Given: $\bar{X} = 400, N = 60$

As $\bar{X} = \frac{\Sigma X}{N} \Rightarrow \Sigma X = N \cdot \bar{X}$

Total Salary Bill $= (\Sigma X) = 60 \times 400 =$ Rs. 24,000

Example 23. If the arithmetic mean of series is 28, what will be resultant mean if each item of a series is increased by 3, decreased by 5 or divided by 4 or is multiplied by 10?

Solution: When each item is increased by 3, the mean $\bar{X} = 28 + 3 = 31$

When each item is decreased by 5, the mean $\bar{X} = 28 - 5 = 23$

When each item is multiplied by 10, the mean $\bar{X} = 28 \times 10 = 280$

When each item is divided by 4, the mean $\bar{X} = \frac{28}{4} = 7$

Example 24. Prove that $\Sigma(X - \bar{X}) = 0$

Solution: $\Sigma(X - \bar{X}) = \Sigma X - \Sigma\bar{X} = \Sigma X - N\bar{X}$

$= N\bar{X} - N\bar{X} = 0$ [As $\bar{X} = \frac{\Sigma X}{N} \Rightarrow \Sigma X = N\bar{X}$]

Hence, proved.

Merits and Demerits of Arithmetic Mean

Merits:

(1) It is easy to calculate and simple to understand.

(2) It is based on all observations.

(3) It is capable of further algebraic treatment.

(4) It is rigidly defined.

(5) It is a calculated value and not based on the position of the series.

Demerits:

(1) It is highly affected by extreme values. It is not advised if the series has a few extreme items.

(2) It cannot be calculated in open ended series.

(3) Sometimes, arithmetic mean gives misleading and surprising results such as average number of children born per married couple is 2.3, etc.

(4) It cannot be ascertained graphically.

(5) It cannot be determined in a situation when any value is missing.

Weighted Arithmetic Mean

Simple Arithmetic Mean, as discussed above gives equal importance (or weights) to each item of the series. But there can be some cases where all the items of a series are not of equal importance. In case of Science, Economics or Technology, each subject has its own importance. We have to provide different weights according to their importance and weighted arithmetic mean is used as an average in such cases. The following formula is used to calculate the weighted arithmetic mean:

$$\bar{X}_W = \frac{\Sigma WX}{\Sigma W}, \quad \text{Where, } \bar{X}_W = \text{Weighted A.M.}$$

$$\Sigma W = \text{Sum of weights}$$

$$X = \text{Variables}$$

Note: In weighted A.M., W is taken instead of f.

▶ Steps for Calculation

(*i*) Multiply the values of the items (X) by the weights (W) and obtain the total, *i.e.*, ΣWX.

(*ii*) Now, divide this total, *i.e.*, ΣWX by the sum total of weights, *i.e.*, ΣW. The resultant value would give the weighted A.M.

The following examples would illustrate the computation of weighted A.M.:

Example 25. Calculate weighted mean from the following data:

Items:	81	76	74	58	70	73
Weight:	2	3	6	7	3	7

Solution:

Calculation of Weighted Mean

X	W	WX
81	2	162
76	3	228
74	6	444
58	7	406
70	3	210
73	7	511
	$\Sigma W = 28$	$\Sigma WX = 1961$

$$\therefore \quad \bar{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{1961}{28} = 70.04$$

Example 26. A student obtained 60 marks in English, 75 in Hindi, 63 in Mathematics, 59 in Economics and 55 in Statistics. Calculate the weighted mean of the marks if weights are respectively 2, 1, 5, 5 and 3.

Solution:

Calculation of Weighted Mean

Marks (X)	Weights (W)	WX
60	2	120
75	1	75
63	5	315
59	5	295
55	3	165
	$\Sigma W = 16$	$\Sigma WX = 970$

$$\bar{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{970}{16} = 60.63$$

Example 27. A train runs 25 miles at a speed of 30 *mph*, another 50 miles at a speed of 40 *mph*, then due to repairs of the track travels for 6 minutes at a speed of 10 *mph* and finally covers the remaining distance of 24 miles at a speed of 24 *mph*. What is the average speed in miles per hour?

Solution: Time taken in covering 25 miles at a speed of 30 *mph* = 50 minutes

Time taken in covering 50 miles at a speed of 40 *mph* = 75 minutes

Distance covered in 6 minutes at a speed of 10 *mph* = 1 mile

Time taken in covering 24 miles at speed of 24 *mph* = 60 minutes.

Therefore, taking the time taken as weights we have the weighted mean as

Speed in *mph* (X)	Time Taken (W)	WX
30	50	1,500
40	75	3,000
10	6	60
24	60	1,440
	$\Sigma W = 191$	$\Sigma WX = 6,000$

$$\therefore \quad \text{Average speed} = \frac{6,000}{191} = 31.41 \text{ } mph$$

Example 28. Comment on the performance of the students of the three universities given below using simple and weighted averages:

University	Bombay		Calcutta		Madras	
Course of Study	Pass %	No. of students (in hundreds)	Pass %	No. of students (in hundreds)	Pass %	No. of students (in hundreds)
M.A.	71	3	82	2	81	2
M.Com.	83	4	76	3	76	3.5
B.A	73	5	73	6	74	4.5
B.Com.	74	2	76	7	58	2
B.Sc.	65	3	65	3	70	7
M.Sc.	66	3	60	7	73	2

Solution:

Calculation of Simple and Weighted Arithmetic Means

University	Bombay			Calcutta			Madras		
Course of Study	Pass %	No.of students (in hundreds)		Pass %	No.of students (in hundreds)		Pass %	No.of students (in hundreds)	
	X	W	WX	X	W	WX	X	W	WX
M.A.	71	3	213	82	2	164	81	2	162
M.Com.	83	4	332	76	3	228	76	3.5	266
B.A	73	5	365	73	6	438	74	4.5	333
B.Com.	74	2	148	76	7	532	58	2	116
B.Sc.	65	3	195	65	3	195	70	7	490
M.Sc.	66	3	198	60	7	420	73	2	146
	ΣX =432	ΣW =20	ΣWX =1,451	ΣX =432	ΣW =28	ΣWX =1,977	ΣX =432	ΣW =21	ΣWX =1,513

Simple and Weighted Arithmetic Means

Bombay : $\bar{X} = \frac{\Sigma X}{N} = \frac{432}{6} = 72;\ \bar{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{1,451}{20} = 72.55$

Calcutta : $\bar{X} = \frac{\Sigma X}{N} = \frac{432}{6} = 72;\ \bar{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{1,977}{28} = 70.61$

Madras : $\bar{X} = \frac{\Sigma X}{N} = \frac{432}{6} = 72;\ \bar{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{1,513}{21} = 72.05$

The simple arithmetic mean is the same for all the three universities, *i.e.*, 72 and hence, it may be concluded that the performance of students is alike. But this will be a wrong conclusion because what we should compare here is the weighted arithmetic mean. On comparing the weighted arithmetic means, we find that for Bombay the mean value is the highest and hence, we can say that in Bombay University the performance of students is best.

EXERCISE 5.4

1. A housewife uses 10 kg of Wheat, 20 kg of Fuel, 5 kg. of Sugar and 2 kg. of Oil. Price (per kg.) of these items are respectively Rs. 1.50, Rs. 0.50, Rs. 2.80 and Rs. 10. Taking quantities used as weights, find out weighted arithmetic average of the prices. **[Ans.** $\bar{X}_w$ = Rs. 1.59 per kg.]

2. Calculate the simple and weighted arithmetic mean price per tonne of coal purchased by an industry for the half year:

Months:	Jan.	Feb.	Mar.	Apr.	May	June
Price per tonne (Rs.)	42.50	51.25	50.00	52.00	44.25	54.00
Tonnes Purchased:	25	30	40	50	10	45

[Hint: Taking quantity purchased as weight]

[Ans. $\bar{X} = 49$, $\bar{X}w = 50.363$]

3. From the following results of Universities A and B, which is better?

Class	University A		University B	
	Appeared	Passed	Appeared	Passed
M.A.	100	90	240	200
M.Com.	60	45	200	160
B.A.	120	75	160	60
B.Com.	200	150	200	140
Total	480	360	800	560

[Hint: Taking number of students appeared as weights and pass percentage of each as X]

[Ans. $\bar{X}_w$(Univ. A) = 75%, $\bar{X}_w$(Univ. B) = 70%; University A is better]

(2) MEDIAN

Median is another important measure of central tendency. It is a positional average. Median is defined as the middle value of the series when arranged either in ascending or descending order. It is a value which divides the arranged series into two equal parts in such a way that the number of observations smaller than the median is equal to the number greater than it. Median is thus, a positional average. In the words of **Connor, "The median is that value of the variable which devides the group into two equal parts, one part comprising all values greater and the other values less than the median."** A point to be noted is that median is always determined by first arranging the series in an ascending or descending manner. Median is denoted by the symbol 'M'.

Calculation of Median

Individual Series

The formula used for calculating median in individual series is:

$$M = \text{Size of} \left(\frac{N+1}{2}\right)\text{th item.}$$

Where, M = median, N = total number of items in the series.

Steps for Calculation

(*i*) Arrange the data in ascending or descending order of the size.

(*ii*) Locate the median item by using the formula $\frac{N+1}{2}$.

(*iii*) The value or size of this item is the median.

Odd Number Series

If the number of items is odd, then the median is the middle value after the items are arranged in ascending or descending order of their magnitude.

Example 29. Calculate median from the following data:

22, 16, 18, 13, 15, 19, 17, 20, 23

Solution: The data is first arranged in ascending order

Sr. No.	Items (X)
1	13
2	15
3	16
4	17
5	18
6	19
7	20
8	22
9	23
$N = 9$	

$$M = \text{Size of}\left(\frac{N+1}{2}\right)\text{th item.}$$

Here, $N = 9$

$$\therefore \quad M = \text{Size of}\left(\frac{9+1}{2}\right)\text{th item.}$$

$$= \text{Size of 5th item} = 18$$

Hence, $M = 18$

Even Number Series

In case of even number of items, median is arithmetic median of two middle values after the items are arranged in ascending or descending order of their magnitude.

Example 30. Calculate median from the following data:

200, 217, 316, 264, 296, 282, 317, 299

Solution: The data is first arranged in ascending order

Sr. No.	Items (X)
1	200
2	217
3	264
4	282
5	296
6	299
7	316
8	317
$N = 8$	

$$M = \text{Size of}\left(\frac{N+1}{2}\right)\text{th item.}$$

Here, $N = 8$

$$\therefore \quad M = \text{Size of}\left(\frac{8+1}{2}\right)\text{th item.}$$

$$= \text{Size of 4.5th item}$$

$$= \frac{\text{Size of 4th items} + \text{Size of 5th item}}{2}$$

$$\therefore \quad M = \frac{282+296}{2} = \frac{578}{2} = 289$$

Hence, $M = 289$.

Example 31. Following are the marks obtained by a batch of 10 students in a certain class test in Statistics (X) and Accountancy (Y):

Roll No.	1	2	3	4	5	6	7	8	9	10
X:	63	64	62	32	30	60	47	46	35	28
Y:	68	66	35	42	26	85	44	80	33	72

In which subject is the level of knowledge of students higher?

Solution: **Calculation of Median**

X:	28	30	32	35	46	47	60	62	63	64
Y:	26	33	35	42	44	66	68	72	80	85

$$\text{Median} = \text{Size of}\left(\frac{N+1}{2}\right)\text{th item.}$$

$$= \frac{10+1}{2}\text{th item} = \frac{11}{2}\text{th} = 5.5\text{th item.}$$

$$\text{Median (X)} = \text{Size of}\left(\frac{5\text{th}+6\text{th}}{2}\right)\text{item} = \frac{46+47}{2} = 46.5 \text{ marks.}$$

$$\text{Median (Y)} = \text{Size of}\left(\frac{5\text{th}+6\text{th}}{2}\right)\text{item} = \frac{44+66}{2} = 55 \text{ marks.}$$

Median marks of Accountancy (Y) are more than that of statistics (X). Therefore, level of knowledge of students is higher in Accountancy.

Discrete Series

The formula used for calculating median in discrete series is

$$\mathbf{M = Size\ of\left(\frac{N+1}{2}\right)th\ item.}$$

▶ **Steps for Calculation**

(*i*) Arrange the data in ascending or descending order of size.

(*ii*) Then find the cumulative frequency column.

(*iii*) Apply the formula,

$$M = \text{Size of}\left(\frac{N+1}{2}\right)\text{th item.}$$

(*iv*) Now locate $\frac{N+1}{2}$th item in the cumulative frequency column. It is done by comparing $\frac{N+1}{2}$ with the cumulative frequency at each stage. The value of the variable is the value of the median.

Example 32. Calculate the median from the following data:

X:	10	12	14	16	18	20	22
f:	2	5	12	20	10	7	3

Solution:

Calculation of Median

X	f	*c.f.*
10	2	2
12	5	7
14	12	19
16	20	39 M
18	10	49
20	7	56
22	3	59
	$N = 59$	

$$M = \text{Size of}\left(\frac{N+1}{2}\right)\text{th item.}$$

$$= \text{Size of}\left(\frac{59+1}{2}\right)\text{th item.}$$

$$= \text{Size of 30th item.}$$

The value of 30th item lies against 39 whose value is 16.

Hence, $M = 16$.

Continuous Series

The formula used for calculating median in continuous series is

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

Where, l_1 = lower limit of the median class; *c.f.* = cumulative frequency of the class preceeding the median class; f = frequency of the median class; i = size of class interval of the median class.

▶ **Steps for Calculation**

(*i*) Firstly, calculate cumulative frequency.

(*ii*) Then, find out the median size by using the formula: $\frac{N}{2}$

(*iii*) Determine the median class in which median lies.

(*iv*) Substitute the values in the above formula.

Example 33. Calculate median from the following data:

Marks:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40	40—45
No. of students:	6	12	17	30	10	10	8	5	2

Solution:

Marks	f	*c.f.*
0—5	6	6
5—10	12	18
10—15	17	35 *c.f.*
15—20	30 f	65 M
20—25	10	75
25—30	10	85
30—35	8	93
35—40	5	98
40—45	2	100
	N = 100	

Median item = Size of $\left(\frac{N}{2}\right)$th item = $\frac{100}{2}$th = 50th item.

50th item lies in class 15—20. Hence median class is 15–20.

Applying the formula,

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 15 + \frac{50-35}{30} \times 5$$

$$= 15 + \frac{15}{30} \times 5 = 15 + 2.5 = 17.5$$

Hence, $M = 17.5$.

Cumulative Frequency Series

Example 34. Calculate median from the following data:

Value	Frequency
Less than 10	4
Less than 20	16
Less than 30	40
Less than 40	76
Less than 50	96
Less than 60	112
Less than 70	120
Less than 80	125

Solution: Since we are given cumulative frequencies, firstly we find simple frequencies.

Calculation of Median

Value	f	c.f.
0—10	4	4
10—20	12	16
20—30	24	40
30—40	36	76
40—50	20	96
50—60	16	112
60—70	8	120
70—80	5	125
	N = 125	

Median item = Size of $\left(\frac{N}{2}\right)$th item $= \frac{125}{2} = 62.5$th item.

$\therefore$ Median lies in the class 30—40

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 30 + \frac{62.5 - 40}{36} \times 10$$

$$= 30 + \frac{22.5}{36} \times 10$$

$$= 30 + 6.25$$

$$= 36.25$$

Inclusive Series

Example 35. Calculate median from the following data:

Value:	1—10	11—20	21—30	31—40	41—50
Frequency:	4	12	20	9	5

Solution: Since we are given inclusive series, firstly we convert it into exclusive one by deducting 0.5 from the lower limits and adding 0.5 to the upper limits.

Calculation of Median

Value	f	c.f.
0.5—10.5	4	4
10.5—20.5	12	16
20.5—30.5	20	36
30.5—40.5	9	45
40.5—50.5	5	50
	$N = 50$	

Median item = Size of $\left(\frac{N}{2}\right)$th item $= \frac{50}{2}$th = 25th item.

$\therefore$ Median lies in the class 20.5—30.5.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20.5 + \frac{25 - 16}{20} \times 10 = 20.5 + \frac{9}{20} \times 10 = 20.5 + 4.5 = 25$$

$\therefore$ $M = 25$.

Unequal Class Interval

Example 36. Amend the following table and calculate the median from the amended table:

Size :	10—15	15—17.5	17.5—20	20—30	30—35	35—40	40 and onwards
f :	10	15	17	25	28	30	40

Solution: Since the class intervals are unequal, let us first convert it into a series with equal class intervals by adjusting the frequencies correspondingly.

Calculation of Median

Size	f	c.f.
10—20	10+15+17 = 42	42
20—30	25	67
30—40	28+30 = 58	125
40 and onwards	40	165
	$N = 165$	

$$\text{Median item} = \text{Size of}\left(\frac{N}{2}\right)\text{th item} = \frac{165}{2} = 82.50 \text{ item.}$$

$\therefore$ Median lies in class 30—40

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 30 + \frac{82.50 - 67}{58} \times 10$$

$$= 30 + 2.67 = 32.67$$

$$\therefore \quad M = 32.67$$

Descending Class Intervals

Example 37. Find the median from the distribution of marks obtained in Economics:

Marks:	30—35	25—30	20—25	15—20	10—15	5—10	0—5
Number of students:	4	8	12	16	10	6	4

Solution: **Ist Method.** Converting the decending order series into ascending order series.

Marks	f	$c.f.$
0—5	4	4
5—10	6	10
10—15	10	20 $c.f.$
15—20	16 f	36 M
20—25	12	48
25—30	8	56
30—35	4	60
	N=60	

$$\text{Median item} = \frac{N}{2} = \frac{60}{2} = 30\text{th item.}$$

Median lies in the class 15–20.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 15 + \frac{30 - 20}{16} \times 5$$

$$= 15 + \frac{50}{16} = 15 + 3.125 = 18.125$$

$$\therefore \quad \text{Median} = 18.125$$

2nd Method. Median can also be calculated by keeping the series in descending order. The formula used is:

$$M = l_2 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

Where, l_2= upper limit of the median class.

Marks	f	$c.f.$
30—35	4	4
25—30	8	12
20—25	12	24 $c.f.$
15—20	16 f	40 M
10—15	10	50
5—10	6	56
0—5	4	60
	N=60	

$$\text{Median item} = \frac{N}{2} = \frac{60}{2} = 30\text{th item.}$$

Median lies in the class 15–20.

Applying the modified formula:

$$M = l_2 - \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 - \frac{30 - 24}{16} \times 5$$

$$= 20 - \frac{6}{16} \times 5 = 20 - \frac{30}{16} = 20 - 1.875 = 18.125$$

$$\therefore \quad \text{Median} = 18.125$$

Note: Students find that the answer of median will however, remain the same as is computed in ascending or descending order.

Mid-value Series

Example 38. Compute median from the following data:

Mid-value:	5	15	25	35	45	55	65	75
Frequency:	15	7	11	10	13	8	20	16

Solution: Since, we are given the mid-values, we should first find out the upper and lower limits of the various classes. As the mid-values are 5, 15, 25, 35, 45, 55, 65, 75, so, class size is 10. For determining limits of different classes, applying the formula.

$l_1 = m - \frac{i}{2}$ and $l_2 = m + \frac{i}{2}$, where, i = difference between two mid-values.

In this case, the upper and the lower limits of the first mid-value are:

$l_1 = 5 - \frac{10}{2} = 0$ and $l_2 = 5 + \frac{10}{2} = 10$, *i.e.*, 0–10 class interval.

Similarly class intervals for other mid-values are obtained as:

Classes	f	*c.f.*
0—10	15	15
10—20	7	22
20—30	11	33
30—40	10	43 *c.f.*
40—50	13 f	56 M
50—60	8	64
60—70	20	84
70—80	16	100
	$N = 100$	

Median item = Size of $\left(\frac{N}{2}\right)$th item $= \frac{100}{2} = 50$th item.

$\therefore$ Median lies in the class 40—50.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 40 + \frac{50 - 43}{13} \times 10$$

$$= 40 + 5.385 = 45.385$$

IMPORTANT TYPICAL EXAMPLES

▶ To Locate Missing Frequency

Example 39. Find the missing frequency in the following distribution if $N = 100$ and $M = 30$.

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	10	?	25	30	?	10

Solution: Two frequencies are missing and let the missing frequencies be denoted by f_1 and f_2. We need two equations, we will get one from summation of frequencies and one from median formula.

Marks	f	*c.f.*
0—10	10	10
10—20	f_1	$10 + f_1$
20—30*	25	$35 + f_1$
30—40	30	$65 + f_1$
40—50	f_2	$65 + f_1 + f_2$
50—60	10	$75 + f_1 + f_2$
	$N = 100$	

1st Equation: From summation of frequencies,

$$75 + f_1 + f_2 = 100 \Rightarrow f_1 + f_2 = 25 \qquad ...(i)$$

2nd Equation: From information regarding median,

$M = 30$, Median class is 30–40.

Now, $$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$30 = 30 + \frac{50 - (35 + f_1)}{30} \times 10$$

$$0 = \frac{15 - f_1}{3}$$

or $$f_1 = 15 \qquad ...(ii)$$

Substituting $f_1 = 15$ in equation (*i*),

$\therefore$ $f_1 + f_2 = 25$

Put $f_1 = 15$

$\therefore$ $15 + f_2 = 25$

$\therefore$ $f_2 = 25 - 10 = 15$

Thus, $f_1 = 15$, $f_2 = 10$

Example 40. The following table gives distribution of marks secured by some students:

Marks:	0—20	20—30	30—40	40—50	50—60	60—70	70—80
No. of students:	42	38	120	84	48	36	31

Find (*i*) Median marks (*ii*) The percentage of failure if minimum for a pass is 35 marks.

Solution: (*i*) **Calculation of Median**

Marks	f	*c.f.*
0—20	42	42
20—30	38	80
30—40	120	200

* [illegible] 30 would [illegible] and 30 would be covered in class 30—40

40—50	84	284
50—60	48	332
60—70	36	368
70—80	31	399
	$N = 399$	

Median item = Size of $\frac{N}{2}$th item $= \frac{399}{2} = 199.5$th item.

Hence, median lies in the class 30—40.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 30 + \frac{199.5 - 80}{120} \times 10$$

$$= 30 + 9.96 = 39.96 \approx 40$$

(*ii*) **Percentage of failure if the minimum for a pass is 35 marks.**

Under the assumption that observations in a class are uniformly distributed, the number of students getting less than 35 marks are:

$$= 42 + 38 + \left(\frac{35 - 30}{10}\right) \times 120$$

$$= 42 + 38 + 60 = 140$$

$\therefore$ Percentage of failure $= \frac{140}{399} \times 100 = 35.1$

Thus, the percentage of failure for a minimum pass marks is 35.1.

Example 41. The median of a few number of observations is given to be 48.67. The six items whose values are 33.5, 38.9, 45.57, 49.03, 53.43 and 59.95 were added to given series, what will be the new median?

Solution: $M = 48.67$ (given)

Number of items less than median = 3

Number of items more than median = 3

Since, three items are less than given median and the same number of items are more than median, hence the position of the given median, *i.e.*, 48.67 remains unchanged.

Graphical Location of Median

Median can also be located graphically with the help of ogive curve (or cumulative frequency curve). Following are the steps involved in it:

(*i*) Firstly draw an ogive curve either 'less than ogive' or 'more than ogive'.

(*ii*) Compute $\frac{N}{2}$ and locate it (median item) on Y-axis.

(*iii*) Now draw a horizontal line from this point on Y-axis so as to meet at ogive curve.

(*iv*) Then draw a perpendicular line from the point of intersection on X-axis.

(*v*) The point of intersection on X-axis gives the value of median.

Aliter

Median can also be located by making use of both the ogive curves together. The point where 'less than ogive curve' and 'more than ogive curve' meet will give us the value of median.

Example 42. Locate the median graphically from the following data:

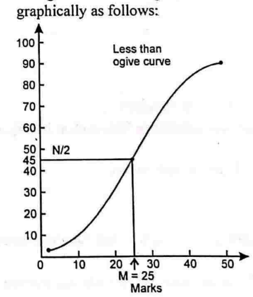
Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	10	20	30	20	10

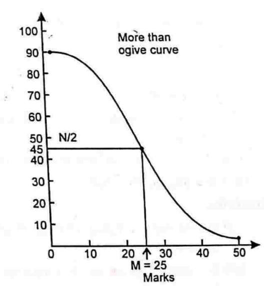
Solution: Firstly we convert the given distribution into "less than" and "more than" cumulative frequencies distribution.

Less than method	
Marks	***c.f.***
Less than 10	10
Less than 20	30
Less than 30	60
Less than 40	80
Less than 50	90

More than method	
Marks	***c.f.***
More than 0	90
More than 10	80
More than 20	60
More than 30	30
More than 40	10

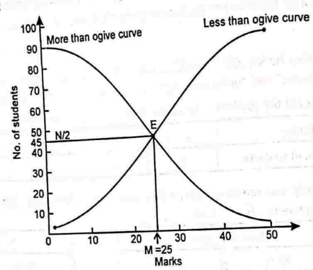
Method I: Using the "less than ogive" and "more than ogive" method, the median is computed graphically as follows:





M = Size of $\frac{N}{2}$th item $= \frac{90}{2} = 45$th item. It lies in the class interval 20—30.

Method II: By making use of both the ogive curves together, the median is computed graphically as follows:



It is clear from the above figure that less than ogive curve and more than ogive curve meet each other at point E. The value of median is 25.

An Important Property

An important property of median is that the sum of the absolute deviations of the items from the median is less than the sum from any other value or average.

Merits and Demerits of Median

Merits:

(*i*) It is easy to understand.
(*ii*) It is easy to locate or compute.
(*iii*) It is not affected by extreme items.
(*iv*) Median can be located graphically with the help of ogives.
(*v*) It is the most suitable average in dealing with qualitative facts such as beauty, intelligence, honesty, etc.
(*vi*) It is most appropriate average in case of open ended classes.
(*vii*) It is rigidly defined.

Demerits:

(*i*) It requires arranging of data in ascending or descending order but other averages do not need this.
(*ii*) It is not based on all observations of the series, since it is a positional average.
(*iii*) It is not capable of further algebraic treatment like arithmetic mean.
(*iv*) It cannot be computed exactly where the number of items in a series is even.
(*v*) It is very difficult to calculate if the number of items is very small or large.

EXERCISE 5.5

1. Calculate the median of the following items:
25, 20, 15, 45, 18, 7, 10, 38, 12
[Ans. $M = 18$]

2. Calculate the median of the following items:
15, 20, 20, 23, 23, 25, 26, 27, 35, 40
[Ans. $M = 24$]

3. Calculate the median from the following data:

X:	15	20	25	30	35	40
f:	10	15	25	5	5	20

[Ans. $M = 25$]

4. Find out median marks from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	8	30	40	12	10

[Ans. $M = 23$]

5. Calculate the median from the following data:

Weight in gms.:	410—419	420—429	430—439	440—449	450—459	460—469	470—479
No. of apples:	14	20	42	54	45	18	7

[Ans. $M = 443.94$]

6. Find out median from the following table:

Monthly Wages (in Rs.):	50—80	80—100	100—110	110—120	120—130	130—150	150—180	180—200
No. of workers:	30	127	140	240	176	135	20	3

A fund is to be raised and it is decided that workers getting less than Rs. 120 should contribute 5% of their wages and those getting more than Rs. 120 should contribute 10% of their wages. What sum will be collected? [Ans. $M = 115.77$, Total Fund = 7261]

7. Following is the distribution of marks in statistics obtained by 50 students:

Mid-value:	5	15	25	35	45	55
Frequency :	4	6	10	7	3	2

Calculate the median marks. [Ans. $M = 26$]

8. Find the missing frequency of the group 20–30 when the median is 24.

Sale (in '000 Rs.)	0—10	10—20	20—30	30—40	40—50
No. of shops	5	25	—	18	7

[Ans. Missing frequency is 25]

9. The median of the following incomplete table is 46. Find the missing frequencies and calculate the arithmetic mean of the completed table:

Class interval:	10—20	20—30	30—40	40—50	50—60	60—70	70—80	Total
Frequency:	12	30	?	65	?	25	18	229

[**Ans.** $f_1 = 33 \cdot 5 \simeq 34$, $f_2 = 45$ and $\overline{X} = 45.82$]

10. Calculate median from the following data:

Marks less than:	80	70	60	50	40	30	20	10
No. of students:	100	90	80	60	32	20	13	5

[**Ans.** $M = 46.43$]

11. Below are given the marks obtained by 65 students in an examination. Find the median marks:

Marks more than:	70%	60%	50%	40%	30%	20%
No. of students:	7	18	40	40	63	65

[**Ans.** $M = 53.4$]

12. The following table gives the frequency distribution of the marks of 400 candidates in an examination:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90	90—100
No.of Candidates:	5	20	40	70	85	65	50	35	20	10

Find (*i*) Median marks, (*ii*) The percentage of pass if the minimum pass marks are 35.

[**Ans.** M = 47.64, % of Pass = 75%]

13. Find the median income graphically from the following data. Also verify the result algebraically.

Marks:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
No. of students:	4	6	10	10	25	22	18	5

[**Ans.** $M = 24$]

(3) PARTITION VALUES—QUARTILES, DECILES AND PERCENTILES

Just as median divides the series into two equal parts, there are other useful measures which divides the series into 4, 10 or 100 equal parts. They are called quartiles, deciles and percentiles.

(1) Quartiles: Quartiles divide a series into 4 equal parts. For any series there are three quartiles denoted by Q_1, Q_2 and Q_3. Q_1 is known as first or lower quartile, covering 25% items. The second quartile or Q_2 is the same as Median of the series. Q_3 is called third or upper quartile, covering 75% items.

(2) Deciles: Deciles divide a series into 10 equal parts. For any series, there are 9 deciles denoted by $D_1, D_2 \ldots D_9$. These are called as first decile, second decile and so on.

(3) Percentiles: Percentiles divide a series into 100 equal parts. For any series, there are 99 percentiles denoted by $P_1, P_2, P_3 \ldots P_{99}$.

Calculation of Quartiles, Deciles and Percentiles

The calculation of quartiles, deciles and percentiles is done in the same manner as the calculation of median.

For Individual and Discrete Series	For Continuous Series	Formula to be used in continuous series:
$Q_1 = \text{Size of } \frac{N+1}{4}\text{th item}$	$Q_1 = \text{Size of } \frac{N}{4}\text{th item}$	$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$
$Q_3 = \text{Size of } \frac{3(N+1)}{4}\text{th item}$	$Q_3 = \text{Size of } \frac{3N}{4}\text{th item}$	$Q_3 = l_1 + \frac{\frac{3}{4}N - c.f.}{f} \times i$
$D_1 = \text{Size of } \frac{N+1}{10}\text{th item}$	$D_1 = \text{Size of } \frac{N}{10}\text{th item}$	$D_1 = l_1 + \frac{\frac{N}{10} - c.f.}{f} \times i$
$D_9 = \text{Size of } \frac{9(N+1)}{10}\text{th item}$	$D_9 = \text{Size of } \frac{9N}{10}\text{th item}$	$D_9 = l_1 + \frac{\frac{9}{10}N - c.f.}{f} \times i$
$P_1 = \text{Size of } \frac{N+1}{100}\text{th item}$	$P_1 = \text{Size of } \frac{N}{100}\text{th item}$	$P_1 = l_1 + \frac{\frac{N}{100} - c.f.}{f} \times i$
$P_{99} = \text{Size of } \frac{99(N+1)}{100}\text{th item}$	$P_{99} = \text{Size of } \frac{99}{100}N\text{th item}$	$P_{99} = l_1 + \frac{\frac{99}{100}N - c.f.}{f}$

Individual Series

Example 43. From the following data, calculate Q_1, Q_3, D_5 and P_{25}

21, 15, 40, 30, 26, 45, 50, 54, 60, 65, 70

Solution: The data is first arranged in ascending order:

Sr. No.	X
1	15
2	21
3	26
4	30
5	40
6	45
7	50
8	54
9	60
10	65
11	70
$N = 11$	

$Q_1 = \text{Size of}\left(\frac{N+1}{4}\right)\text{th item} = \text{Size of}\left(\frac{11+1}{4}\right)\text{th item} = \text{Size of 3rd item} = 26$

Thus, $Q_1 = 26$

$Q_3 = \text{Size of}\frac{3(N+1)}{4}\text{th item} = \text{Size of}\frac{3(11+1)}{4}\text{th item} = \text{Size of 9th item} = 60$

Thus, $Q_3 = 60$

$D_5 = \text{Size of}\frac{5(N+1)}{10}\text{th item} = \text{Size of}\frac{5(11+1)}{10}\text{th item}$

$= \text{Size of 6th item} = 45$

Thus, $D_5 = 45$

$P_{25} = \text{Size of}\frac{25(N+1)}{100}\text{th item} = \text{Size of}\frac{25(11+1)}{100}\text{th item}$

$= \text{Size of 3rd item} = 26$

Thus, $P_{25} = 26$

Example 44. Calculate Q_1, Q_3, D_9 and P_{70} from the following data:

120, 150, 170, 180, 181, 187, 190, 192, 200, 210

Solution: The data is first arranged in ascending order:

Sr. No.	X
1	120
2	150
3	170
4	180
5	181
6	187
7	190
8	192
9	200
10	210
$N = 10$	

$Q_1 = \text{Size of}\frac{N+1}{4}\text{th item} = \left(\frac{10+1}{4}\right)\text{th item}$

$= \text{Size of 2.75th item.}$

$= \text{Size of 2nd item} + 0.75\ (\text{Size of 3rd item} - \text{Size of 2nd item})$

$= 150 + \frac{3}{4}\ (170 - 150) = 150 + 15 = 165$

$Q_3 = \text{Size of }\frac{3(N+1)}{4}\text{th item} = \frac{3(10+1)}{4}$

$= \text{Size of 8.25th item.}$

$= \text{Size of 8th item} + 0.25\ (\text{Size of 9th item} - \text{8th item})$

$= 192 + \frac{1}{4}\ (200 - 192) = 192 + 2 = 194.$

$D_9 = \text{Size of }\frac{9(N+1)}{10}\text{th item} = \frac{9(10+1)}{10} = \text{Size of 9.9 item.}$

$= \text{Size of 9th item} + 0.9\ (\text{Size of 10th item} - \text{9th item})$

$= 200 + \frac{9}{10}(210 - 200)$

$= 200 + 9 = 209$

$P_{70} = \text{Size of }\frac{70(N+1)}{100}\text{th item} = \frac{70(10+1)}{100} = \text{Size of 7.7th item}$

$= \text{Size of 7th item} + 0.7\ (\text{Size of 8th item} - \text{Size of 7th item})$

$= 190 + \frac{7}{10}(192 - 190) = 190 + 1.40 = 191.40$

Thus, $Q_1 = 165, Q_3 = 194, D_9 = 209, P_{70} = 191.40$

Discrete Series

Example 45. Calculate Q_1, Q_3, D_6 and P_{85} from the following data:

X:	10	11	12	13	14	15	16	17	18
f:	3	4	5	12	10	7	5	2	1

Solution:

X	f	c.f.
10	3	3
11	4	7
12	5	12
13	12	24
14	10	34
15	7	41
16	5	46
17	2	48
18	1	49
	N = 49	

$Q_1 = \text{Size of}\frac{N+1}{4}\text{th item} = \frac{49+1}{4} = \text{Size of 12.5th item} = 13$

$$Q_1 = \text{Size of}\left(\frac{N+1}{4}\right)\text{th item} = \text{Size of}\left(\frac{11+1}{4}\right)\text{th item} = \text{Size of 3rd item} = 26$$

Thus, $Q_1 = 26$

$$Q_3 = \text{Size of}\frac{3(N+1)}{4}\text{th item} = \text{Size of}\frac{3(11+1)}{4}\text{th item} = \text{Size of 9th item} = 60$$

Thus, $Q_3 = 60$

$$D_5 = \text{Size of}\frac{5(N+1)}{10}\text{th item} = \text{Size of}\frac{5(11+1)}{10}\text{th item}$$

$= \text{Size of 6th item} = 45$

Thus, $D_5 = 45$

$$P_{25} = \text{Size of}\frac{25(N+1)}{100}\text{th item} = \text{Size of}\frac{25(11+1)}{100}\text{th item}$$

$= \text{Size of 3rd item} = 26$

Thus, $P_{25} = 26$

Example 44. Calculate Q_1, Q_3, D_9 and P_{70} from the following data:

120, 150, 170, 180, 181, 187, 190, 192, 200, 210

Solution: The data is first arranged in ascending order:

Sr. No.	X
1	120
2	150
3	170
4	180
5	181
6	187
7	190
8	192
9	200
10	210
$N = 10$	

$$Q_1 = \text{Size of}\frac{N+1}{4}\text{th item} = \left(\frac{10+1}{4}\right)\text{th item}$$

= Size of 2.75th item.

= Size of 2nd item + 0.75 (Size of 3rd item – Size of 2nd item)

$$= 150 + \frac{3}{4}(170 - 150) = 150 + 15 = 165$$

$$Q_3 = \text{Size of }\frac{3(N+1)}{4}\text{th item} = \frac{3(10+1)}{4}$$

= Size of 8.25th item.

= Size of 8th item + 0.25 (Size of 9th item – 8th item)

$$= 192 + \frac{1}{4}(200 - 192) = 192 + 2 = 194.$$

$$D_9 = \text{Size of }\frac{9(N+1)}{10}\text{th item} = \frac{9(10+1)}{10} = \text{Size of 9.9 item.}$$

= Size of 9th item + 0.9 (Size of 10th item – 9th item)

$$= 200 + \frac{9}{10}(210 - 200)$$

$$= 200 + 9 = 209$$

$$P_{70} = \text{Size of }\frac{70(N+1)}{100}\text{th item} = \frac{70(10+1)}{100} = \text{Size of 7.7th item}$$

= Size of 7th item + 0.7 (Size of 8th item – Size of 7th item)

$$= 190 + \frac{7}{10}(192 - 190) = 190 + 1.40 = 191.40$$

Thus, $Q_1 = 165, Q_3 = 194, D_9 = 209, P_{70} = 191.40$

Discrete Series

Example 45. Calculate Q_1, Q_3, D_6 and P_{85} from the following data:

X:	10	11	12	13	14	15	16	17	18
f:	3	4	5	12	10	7	5	2	1

Solution:

X	f	c.f.
10	3	3
11	4	7
12	5	12
13	12	24
14	10	34
15	7	41
16	5	46
17	2	48
18	1	49
	N = 49	

$$Q_1 = \text{Size of}\frac{N+1}{4}\text{th item} = \frac{49+1}{4} = \text{Size of 12.5th item} = 13$$

$$Q_3 = \text{Size of } \frac{3(N+1)}{4}\text{th item} = \frac{3(49+1)}{4} = \text{Size of 37.5th item} = 15$$

$$D_6 = \text{Size of } \frac{6(N+1)}{10}\text{th item} = \frac{6(49+1)}{10} = \text{Size of 30th item} = 14$$

$$P_{85} = \text{Size of } \frac{85(N+1)}{100}\text{th item} = \frac{85(49+1)}{100} = \text{Size of 42.5th item} = 16$$

Thus, $Q_1 = 13$, $Q_3 = 15$, $D_6 = 14$, $P_{85} = 16$

Continuous Series

Example 46. Calculate the values of Q_1, Q_3, D_8 and P_{56} from the following data:

Wages:	0—10	10—20	20—30	30—40	40—50
No. of workers:	22	38	46	35	19

Solution: Calculation of Q_1, Q_3, D_8 and P_{56}

Wages	f	c.f.
0—10	22	22
10—20	38	60
20—30	46	106
30—40	35	141
40—50	19	160
	N = 160	

$$Q_1 = \text{Size of } \frac{N}{4}\text{th item} = \text{Size of } \frac{160}{4}\text{th item} = \text{Size of 40th item.}$$

Q_1 lies in the class 10—20

Thus, $$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

$$= 10 + \frac{40-22}{38} \times 10 = 10 + 4.74 = 14.74$$

$\therefore$ $Q_1 = 14.74$

$$Q_3 = \text{Size of } \frac{3N}{4}\text{th item} = \frac{3(160)}{4}\text{th item} = \text{Size of 120th item}$$

Q_3 lies in the class 30—40.

Thus, $$Q_3 = l_1 + \frac{\frac{3}{4}N - c.f.}{f} \times i$$

$$= 30 + \frac{120-106}{35} \times 10 = 34$$

$\therefore$ $Q_3 = 34$

$$D_8 = \text{Size of } \frac{8N}{10}\text{th item} = \text{Size of } \frac{8\,(160)}{10}\text{ th item}$$

$$= \text{Size of 128th item}$$

D_8 lies in the class 30—40

Thus, $$D_8 = l_1 + \frac{\frac{8N}{10} - c.f.}{f} \times i$$

$$= 30 + \frac{128-106}{35} \times 10 = 36.29$$

$\therefore$ $D_8 = 36.29$

$$P_{56} = \text{Size of } \frac{56N}{100}\text{th item} = \frac{56(160)}{100}\text{th item}$$

$$= \text{Size of 89.6th item.}$$

P_{56} lies in the class 20—30.

Thus, $$P_{56} = l_1 + \frac{\frac{56}{100}N - c.f.}{f} \times i$$

$$= 20 + \frac{89.6-60}{46} \times 10 = 26.43$$

$\therefore$ $P_{56} = 26.43$

Example 47. Calculate Median, Quartiles, 6th decile and 70th percentile from the following data:

Marks less than:	80	70	60	50	40	30	20	10
No. of students:	100	90	80	60	32	20	13	5

Solution: The data is given in the form of a cumulative frequency distribution. First we convert it into simple frequency distribution and write it in ascending order:

Marks	f	c.f.
0—10	5	5
10—20	8	13
20—30	7	20
30—40	12	32
40—50	28	60
50—60	20	80
60—70	10	90
70—80	10	100
	$N = 100$	

$$\text{Median} = \text{Size of } \frac{N}{2}\text{th item} = \frac{100}{2} = \text{Size of 50th item}$$

Median lies in the class 40—50

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 40 + \frac{50-32}{28} \times 10$$

$$= 40 + \frac{18}{28} \times 10 = 40 + 6.42 = 46.42$$

$\therefore$ Median = 46.42

$$Q_1 = \text{Size of } \frac{N}{4}\text{th item} = \frac{100}{4} = 25\text{th item}$$

Q_1 lies in the class 30—40

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 30 + \frac{25-20}{12} \times 10$$

$$= 30 + \frac{5}{12} \times 10 = 30 + 4.16 = 34.16$$

$\therefore$ $Q_1 = 34.16$

$$Q_3 = \text{Size of } \frac{3N}{4}\text{th item} = \frac{3 \times 100}{4} = 75\text{th item.}$$

Q_3 lies in the class 50—60

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 50 + \frac{75-60}{20} \times 10$$

$$= 50 + \frac{15}{20} \times 10 = 50 + 7.5 = 57.5$$

$\therefore$ $Q_3 = 57.5$

$$D_6 = \text{Size of } \frac{6N}{10}\text{th item} = \frac{6 \times 100}{10} = 60\text{th item.}$$

D_6 lies in the class 40—50.

$$D_6 = l_1 + \frac{\frac{6}{10}N - c.f.}{f} \times i = 40 + \frac{60-32}{28} \times 10$$

$$= 40 + \frac{28}{28} \times 10 = 50$$

$\therefore$ $D_6 = 50$

$$P_{70} = \text{Size of } \frac{70N}{100}\text{th item} = \frac{70 \times 100}{100} = 70\text{th item}$$

P_{70} lies in the class 50—60

$$P_{70} = l_1 + \frac{\frac{70N}{100} - c.f.}{f} \times i = 50 + \frac{70-60}{20} \times 10$$

$$= 50 + \frac{10}{20} \times 10 = 55$$

$\therefore$ $P_{70} = 55$

IMPORTANT TYPICAL EXAMPLES

Example 48. Following is the distribution of marks in Economics obtained by 50 students:

Marks (more than):	0	10	20	30	40	50
No. of students:	50	46	40	20	10	3

Calculate the median marks. If 60% of the students pass this test, find the minimum marks obtained by a pass candidate.

Solution: Let us first of all convert cumulative frequencies into simple frequencies.

Marks	f	c.f.
0—10	50 − 46 = 4	4
10—20	46 − 40 = 6	10
20—30	40 − 20 = 20	30
30—40	20 − 10 = 10	40
40—50	10 − 3 = 7	47
50 and above	3	50
	N = 50	

(*i*) Median item = Size of $\left(\frac{N}{2}\right)$th item = Size of $\frac{50}{2}$th item = Size of 25th item.

Median lies in the class 20—30.

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{25-10}{20} \times 10$$

$$= 20 + \frac{15}{2}$$

$$= 20 + 7.5 = 27.5$$

(ii) Since, 60% candidates are passed and 40% have failed, the minimum pass marks are given by P_{40}. In other words, we have to find the value of P_{40}.

$$P_{40} = \text{Size of}\left(\frac{40N}{100}\right)\text{th item} = \text{Size of}\left(\frac{40\times 50}{100}\right)\text{th item}$$

$$= \text{Size of 20th item.}$$

P_{40} lies in the class 20—30

$$\therefore \quad P_{40} = l_1 + \frac{\frac{40}{100}N - c.f.}{f} \times i$$

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

Thus, the minimum marks obtained in a test by a pass candidate are 25.

Example 49. The first and third quartiles of the following data are given to be 25 marks and 50 marks respectively out of the data given below:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80
Frequency:	4	8	—	19	—	10	5	—

Find the missing frequencies when $N = 72$.

Solution: Let the missing frequencies be x, y and z. The given table is written as:

Marks	Frequency (f)	c.f.
0—10	4	4
10—20	8	12
20—30	x	$12+x$
30—40	19	$31+x$
40—50	y	$31+x+y$
50—60	10	$41+x+y$
60—70	5	$46+x+y$
70—80	z	$46+x+y+z$
	$N=72$	

Since, $N = 72$

$$\therefore \quad x+y+z+46 = 72$$

$$\Rightarrow \quad x+y+z = 72-46 = 26 \qquad \ldots(i)$$

$$Q_1 = \left(\frac{N}{4}\right)\text{th item} = \frac{72}{4} = 18\text{th item.}$$

Since x is still unknown, therefore, we cannot locate the position of Q_1 from c.f. column. As $Q_1 = 25$

$\therefore$ Q_1 lies in class 20–30.

Now,
$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

$$25 = 20 + \frac{18-12}{x} \times 10$$

or
$$5 = \frac{6}{x} \times 10$$

$$\Rightarrow \quad x = 12 \qquad \ldots(ii)$$

$$Q_3 = \left(\frac{3N}{4}\right)\text{th item} = \frac{3\times 72}{4} = 54\text{th item.}$$

Since, the value of y is still unknown, therefore, we cann not locate the position of Q_3 from c.f. column. As $Q_3 = 50$, Q_3 lies in 50–60 class.

Now,
$$Q_3 = l_1 + \frac{\left(\frac{3N}{4} - c.f.\right)}{f} \times i$$

$$50 = 50 + \frac{(54-31-x-y)}{10} \times 10$$

$$0 = 54-31-x-y$$

$$\Rightarrow \quad x+y = 23 \qquad \ldots(iii)$$

Putting the value of x from (ii) in equation (iii), we have

$$12 + y = 23$$

$$\Rightarrow \quad y = 11$$

Now,
$$x+y+z = 26$$

$$12 + 11 + z = 26$$

$$\Rightarrow \quad z = 3$$

$\therefore$ $x = 12$, $y = 11$ and $z = 3$

Example 50. Given the following distribution of income:

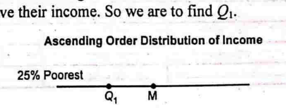
Income (thousands):	0—1	1—2	2—3	3—4	4—5	5—6	6—7	7—8
Number of families	4	6	10	15	8	5	4	2

Find out (i) Highest income among the poorest 25%.

(ii) Lowest income among the richest 30%.

Solution: (i) We are interested in finding that level of income less than or equal to which 25% of families have their income. So we are to find Q_1.

See Diagram



Income (thousands) (X)	Number of families (f)	c.f.
0—1	4	4
1—2	6	10
2—3	10	20
3—4	15	35
4—5	8	43
5—6	5	48
6—7	4	52
7—8	2	54
	$N = 54$	

$$Q_1 = \text{Size of } \frac{N}{4}\text{th item}$$

$$Q_1 = \text{Size of} \left(\frac{54}{4}\right) \text{th item} = 13.5\text{th item}$$

Q_1 lies in interval 2—3.

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - c.f.\right)}{f} \times i$$

$$Q_1 = 2 + \frac{13.5 - 10}{10} \times 1 = 2 + 0.35 = \text{Rs. } 2.35 \text{ thousands.}$$

So, highest income among the poorest 25% is Rs. 2.35 thousand.

(*ii*) We are interested in finding that level of income equal to or above which have 30% of the families have their income. So we are to find P_{70} or D_7.

See Diagram **Ascending Order Distribution of Income**

Richest 30% families

D_7 or P_{70}

Income (thousands) (X)	Number of families (f)	c.f.
0—1	4	4
1—2	6	10
2—3	10	20
3—4	15	35
4—5	8	43
5—6	5	48
6—7	4	52
7—8	2	54
	$N = 54$	

D_7 is equal to the size of $\frac{(7 \times 54)}{10}$ th item = 37.8th item

It lies in interval 4—5.

$$D_7 = l_1 + \frac{\frac{7N}{100} - c.f.}{f} \times i$$

$$D_7 = 4 + \frac{37.8 - 35}{8} \times 1$$

$$D_7 = 4 + \frac{2.8}{8} = \text{Rs. } 4.35 \text{ thousands.}$$

Hence, the lowest income among the richest 30% is 4.35 thousand rupees.

Example 51. An investigator collected the following information regarding the wage distribution:

Wages (Rs. week):	0—500	500—1000	1000—1500	1500—2000	2000—2500
Number of workers:	4	8	12	15	20
Wages (Rs. week):	2500—3000	3000—3500	3500—4000	4000—4500	4500—5000
Number of workers:	27	15	8	5	3

(*i*) Find out median wages.

(*ii*) Find out the highest wage among the lowest paid 30% workers.

(*iii*) Within what limits do the middle 20% workers have their wages?

(*iv*) The management decides to give bonus to workers with wages less than or equal to Rs. 3300 per week. How many workers will get the bonus?

Solution:

Wages (Rs. week) (X)	No. of workers (f)	c.f.
0—500	4	4
500—1000	8	12
1000—1500	12	24
1500—2000	15	39
2000—2500	20	59
2500—3000	27	86
3000—3500	15	101
3500—4000	8	109
4000—4500	5	114
4500—5000	3	117
	$N = 117$	

(i) $$\text{Median} = \text{Size of}\left(\frac{N}{2}\right)\text{th item} = \frac{117}{2} = 58.5\text{th item}$$

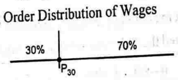
Median lies in the interval 2000—2500.

$$M = l_1 + \frac{\left(\frac{N}{2} - c.f.\right)}{f} \times i$$

$$M = 2000 + \frac{58.5 - 39}{20} \times 500 = 2000 + 487.5$$

$$M = \text{Rs. } 2487.5$$

(ii) Ascending Order Distribution of Wages



As shown here, the highest wage among the lowest paid 30% workers is given by P_{30}.

$$P_{30} = \text{Size of}\left(\frac{30N}{100}\right)\text{th item} = \frac{(30)(117)}{100} = 35.1\text{th item.}$$

P_{30} will lie in the class interval 1500—2000.

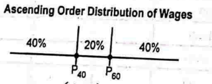
$$P_{30} = l_1 + \frac{\frac{30N}{100} - c.f.}{f} \times i$$

$$P_{30} = 1500 + \frac{35.1 - 24}{15} \times 500 = 1500 + \frac{11.1}{15} \times 1500$$

$$P_{30} = \text{Rs. } 1870.$$

(iii) The middle 20% of the workers will have wages between P_{40} and P_{60} as shown below:

Ascending Order Distribution of Wages



$$P_{40} = \text{Size of}\left(\frac{40N}{100}\right)\text{th item} = \frac{40 \times 117}{100} = 46.8\text{th item.}$$

It lies in the interval 2000—2005.

$$P_{40} = 2000 + \frac{46.8 - 39}{20} \times 500 = \text{Rs. } 2195$$

$$P_{60} = \text{Size of}\left(\frac{60N}{100}\right)\text{th item} = \frac{60 \times 117}{100} = 70.2\text{th item.}$$

It lies in the interval 2500—3000.

$$P_{60} = 2500 + \frac{70.2 - 59}{27} \times 500 = \text{Rs. } 2707.41$$

The limits within which the middle 20% of the workers will have their wages are Rs. 2195 and Rs 2707.41.

(iv) Suppose K workers have wages less than Rs. 3300. The class interval under question is 3000—3500. The value of K will be determined as below:

$$3300 = 3000 + \frac{K - c.f.}{f} \times i$$

$$3300 = 3000 + \frac{K - 86}{15} \times 500 \quad \Rightarrow \quad 300 = \frac{K - 86}{15} \times 500$$

$$\Rightarrow \quad K - 86 = \frac{300 \times 15}{500} = 9 \quad \Rightarrow \quad K = 95$$

The number of workers with wages equal to or less than Rs. 3300 is 95. So 95 workers will have to paid bonus.

Note: It is being assumed that distribution is uniform within class interval.

Alternative Approach to (d) Part:

Number of workers with wages from 0—3000 = 4 + 8 + 12 + 15 + 20 + 27
= 86 (given)

Number of workers within 3000—3500 = 15 (given)

Number of workers within 3000—3300 = $\frac{15}{500} \times 300 = 9$

Therefore, number of workers with wages less than or equal to Rs. 3300
= 86 + 9 = 95

Example 52. From the following data, calculate the percentage of workers getting wages:
(i) more than Rs. 44.
(ii) between Rs. 22 and Rs. 58.

Wage (in Rs.):	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80
Number of workers:	20	45	85	160	70	55	35	30

Solution:

Wage (in Rs.)	Number of workers (f)	$c.f.$
0—10	20	20
10—20	45	65
20—30	85	150
30—40	160	310
40—50	70	380
50—60	55	435
60—70	35	470
70—80	30	500
	$N = 500$	

(i) Let K% workers get less than Rs. 44. $\therefore$ $P_K = 44$.

$\therefore$ P_K class is 40—50.

Now, $P_K = l_1 + \left(\dfrac{K\left(\frac{N}{100}\right) - c.f.}{f}\right) \times i \Rightarrow 44 = 40 + \left(\dfrac{K\left(\frac{500}{100}\right) - 310}{70}\right) \times 10$

$\Rightarrow \quad 4 = \dfrac{5K - 310}{7} \quad \Rightarrow 5K = 338$

$\Rightarrow \quad K = 67.6$

$\therefore$ 67.6% workers get less than Rs. 44.

$\therefore$ Percentage of workers getting more than Rs. 44

$= 100\% - 67.6\%$

$= 32.4\%$

Alternative Method:

The number of workers getting more than 44 can be obtained directly as :

$= \dfrac{50-44}{10} \times 70 + 55 + 35 + 30$

$= 42 + 55 + 35 + 30$

$= 162$

Percentage of workers $= \dfrac{162}{500} \times 100 = 32.4\%$

(ii) Let K% workers get less than Rs. 22. $\therefore$ $P_K = 22$.

$\therefore$ P_K class is 20—30.

Now, $P_K = l_1 + \left(\dfrac{K\left(\frac{N}{100}\right) - c.f.}{f}\right) \times i$

$\Rightarrow \quad 22 = 20 + \left(\dfrac{K\left(\frac{500}{100}\right) - 65}{85}\right) \times 10$

$\Rightarrow \quad 2 = \dfrac{(5K - 65) \times 2}{17} \quad \Rightarrow \quad 34 = 10K - 130$

$\Rightarrow \quad K = 16.4.$

$\therefore$ 16.4% workers get less than Rs. 22.

Let K% workers get less than Rs. 58. $\therefore$ $P_K = 58$

$\therefore$ P_K class is 50—60.

Now, $P_K = l_1 + \left(\dfrac{K\left(\frac{N}{100}\right) - c.f.}{f}\right) \times i$

$\Rightarrow \quad 58 = 50 + \left(\dfrac{K\left(\frac{500}{100}\right) - 380}{55}\right) \times 10$

$\Rightarrow \quad 8 = \dfrac{(5K - 380) \times 2}{11} \quad \Rightarrow \quad 88 = 10K - 760$

$\Rightarrow \quad K = 84.8$

$\therefore$ 84.8% workers get less than Rs. 58.

$\therefore$ Percentage of workers getting between Rs. 22 and Rs. 58

$= 84.8\% - 16.4\% = 68.4\%$

Alternative Method:

The number of workers getting between Rs. 22 and Rs. 58 can be obtained directly as:

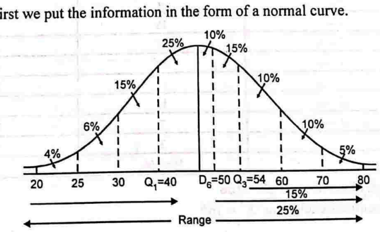
$= \dfrac{30-22}{10} \times 85 + 160 + 70 + \dfrac{58-50}{10} \times 55$

$= 68 + 160 + 70 + 44 = 342$

Percentage of workers $= \dfrac{342}{500} \times 100 = 68.4\%$

Example 53. For a group of 5,000 workers, the weekly wages vary from Rs. 20 and Rs.80. The wages of 4 per cent of workers are under Rs. 25 and those of 10 per cent are under Rs. 30. 15 per cent of the workers each Rs. 60and over and 5 per cent of them get Rs. 70 and over. The quartile wages are Rs. 40 and Rs. 54. The sixth Decile is Rs. 50. Put this information in a frequency table and find mean wages.

Solution: First we put the information in the form of a normal curve.



Calculation of Mean Wages

Weakly wages (Rs.)	Percentage(%) of workers	No.of workers (f)	Mid-value (m)	fm
20-25	4%	$\frac{4\times5000}{100}=200$	22.5	4500
25-30	6%	$\frac{6\times5000}{100}=300$	27.5	8250
30-40	15%	$\frac{15\times5000}{100}=750$	35	26250
40-50	35%	$\frac{35\times5000}{100}=1750$	45	78750
50-54	15%	$\frac{15\times5000}{100}=750$	52	39000
54-60	10% (balance)	$\frac{10\times5000}{100}=500$	57	28500
60-70	10%	$\frac{10\times5000}{100}=500$	65	32500
70-80	5%	$\frac{5\times5000}{100}=250$	75	18750
		N = 5000		$\Sigma fm = 236500$

$$\overline{X}=\frac{\Sigma fm}{N}=\frac{236500}{5000}=47\cdot3$$

Example 54. The age distribution of the members of a certain children's club is as follows:

Age as on last birth day (in years):	4	5	6	7	8	9	10	11	12
Number of children:	5	9	18	35	42	32	15	7	3

There is a member 'A' such that there are twice as many members older than 'A' as there are members younger than 'A'. Estimate the age of A.

Solution: As the age on the last birthday is given, we can put the data in the exclusive form of class intervals as under:

Age in years	Number of children	c.f.
4—5	5	5
5—6	9	14
6—7	18	32
7—8	35	67
8—9	42	109
9—10	32	141
10—11	15	156
11—12	7	163
12—13	3	166
	N = 166	

Total members of the club = 166

Total members excluding 'A' = 165

Let total members younger than 'A' = x

The members older than 'A' = $2x$

Thus, $x+2x=165 \therefore x=55$

Thus, 'A's position is at 56th in order of age.

$\therefore$ A's age lies in 7—8 age group

$$\text{A's age} = l_1+\frac{\text{A's position}-c.f.}{f}\times i$$

$$=7+\frac{56-32}{35}\times1=7+0.6857=7.6857=7.6 \text{ years}$$

Graphical Location of Quartiles, Deciles and Percentiles

The various partition values *viz.*, quartiles (Q_1 and Q_3), deciles and percentiles can be easily located graphically with the help of less than cumulative frequency curve or ogive. For first and third quartiles, we mark the values $\frac{N}{4}$ and $\frac{3N}{4}$ on the Y-axis. For rth decile we mark the value $\left(\frac{N}{10}\right)r$ and for rth percentile we mark $\left(\frac{N}{100}\right)$ r on Y-axis. Rest of the steps are the same as explained for drawing median by using the lees than ogive.

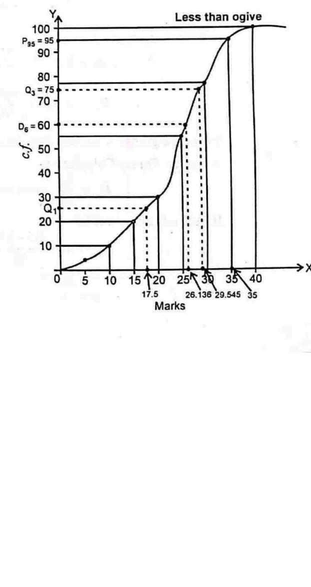
Example 55. The marks obtained by 100 students of a university are given below:

Marks:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
No. of students:	4	6	10	10	25	22	18	5

Draw a less than ogive from the data given above and hence find out Q_1, Q_3, D_6 and P_{95}. Also verify your results by direct formula calculation.

Solution:

Marks	c.f.
Less than 5	4
10	10
15	20
20	30
25	55
30	77
35	95
40	100



$$Q_1 = \text{Size of}\left(\frac{N}{4}\right)\text{th item} = \frac{100}{4} = 25\text{th item}$$

From the graph, it is clear that size of 25th item = 17.5.

Value by Direct Calculation

$$Q_1 = \text{Size of}\left(\frac{N}{4}\right)\text{th item} = \frac{100}{4} = 25\text{th item}$$

It lies in the interval 15—20.

$$Q_1 = 15 + \frac{25-20}{10} \times 5 = 17.5$$

$$Q_3 = \text{Size of}\left(\frac{3N}{4}\right)\text{th item} = \frac{3 \times 100}{4} = 75\text{th item.}$$

From the graph, it is clear that size of 75th item = 29.545.

Value by Direct Calculation

$$Q_3 = \text{Size of}\left(\frac{3N}{4}\right)\text{th item} = \frac{3 \times 100}{4} = 75\text{th item.}$$

It lies in the interval 25—30.

$$Q_3 = 25 + \frac{75-55}{22} \times 5 = 29.545$$

$$D_6 = \text{Size of}\left(\frac{6N}{10}\right)\text{th item} \; \frac{6 \times 100}{10} = 60\text{th item.}$$

From the graph, it is clear that the size of 60th item = 26.136.

Value by Direct Calculation

$$D_6 = \frac{6}{10} \times 100 = 60\text{th item.}$$

It lies in the interval 25—30.

$$D_6 = 25 + \frac{(60-55)}{22} \times 5 = 26.136$$

$$P_{95} = \text{Size of}\left(\frac{95N}{100}\right)\text{th item} = \frac{95 \times 100}{100} = 95\text{th item}$$

From the graph, it is clear that the size of 95th item = 35.

Value by Direct Calculation

$$P_{95} = \frac{95}{100} \times 100 = 95\text{th item.}$$

It lies in the interval 30—35.

$$P_{95} = 30 + \frac{95-77}{18} \times 5$$
$$= 35$$

Example 56. Given below is the pre-tax monthly income of residents of an industrial town:

Pre-tax income (Rs.)	No. of Residents (in thousands)
More than 7,000	2
More than 6,000	8
More than 5,000	10
More than 4,000	15
More than 3,000	25
More than 2,000	40
More than 1,000	55
More than 0	60

Draw a 'less than ogive' and hence find out.

(*i*) the highest income of the lowest 50% of the residents; and

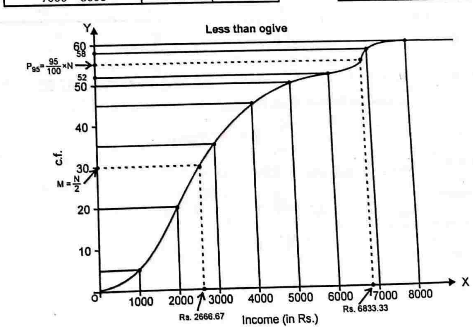
(*ii*) the minimum income earned by the top 5% of the residents.

Also, verify you results by direct formula calculation.

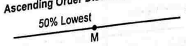
Solution: First we convert the given cumulative frequency distribution into simple frequency distribution and then into less than cumulative frequency distribution.

Income (Rs.)	f	c.f.
0—1000	5	5
1000—2000	15	20
2000—3000	15	35
3000—4000	10	45
4000—5000	5	50
5000—6000	2	52
6000—7000	6	58
7000—8000	2	60

Income	c.f.
Less than 1000	5
Less than 2000	20
Less than 3000	35
Less than 4000	45
Less than 5000	50
Less than 6000	52
Less than 7000	58
Less than 8000	60



(i) Ascending Order Distribution of Income



We are interested in finding the highest income of the lowest 50% of the residents. So we are find M or D_5 or P_{50}

$$M = \text{Size of} \left(\frac{N}{2}\right) \text{th item} = \frac{60}{2} = 30\text{th item.}$$

From the graph, it is clear that size of 30th item is 2667.

Value by Direct Calculation

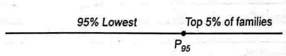
$$\text{Median} = \frac{N}{2} = \frac{60}{2} = 30\text{th item.}$$

It lies in the class interval 2000—3000.

$$M = 2000 + \frac{30-20}{15} \times 1000$$

$$= 2666.67 \approx 2667 \text{ approx.}$$

Ascending Order Distribution of Income



(ii) We are interested in finding the income earned by 5% of the families. So we are to find P_{95}.

$$P_{95} = \frac{95}{100} \times 60$$

$$= 57\text{th item.}$$

From the graph, it is clear that size of 57th item is 6833.33.

Value by Direct Calculation

$$P_{95} = \frac{95}{100} \times 60 = 57\text{th item.}$$

It lies in the interval 6000—7000.

$$P_{95} = 6000 + \frac{57-52}{6} \times 1000$$

$$= 6833.33.$$

Example 57. The monthly salary distribution of 250 families in a certain locality in Agra is given below:

Monthly salary	No. of families
More than 0	250
More than 500	200
More than 1,000	120
More than 1,500	80
More than 2,000	55
More than 2,500	30
More than 3,000	15
More than 3,500	5

Draw a 'less than' ogive for the data given above and hence find out:

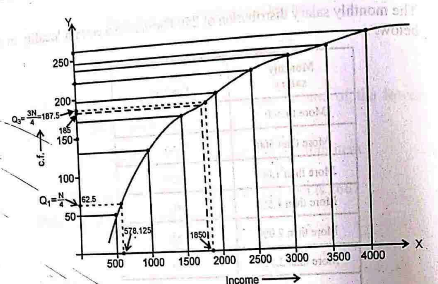
(i) Limits of the income of middle 50% of the families; and

(ii) If income-tax is to be levied on families, whose income exceeds Rs. 1,800 p.m., calculate the percentage of families, which will be paying income-tax.

Solution: First we convert the given frequency distribution into simple frequency distribution and then into less than cumulative frequency distribution.

Monthly salary	No. of families (f)	c.f.
0—500	50	50
500—1000	80	130
1000—1500	40	170
1500—2000	25	195
2000—2500	25	220
2500—3000	15	235
3000—3500	10	245
3500—4000	5	250

Monthly salary	c.f.
Less than 500	50
Less than 1000	130
Less than 1500	170
Less than 2000	195
Less than 2500	220
Less than 3000	235
Less than 3500	245
Less than 4000	250



(*i*) The middle 50% of the families will have income between Q_3 and Q_1 as shown below:

25% 50% 75%
Q_1 Q_3

$$Q_1 = \text{Size of}\left(\frac{N}{4}\right)\text{th item} = \frac{250}{4} = 62.5\text{th item.}$$

From the graph, it is clear that $Q_1 = 578.125$.

Value by Direct Calculation

$$Q_1 = \text{Size of}\left(\frac{N}{4}\right)\text{th item} = \frac{250}{4} = 62.5\text{th item.}$$

It lies in the class interval 500—1000.

$$Q_1 = 500 + \frac{62.5-50}{80}\times 500 = 578.125$$

$$Q_3 = \text{Size of}\left(\frac{3N}{4}\right)\text{th item}\ \frac{3\times 250}{4} = 187.5\text{th item.}$$

From the graph, it is clear that $Q_3 = 1850$.

Value by Direct Calculation

$$Q_3 = \text{Size of}\left(\frac{3N}{4}\right)\text{th item}\ \frac{3\times 250}{4} = 187.5\text{th item.}$$

It lies in the class interval 1500—2000.

$$Q_3 = 1500 + \frac{187.5-170}{25}\times 500 = \text{Rs } 1{,}850.$$

Income limit of middle 50% families = 1850 − 578.12 = Rs. 1,271.88

(*ii*) From the graph, it is clear that number of families having income less than or equal to 1800 = 185

No. of families having income more than 1800 = 250 − 185 = 65

Percentage of families having income more than 1800 $= \frac{65}{250}\times 100 = 26$

Value by Direct Calculation

Let K% of families getting less or equal to 1800; $\therefore$ $P_K = 1800$

$\therefore$ P_K class is 1500—2000.

$$\text{Now,}\quad P_K = 1500 + \frac{\frac{K\times 250}{100}-170}{25}\times 500$$

$$\Rightarrow\quad 1800 = 1500 + \frac{2.5K-170}{25}\times 500$$

$$\frac{300\times 25}{500} = 2.5K - 170$$

$$2.5K - 170 = 15$$

$$K = \frac{185}{2.5} = 74\%.$$

Percentage of families getting less than 1800 = 74%.

Percentage of families getting more than 1800 = 100 − 74 = 26%.

Another Method:

Percentage of families having income more than in 1800

$$= \frac{2000-1800}{500}\times 25 + 25 + 15 + 10 + 5$$

$$= 10 + 25 + 15 + 10 + 5 = 65$$

$$\text{Percentage of families} = \frac{65}{250}\times 100 = 26\%$$

EXERCISE 5.6

1. 19 students of B. Com. II class secured following marks in Economics:
   18, 20, 25, 17, 9, 11, 23, 37, 38, 42, 36, 35, 8, 10, 11, 21, 20, 41, 35
   Calculate Q_1, Q_3, D_7 and P_{78} [Ans. $Q_1 = 11, Q_3 = 36, D_7 = 35, P_{78} = 36.6$]

2. From the following data calculate Q_1, Q_3 and P_{60}

Sr.No.	1	2	3	4	5	6	7	8	9	10
Marks:	12	30	20	15	25	10	2	40	4	8

[Ans. $Q_1 = 7, Q_3 = 26.25, P_{60} = 18$]

3. Calculate Median, Q_1 and Q_3, D_6 and P_{89} from the following data:

X:	1	2	3	4	5	6	7	8	9	10	11
f:	5	8	15	22	36	44	28	17	12	9	3

[Ans. $M = 6, Q_1 = 4, Q_3 = 7, D_6 = 6, P_{89} = 9$]

4. Calculate median, lower and upper quartiles, first decile and 68th percentile from the following data:

Size:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
f:	6	7	9	18	25	15	12	8

[Ans. M = 22, Q_1 = 15.83, Q_3 = 28.33, D_1 = 7.86, P_{68} = 26]

5. Calculate the median and quartiles for the following:

Marks below:	10	20	30	40	50	60	70	80
No. of students:	15	35	60	84	96	127	198	250

[Ans. M = 59.35, Q_1 = 31.04, Q_3 = 68.52]

6. The first quartile of the following data is given to be 21.5 marks.

Marks:	10—15	15—20	20—25	25—30	30—35	35—40	40—45	45—50
Frequency:	24	?	90	122	?	56	20	33

Find the missing frequencies when N = 460. [Ans. $f_1 = 64$ and $f_2 = 51$]

7. For a certain group of 100 saree-weavers of Varanasi, the median and quartile earnings per week are Rs. 44.3, Rs. 43.0 and Rs. 45.9 respectively. The earnings for the group range between Rs. 40 and Rs. 50. 10 per cent of the group earn under Rs. 42.13 per cent earn Rs. 47 and over and 6 per cent Rs. 48 and other. Put these figures in the form of frequency distribution and obtain the value of the mean wage. [Ans. $\overline{X} = 44.5$]

8. The following series relates to the daily income of working employed in a firm. Compute (*i*) highest income of lowest 50% workers (*ii*) minimum income earned by the top 25% of the workers, and (*iv*) maximum income earned by lowest 25% workers:

Daily Income (Rs.):	10—14	15—19	20—24	25—29	30—34	35—39
No. of Workers:	5	10	15	20	10	5

[Hint: Compute Median, Upper Quartile and Lower Quartile]

[Ans. (*i*) Rs. 25.11 (*ii*) Rs. 29.19 (*iii*) Rs. 19.92]

9. From the following data of wages, find the percentage of workers getting wages (*i*) between Rs. 125 and Rs. 260; (*ii*) more than Rs. 340.

Weekly wages (Rs.)	50—100	100—150	150—200	200—250	250—300
No. of workers	15	40	35	60	125
Weekly wages (Rs.)	300—350	350—400	400—450	450—500	
No. of workers	100	70	40	15	

[Ans. (*i*) 28% (*ii*) 29%]

10. If the quartiles for the following distribution are $Q_1 = 23.125$ and $Q_3 = 43.5$, find the missing frequencies and median of the distribution:

Daily wages:	0—10	10—20	20—30	30—40	40—50	50—60
No. of workers:	5	—	20	30	—	10

[Ans. $f_1 = 11, f_2 = 13.669 \approx 14$ (approx.)]

11. Find the percentage of workers who earned more than Rs. 840.

Income (in Rs.):	500—600	600—700	700—800	800—900	900—1000	1000—1100
No. of workers:	12	17	22	29	12	6

[Ans. 36.12%]

12. Given below is the distribution of works obtained by 50 students in a class test:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	3	5	9	12	18	3

If 70% students pass the test, find the minimum marks needed by a student to pass the examination.

[Hint: Compute P_{70}] [Ans. 28 Marks]

13. Draw a less than ogive from the following data:

Weekly income (Rs.) (equal to or more than)	12,000	11,000	10,000	8,000	6,000	4,000	3,000	2,000	1,000
No. of families	0	6	14	26	42	54	62	70	80

From the graph estimate the number of families in the income range of Rs. 2,500 and Rs. 10,500. Also find maximum income of the lowest 25% of the families. [Ans. (*i*) 56, (*ii*) Rs. 3250]

14. With the help of given data, find (*i*) value of middle 50% class, (*ii*) value of exactly 50% class, (*iii*) the value of P_{40} and D_6, (*iv*) graphically with the help of ogive curves, the values of Q_1, Q_3, median, P_{40} and D_6.

C.I.:	10—14	15—19	20—24	25—29	30—34	35—39
f:	5	10	15	20	10	5

[Ans. $Q_3 - Q_1 = 9.2375$, $M(= Q_2) = 25.125$, $P_{40} = 23.167$, $D_6 = 26.75$]

15. The following table gives the frequency distribution of 800 candidates in an examination:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90	90—100
No. of candidates	10	40	80	140	170	130	100	70	40	20

Draw less than ogive and answer the following questions from the graph:

(*i*) If the minimum works required for passing are 35, what percentage of candidates pass the examination?

(*ii*) If it is decided to allow 80% of the candidates to pass, what should be the minimum marks for passing?

[Hint: See Example 122] [Ans. (*i*) 75% (*ii*) 32 approx.]

(4) MODE

Mode is another important measure of central tendency. Mode is defined as the value which occurs most frequently in a series. In other words, it is a value which has the greatest frequency in a distribution. For example, the mode of the series 20, 21, 23, 23, 23, 23, 23, 25, 26, 26 would be 23, since this value occurs most frequently than any of other values. In the words of **Kenny and Keeping**, "*The value of the variable which occurs most frequently in a distribution is called the mode*". According to **A.M. Tuttle**, "*Mode is the value which has the greatest frequency density*". The above definitions indicate that mode is the value around which there is greatest concentration of items. Mode is denoted by the symbol 'Z'.

Calculation of Mode

Individual Series

In case of individual series, mode can be computed by applying any of the two methods:

(1) Inspection Method: This method involves an inspection of the items. One is to simply identify the value that occurs most frequently in a series. Such a value is called mode.

Example 58. Find the mode from the following data:

8, 10, 5, 8, 12, 7, 8, 9, 11, 7

Solution: An inspection of the series shows that the value 8 occurs most frequently in the series. Hence, Mode (Z) = 8.

(2) By changing the Individual series into Discrete Series: When the numbers of items in a series is very large, individual series is first converted into discrete series. Then we identify the value corresponding to which there is highest frequency. Such a value is called mode.

Example 59. Find the mode from the following data:

11.1, 10.9, 10.7, 11.1, 10.6, 11.3, 10.6, 10.7, 10.6, 10.9, 10.6, 10.5, 10.4, 10.6

Solution: First we convert the given series into a discrete series in ascending order as follows:

Size:	10.4	10.5	10.6	10.7	10.9	11.1	11.3
Frequency:	1	1	5	2	2	2	1

The modal value is 10.6 since it appears maximum number of times in the series.

Discrete Series

For calculating mode in discrete series, the following two methods may be used:

(1) Inspection Method

(2) Grouping Method.

(1) Inspection Method: In this method, the value of mode is determined by inspecting the series. The value whose frequency is maximum is mode. Generally, inspection method is used in only those cases where frequency increase upto a point and after reaching the maximum, decline.

Example 60. Find out the mode from the following data:

Income (Rs.):	110	120	130	140	150	160
No. of persons:	2	4	8	10	5	4

Solution: An inspection of the series reveals that the value 140 has the maximum frequency, *i.e.*, 10. Thus, mode is 140.

(2) Grouping Method: In some cases, it is possible that value having the highest frequency may not be the modal value. This will specially be so where the difference between the maximum frequency and the frequency preceeding or succeeding it is very small and items are heavily concentrated on either side. Inspection method will also be of no use when the frequencies in the immediate neighbourhood of the highest frequency are very low. In such cases, mode can be determined only by grouping method. Under grouping method, modal value is determined by preparing two tables—(*i*) grouping table and (*ii*) analysis table.

Preparation of Grouping Table

A grouping table has six columns. The various steps in its preparation are as follows:

(*i*) In the column first, the maximum frequency is marked, underlined or put in a circle.

(*ii*) In column second, frequencies are grouped in two's, starting with the first two frequencies of the series.

(*iii*) In the column third, first frequency is left out and the remaining are grouped in two's.

(*iv*) In fourth column, frequencies are grouped in three's starting with the first three frequencies.

(*v*) In column fifth, leave the first frequency and group the remaining in three's.

(*vi*) In the column sixth, leave the first two frequencies and group the other frequencies in three's.

We mark, underline or circle the maximum frequency in each column. The six columns are to serve as the basis for the preparation of analysis table.

Preparation of Analysis Table

After the preparation of grouping table, the analysis table is prepared. Following steps are followed for its preparation:

(*i*) Put the column numbers on the left hand side.

(*ii*) Put the probable values of the mode on the right hand side.

(*iii*) Now enter into columns the highest frequencies marked in the grouping table.

(*iv*) Take the total of each column to find out the value repeated maximum number of times. This value against which the total is highest is the mode.

Example 61. From the following data, determine the mode by grouping method:

X:	7	8	9	10	11	12	13	14	15	16	17
f:	2	3	6	12	20	24	25	7	5	3	1

Solution:

Grouping Table

X	I (f)	II (1+2)	III (2+3)	IV (1+2+3)	V (2+3+4)	VI (3+4+5)
7	2					
		5				
8	3			11		
			9			
9	6				21	
		18				
10	12					38
			32			
11	20			56		
		44				
12	24				69	
			49			
13	25					56
		32				
14	7			37		
			12			
15	5				15	
		8				
16	3					
			4			9
17	1					

Analysis Table

Col. No.	7	8	9	10	11	12	13	14	15	16	17
						Size (X)					
I							√				
II					√	√					
III						√	√				
IV				√	√	√					
V					√	√	√				
VI						√	√	√			
Total				1	3	5	4	1			

It is clear from the above table that the size 12 occurs the maximum number of times, *i.e.* 5 times. Thus, the mode (Z) is 12.

Continuous Series

The following steps are taken for the determination of mode in a continuous series:

(*i*) Firstly, modal group is ascertained either by using inspection method or by grouping method. The procedure to be followed will remain the same as in discrete series.

(*ii*) After determining the modal group, mode can be found by using the following formula:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Where, Z = Mode, l_1 = lower limit of the modal class, f_1 = frequency of the modal class, f_0 = frequency of the pre modal class, f_2 = frequency of the post modal class, i = size of the modal group.

Aliter: The above mentioned formula can also be written as:

$$Z = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

Where, l_1=lower limit of the modal class, $\Delta_1 = |f_1 - f_0|$ and $\Delta_2 = |f_1 - f_2|$.

For the calculation of Δ_1 and Δ_2, signs are ignored.

Note 1. If the first class is the modal class, then f_0 is taken as zero. Similarly, if the last class is modal class, then f_2 is taken as zero.

Note 2. If the modal value lies outside the model class, the following formula is used to calculate the mode:

$$Z = l_1 + \frac{f_2}{f_0 + f_2} \times i$$

Note 3. If mode is ill-defined, then we use the following formula $Z = 3M - 2\overline{X}$ where, M = Median, $\overline{X}$ = Mean.

Example 62. Calculate the mode from the following data:

Wages (in Rs.):	0—5	5—10	10—15	15—20	20—25	25—30	30—35
No. of workers:	3	7	15	30	20	10	5

Solution: Since, the series is regular, we may not do grouping for the location of model group. By inspection, modal class is 15—20

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Here, $l_1 = 15, f_1 = 30, f_0 = 15, f_2 = 20, i = 5$

Substituting the values, we get

$$Z = 15 + \frac{30 - 15}{2(30) - 15 - 20} \times 5 = 15 + \frac{15}{25} \times 5 = 18$$

Thus, mode = 18.

Aliter: Mode can also be calculated by using the formula:

$$Z = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

Here, $l_1 = 15,\ \Delta_1 = |f_1 - f_0| = |30 - 15| = 15, \Delta_2 = |f_1 - f_2| = |30 - 20| = 10, i = 5$

$$\therefore \quad Z = 15 + \frac{15}{15+10} \times 5 = 15 + \frac{15}{25} \times 5 = 15 + 3 = 18$$

$$\therefore \quad Z = 18$$

Cumulative Frequency Series

Example 63. Calculate mode from the following data:

Marks between	No. of students
10 and 15	4
10 and 20	12
10 and 25	30
10 and 30	60
10 and 35	80
10 and 40	90
10 and 45	95
10 and 50	97

Solution: Since, this is a cumulative frequency series, we first convert it into simple frequency series:

Marks	f
10—15	4
15—20	8
20—25	18
25—30	30
30—35	20
35—40	10
40—45	5
45—50	2

By inspection, the modal class is 25—30

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$Z = 25 + \frac{30 - 18}{2(30) - 18 - 20} \times 5 = 25 + \frac{12}{60 - 18 - 20} \times 5$$

$$= 25 + \frac{60}{22} = 25 + 2.73 = 27.73$$

$$\therefore \quad Z = 27.73$$

Aliter: Mode can also be located by using the formula:

$$Z = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

Here, $\Delta_1 = |f_1 - f_0| = |30 - 18| = 12,\ \Delta_2 = |f_1 - f_2| = |30 - 20| = 10$

$$\therefore \quad Z = 25 + \frac{12}{12+10} \times 5 = 25 + \frac{60}{22} = 25 + 2.73 = 27.73$$

$$\therefore \quad Z = 27.73$$

Inclusive Series

Example 64. Calculate the mode from the following data:

Class:	20—24	25—29	30—34	35—39	40—44	45—49	50—54	55—59
Frequency:	3	5	10	20	12	6	3	1

Solution: Since we are given inclusive class intervals, we first convert it into exclusive one.

Class	Frequency (f)
19.5—24.5	3
24.5—29.5	5
29.5—34.5	10 f_0
34.5—39.5	20 f_1
39.5—44.5	12 f_2
44.5—49.5	6
49.5—54.5	3
54.5—59.5	1

By inspection, modal class is 34.5—39.5

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 34.5 + \frac{20 - 10}{2 \times 20 - 10 - 12} \times 5$$

$$= 34.5 + \frac{10}{40 - 22} \times 5 = 34.5 + \frac{50}{18} = 34.5 + 2.77 = 37.27$$

$$\therefore \quad Z = 37.27$$

Unequal Class Intervals

Example 65. Calculate the mode from the following data:

X:	0—5	5—10	10—20	20—25	25—30	30—40	40—44	44—50	50—70
f:	4	8	10	9	13	30	6	9	12

Solution: Since the class intervals are unequal, we make them equal before calculating the value of mode.

X	f
0—10	4 + 8 = 12
10—20	10
20—30	9 + 13 = 22 f_1
30—40	30 f_0
40—50	6 + 9 = 15 f_2
50—60	6
60—70	6

By inspection, modal class is 30—40

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 30 + \frac{30 - 22}{2 \times 30 - 22 - 15} \times 10$$

$$= 30 + \frac{8}{60 - 22 - 15} \times 10$$

$$= 30 + \frac{80}{23} = 30 + 3.47 = 33.47$$

Mid-value Series

Example 66. Calculate the mode from the following data:

Class:	1	2	3	4	5	6	7	8	9	10
Frequency:	8	6	10	12	20	12	5	3	2	4

Solution: Since we are given mid-values (*i.e.*, central size), first we determine the lower and upper limits of the classes by using the formula:

$l_1 = m - i/2 \quad l_2 = m + i/2$ where, i = difference between two mid-values.
m = mid-values (or central size)

Here, $i = 1$, $m = 1$,

$\therefore \; l_1 = 1 - 1/2 = 0.5, \; l_2 = 1 + 1/2 = 1.5$

Hence the first class would be 0.5—1.5

Class	Frequency (f)
0.5—1.5	8
1.5—2.5	6
2.5—3.5	10
3.5—4.5	12
4.5—5.5	20 = f_1
5.5—6.5	12
6.5—7.5	5
7.5—8.5	3
8.5—9.5	2
9.5—10.5	4

By inspection, modal class is 4.5—5.5

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Here, $l_1 = 4.5, f_1 = 20, f_0 = 12, f_2 = 12, i = 1$

$$\therefore \quad Z = 4.5 + \frac{20 - 12}{2 \times 20 - 12 - 12} \times 1$$

$$= 4.5 + \frac{8}{16}$$

$$= 4.5 + 0.5$$

$$= 5.0$$

Hence, $Z = 5$.

Bi-Modal Series

Example 67. Calculate the value of mode from the following data:

Marks:	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90
No. of students:	4	6	20	32	33	17	8	2

Solution: By inspection, it is difficult to say which is the modal class. Hence we use grouping method to locate the modal class.

Grouping Table

Marks	I (f)	II (1+2)	III (2+3)	IV (1+2+3)	V (2+3+4)	VI (3+4+5)
10—20	4					
		10		30		
20—30	6					
			26		58	
30—40	20					85
		52				
40—50	32					
			65	82		
50—60	33					
		50				
60—70	17				58	
			25			
70—80	8					27
		10				
80—90	2					

Analysis Table

Col. No.	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90
I					√			
II			√	√				
III				√	√			
IV				√	√	√		
V		√	√	√	√	√	√	
VI			√	√	√			
Total		1	3	5	5	2	1	

As the maximum frequency occurs twice, it is a bi-modal series. Hence, mode is determined applying the formula:

Mode = 3 Median – 2 Mean

Calculation of Mean and Median

Marks	f	M.V. (m)	A = 55 d	$d' = d/10$	fd'	$c.f.$
10—20	4	15	–40	–4	–16	4
20—30	6	25	–30	–3	–18	10
30—40	20	35	–20	–2	–40	30
40—50	32	45	–10	–1	–32	62
50—60	33	55A	0	0	0	95
60—70	17	65	+10	+1	17	112
70—80	8	75	+20	+2	16	120
80—90	2	85	+30	+3	6	122
	$N = 122$				$\Sigma fd' = -67$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

$$= 55 - \frac{67}{122} \times 10 = 55 - \frac{670}{122}$$

$$= 49.51$$

Median item = Size of $\frac{N}{2}$th item = $\frac{122}{2}$ = 61th item which lies in 40—50.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 40 + \frac{61 - 30}{32} \times 10$$

$$M = 40 + \frac{310}{32} = 40 + 9.69 = 49.69$$

Thus, $\overline{X} = 49.51$, $M = 49.69$

Now, $Z = 3M - 2\overline{X}$

$= 3(49.69) - 2(49.51)$

$\therefore$ $Z = 149.07 - 99.02 = 50.05$

Failure of Formula

Example 68. Calculate mode of the following data:

Wages (Rs.):	25—35	35—45	45—55	55—65	65—75
No.of workers:	4	44	38	28	6
Wages (Rs.):	75—85	85—95	95—105	105—115	
No.of workers:	8	12	2	2	

Solution:

Grouping Table

Wages Rs.	I (f)	II (1+2)	III (2+3)	IV (1+2+3)	V (2+3+4)	VI (3+4+5)
25—35	4					
		48				
35—45	44			86		
			82			
45—55	38				110	
		66				
55—65	28					72
			34			
65—75	6			42		
		14				
75—85	8				26	
			20			
85—95	12					22
		14				
95—105	2			16		
			4			
105—115	2					

Analysis Table

Col. No.	25 – 35	35 – 45	45 – 55	55 – 65	65 – 75	75 – 85	85 – 95	95 – 105	105 - 115
I		√							
II			√	√					
III		√	√						
IV	√	√	√						
V		√	√	√					
VI			√	√	√				
Total	1	4	5	3	1				

As the maximum frequency occurs in the class 45 – 55, hence model class is 45 – 55. Applying the formula:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 45 + \frac{38 - 44}{2 \times 38 - 44 - 28} \times 10$$

$$= 45 + \frac{(-6)}{76 - 44 - 28} \times 10$$

$$= 45 - \frac{60}{4} = 45 - 15 = 30$$

The value of the mode lies outside the model class. In such a case, the following alternative formula is used:

$$Z = l_1 + \frac{f_2}{f_0 + f_2} \times i$$

$$= 45 + \frac{28}{44 + 28} \times 10 = 45 + \frac{28}{72} \times 10$$

$$= 45 + 3.89 = 48.49$$

IMPORTANT TYPICAL EXAMPLES

Example 69. An incomplete distribution families according to their expenditure per week is given below. The median and mode for the distribution are Rs. 25 and Rs. 24 respectively. Calculate the missing frequencies:

Expenditure:	0—10	10—20	20—30	30—40	40—50
No. of families:	14	?	27	?	15

Solution: Two frequencies are missing and let the missing frequencies be x and y respectively.

Expenditure	No. of families	c.f.
0—10	14	14
10—20	x	$14 + x$
20—30	27	$41 + x$
30—40	y	$41 + x + y$
40—50	15	$56 + x + y$
	$N = 56 + x + y$	

(*i*) As $M = 25$ ∴ Median class is 20—30

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$\Rightarrow \quad 25 = 20 + \frac{\left(\frac{56 + x + y}{2} - 14 - x\right)}{27} \times 10$$

$$\Rightarrow \quad \frac{135}{10} = \frac{28 - x + y}{2} \quad \text{or} \quad 27 = 28 - x + y$$

$$\Rightarrow \quad x - y = 1 \Rightarrow y = x - 1$$

(ii) As $Z = 24$ ∴ Modal class is 20—30

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$24 = 20 + \frac{27 - x}{54 - x - y} \times 10$$

$$4(54 - x - y) = (27 - x)10$$

$$2[54 - x - (x - 1)] = 5(27 - x) \quad \text{[Using } y = x - 1\text{]}$$

$$108 - 2x - 2x + 2 = 135 - 5x$$

⇒ $x = 25$

∴ $y = x - 1$

∴ $y = 24$

∴ The missing frequencies are 25 and 24.

Example 70. The median and mode of the following wage distribution of 230 workers are known to be Rs. 33.5 and Rs. 34 respectively. Three frequency values from the table however are missing. Find the missing values:

Wages in Rs.:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	Total
No. of workers:	4	16	—	—	—	6	4	230

Solution: Let the missing frequencies be denoted by x, y and z.

Wages (Rs.)	f	$c.f.$
0—10	4	4
10—20	16	20
20—30	x	$20+x$
30—40	y	$20+x+y$
40—50	z	$20+x+y+z$
50—60	6	$26+x+y+z$
60—70	4	$30+x+y+z$
	$N = 230$	

Now, $230 = 30 + x + y + z$

∴ $x + y + z = 230 - 30 = 200$

and $z = 200 - x - y$ ...(i)

Since median and mode are 33.5 and 34, they both lie in the class 30—40

Median item = Size of $\frac{N}{2}$ th item

$= \frac{230}{2} = 115$th item.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$33.5 = 30 + \frac{115 - (20 + x)}{y} \times 10$$

$$3.5 = \frac{(95 - x)}{y} \times 10$$

$$3.5y + 10x = 950 \quad \text{...(ii)}$$

Now Mode = 34, which lies in class 30—40

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$34 = 30 + \frac{y - x}{2y - x - z} \times 10$$

$$4 = \frac{y - x}{2y - x - (200 - x - y)} \times 10$$

$$4 = \frac{y - x}{2y - x - 200 + x + y} \times 10$$

⇒ $$4 = \frac{y - x}{3y - 200} \times 10$$

$$4(3y - 200) = 10y - 10x$$

⇒ $2y + 10x = 800$ ...(iii)

Solving (ii) and (iii),

$$3.5y + 10x = 950$$
$$2y + 10x = 800 \quad \text{[Subtracting]}$$
$$-\quad -\quad -$$
$$1.5y = 150$$

∴ $y = \frac{150}{1.5} = 100$

Put the value of $y = 100$ in (ii)

$$3.5(100) + 10x = 950$$
$$350 + 10x = 950$$
$$10x = 600$$
$$x = 60$$

Now, $z = 200 - x - y$

$= 200 - 60 - 100 = 40$

Thus, $x = 60$, $y = 100$, $z = 40$.

Example 71. Calculate the mode from the following data:

Class:	100—200	200—300	300—400	400—5000	500—600
f:	27	9	7	3	2

Solution:

Class	f
100—200	27 (f_1)
200—300	9 (f_2)
300—400	7
400-500	3
500—600	2

By inspection, modal class is 100—200

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 100 + \frac{27-0}{2 \times 27 - 0 - 9} \times 100 \quad \text{[Here, } f_0 = 0]$$

$$= 100 + \frac{27}{54-9} \times 100 = 100 + \frac{2700}{45}$$

$$= 100 + 60 = 160$$

Example 72. Find the mode from the following data:

Class:	155—157	157—159	159—161	161—163	163—165
f:	4	8	26	53	89

Solution:

Class	f
155—157	4
157—159	8
159—161	26
161—163	53 (f_0)
163—165	89 (f_1)

By inspection, modal class is 163—165. [Here, $f_2 = 0$]

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 163 + \frac{89-53}{2 \times 89 - 53 - 0} \times 2$$

$$= 163 + \frac{72}{125} = 163.576$$

Graphical Location of Mode

The value of mode in a frequency distribution can be located graphically by means of histogram. It involves the following steps:

(*i*) Present the given data in the form of a histogram.

(*ii*) Find the rectangle whose height is the highest. This will be the modal class.

(*iii*) Join the top corners of the modal rectangle with immediately next corners of the adjacent rectangles.

(*iv*) Locate the point where the joining lines intersect with each other.

(*v*) Then draw a perpendicular line from this point of intersection on the X-axis.

(*vi*) The point where this perpendicular line meets the X-axis gives us the value of the mode.

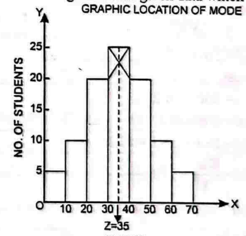
The following examples would illustrate the procedure.

Example 73. Determine the value of mode graphically from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of students:	5	10	20	25	20	10	5

Verify the result by mathematical calculations.

Solution: First we draw the histogram of the given data which is given below:



It is clear from the histogram that the value of the mode is 35.

Calculation of Mode Using Formula

By inspection, mode lies in the class 30—40

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Substituting the values, we get

$$= 30 + \frac{25-20}{50-20-20} \times 10 = 30 + \frac{5}{10} \times 10$$

$$= 35$$

Empirical Relation between Mean, Median and Mode

In a moderately asymmetrical distribution, the difference between $\bar{X}$ and Z is 3 times the difference between $\bar{X}$ and M.

$$\bar{X} - Z = 3(\bar{X} - M)$$

This equation can be expressed as

(a) $Z = 3M - 2\bar{X}$, (b) $M = \frac{1}{3}(2\bar{X} - Z)$, and (c) $\bar{X} = \frac{1}{2}(3M - Z)$.

Example 74. In an asymmetrical distribution, the arithmetic mean and median are respectively 30 and 32. Calculate the mode.

Solution: As we know

$$Z = 3M - 2\bar{X}$$

Given: $M = 32$, $\bar{X} = 30$

$$Z = 3 \times 32 - 2 \times 30$$
$$= 96 - 60 = 36$$

$\therefore$ $Z = 26$

Example 75. The mode and mean are 26.6 and 28.1 respectively in a moderately skewed distribution. Find out the median.

Solution: Given: $Z = 26.6$, $\bar{X} = 28.1$

$$Z = 3M - 2\bar{X}$$

Substituting the values,

$$26.6 = 3M - 2 \times 28.1$$

$\Rightarrow$ $26.6 = 3M - 56.2$

$\therefore$ $3M = 26.6 + 56.2 = 82.8$

$\Rightarrow$ $M = \frac{82.8}{3} = 27.6$

Median value is 27.6.

Merits and Demerits of Mode

Merits:

(*i*) It is easy to understand and simple to calculate.
(*ii*) It is not affected by extreme values.
(*iii*) It can be located graphically with the help of histogram.
(*iv*) It can be easily calculated in case of open ended classes.
(*v*) All the frequencies are not needed for its calculation.
(*vi*) It is true representative of frequency distribution, since it is the value which occurs most frequently.

Demerits:

(*i*) It is not suited to algebraic treatment.
(*ii*) There can be bi-modal frequency series.
(*iii*) It is not based on all the items of the series.
(*iv*) It is not rigidly defined.
(*v*) It has no mathematical property.

EXERCISE 5.7

1. Calculate the mode of the following data:
8, 10, 5, 8, 12, 7, 8, 9, 11, 7 [Ans. $Z = 8$]

2. Calculate the mode of the following data:
(*i*) 10, 15, 20, 25, 30, 35
(*ii*) 7, 8, 10, 15, 10, 22, 20, 26, 20, 34, 20, 6, 10 [Ans. (*i*) No Mode (*ii*) Bi-modal series]

3. Calculate the mode (by grouping method) of the following data:

Wages (in Rs.):	20	25	30	35	40	45	50	55
No. of workers:	1	3	5	9	14	10	6	4

[Ans. Z = 40]

4. Calculate the median and mode of the following data:

Size:	5	10	15	20	25	30	35	40
Frequency:	3	7	15	25	20	19	16	4

[Ans. M = 25, Z = 20]

5. Calculate the mode from the following table by grouping method:

Value:	0—5	5—10	10—15	15—20	20—25	25—30	30—35
Frequency:	1	2	10	4	10	9	2

[Ans. Z = 24.83]

6. Determine the missing frequencies when mode = 36 and total frequency is = 30.:

C.I :	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
f:	—	5	12	—	2

[Ans. $f_1 = 3.67 \approx 4$, $f_2 = 7.33 \approx 7$, the frequencies cannot be in fractions]

7. Calculate Median and Mode from the following:

Income between (Rs.)	No. of Persons
100 and 200	15
100 and 300	33
100 and 400	63
100 and 500	83
100 and 600	100

[Ans. M = 356.67, Z = 354.54]

8. Calculate the median and mode from the following data:

Size:	6—10	11—15	16—20	21—25	26—30
Frequency:	20	30	50	40	10

[Ans. M = 18, Z=18.83]

9. Calculate median and mode (by grouping method) from the following series:

Central Size:	5	10	15	20	25	30	35	40	45
Frequency:	7	13	19	24	32	28	17	8	6

[Ans. M=24.68, Z=25.83]

10. The frequency distribution of marks obtained by 60 students of a 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	16	2

(*i*) Draw Histogram for the distribution and find the modal value.

(*ii*) Draw a cumulative frequency curve and find the marks limits of the middle 50% students.

[Ans. (*i*) Z=47.5, (*ii*) Q_1 =42.5, Q_3 = 52; $Q_3 - Q_1$ = 9.5]

11. The median and mode of the following wage distribution of 230 persons are known to be Rs. 335 and Rs. 340 respectively. Three frequency values from the table are, however, missing. Find out the missing frequencies:

Wages (Rs.):	0—100	100—200	200—300	300—400	400—500	500—600	600—700
No. of persons:	4	16	60	–	–	–	4

[Hint: See Example 70] [Ans. 100, 40, 6]

12. Find out missing frequencies in the following incomplete distribution:-

Class Interval :	0—10	10—20	20—30	30—40	40—50
Frequency:	3	?	20	12	?

The values of Median and Mode are 27 and 26 respectively. [Ans. $f_1 = 8, f_2 = 7, N = 50$

13. Calculate the mode (by grouping method) from the following data:

Monthly wages:	200—250	250—300	300—350	350—400	400—450	450—500	500—550	550—600
No. of workers:	4	6	20	32	33	17	8	2

[Hints: $Z = 3M - 2\overline{X}$] [Ans. Bi-modal series, $\overline{X}$=397.6, M=398.4, Z=4

14. If the mode and mean of a moderately asymmetrical series are 16 and 15.6 respectively, what would be its most probable mean? [Ans. M=15

15. If the mean and median of a moderately asymmetrical series are 26.8 and 27.9 respectively, what would be its most probable mode? [Ans. Z=3

COMBINED EXAMPLES ON MEAN, MEDIAN AND MODE

Example 76. From the data given below, find the mean, median and mode:

Variable	Frequency
10—13	8
13—16	15
16—19	27
19—22	51
22—25	75
25—28	54
28—31	36
31—34	18
34—37	9
37—40	7

Solution: **Calculation of Mean, Median and Mode**

Variable	f	M.V. (m)	d	$d' = \frac{d}{3}$	fd'	c.f.
10—13	8	11.5	–12	–4	–32	8
13—16	15	14.5	–9	–3	–45	23
16—19	27	17.5	–6	–2	–54	50
19—22	51	20.5	–3	–1	–51	101
22—25	75	23.5 A	0	0	0	176
25—28	54	26.5	+3	+1	54	230
28—31	36	29.5	+6	+2	72	266
31—34	18	32.5	+9	+3	54	284
34—37	9	35.5	+12	+4	36	293
37—40	7	38.5	+15	+5	35	300
	N = 300				$\Sigma fd' = 69$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 23.5 + \frac{69}{300} \times 3 = 23.5 + 0.69 = 24.19$$

$$\text{Median} = \text{Size of } \frac{N}{2} \text{ th item} = \frac{300}{2} = 150 \text{ th item.}$$

Median lies in the class 22—25

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 22 + \frac{150 - 101}{75} \times 3 = 22 + 1.96 = 23.96$$

By inspection, mode lies in the class 22—25

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 22 + \frac{75-51}{150-51-54} \times 3$$

$$= 22 + \frac{24 \times 3}{45} = 22 + 1.6 = 23.6$$

Thus, $\overline{X} = 24.19$, M = 23.96, Z = 23.6

Inclusive Series

Example 77. From the data given below, find the mean, median and mode:

Marks	No. of students
1—5	7
6—10	10
11—15	16
16—20	30
21—25	24
26—30	17
31—35	10
36—40	5
41—45	1

Solution: The given data is in inclusive form. Firstly we convert it into exclusive form.

Marks	f	M.V. (m)	d	$d' = \frac{d}{5}$	fd'	c.f.
0.5—5.5	7	3	–20	–4	–28	7
5.5—10.5	10	8	–15	–3	–30	17
10.5—15.5	16	13	–10	–2	–32	33
15.5—20.5	30	18	–5	–1	–30	63
20.5—25.5	24	23 = A	0	0	0	87
25.5—30.5	17	28	+5	+1	+17	104
30.5—35.5	10	33	+10	+2	+20	114
35.5—40.5	5	38	+15	+3	+15	119
40.5—45.5	1	43	+20	+4	+4	120
	N = 120				$\Sigma fd' = -64$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 23 + \frac{(-64)}{120} \times 5 = 23 - \frac{320}{120} = 23 - 2.67 = 20.33$$

$$M = \text{Size of} \left(\frac{N}{2}\right) \text{th item} = \frac{120}{2} = 60\text{th item.}$$

Median lies in the class 15.5—20.5

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 15.5 + \frac{60-33}{30} \times 50$$

$$= 15.5 + \frac{27 \times 5}{30} = 15.5 + \frac{135}{30} = 15.5 + 4.5 = 20$$

By inspection, mode lies in the class 15.5—20.5

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 15.5 + \frac{30-16}{60-16-24} \times 5$$

$$= 15.5 + \frac{70}{20} = 15.5 + 3.5 = 19$$

Thus, $\overline{X} = 20.33$, M = 20, Z = 19.

Example 78. Calculate mean, median and mode for the distribution from the data below:

Central Size	f
35	18
45	37
55	45
65	27
75	15
85	8

Solution: As the central size (mid-value) is given, we have to find the class intervals.

Calculation of Mean, Median and Mode

Classes	f	M.V. (m)	d	$d' = \frac{d}{10}$	fd'	c.f.
30—40	18	35	–20	–2	–36	18
40—50	37	45	–10	–1	–37	55
50—60	45	55 A	0	0	0	100
60—70	27	65	+10	+1	27	127
70—80	15	75	+20	+2	30	142
80—90	8	85	+30	+3	24	150
	N = 150				$\Sigma fd' = 8$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 55 + \frac{8}{150} \times 10 = 55 + 0.53 = 55.53$$

$$M = \text{Size of} \left(\frac{N}{2}\right) \text{th item} = \frac{150}{2} = 75\text{th item}$$

Median lies in the class 50—60

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 50 + \frac{75-55}{45} \times 10$$

$$= 50 + \frac{200}{45} = 50 + 4.44 = 54.44$$

By inspection, mode lies in the class 50—60

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 50 + \frac{45-37}{90-37-27} \times 10$$

$$= 50 + \frac{8}{26} \times 10 = 50 + \frac{80}{26} = 50 + 3.07 = 53.07$$

Thus, $\overline{X} = 55.53$, $M = 54.44$, $Z = 53.07$

Cumulative Frequency Series

Example 79. The grades of 36 students in an Auditing test are given in the following table:

Grades (Less than):	40	50	60	70	80	90	100
No. of students:	3	7	13	23	29	33	36

Find Mean, Median and Mode.

Solution: The given data is in cumulative frequency form. Firstly we convert it into simple frequency series.

Grades	f	M.V. (m)	d	$d' = \frac{d}{10}$	fd'	$c.f.$
30—40	3	35	–30	–3	–9	3
40—50	4	45	–20	–2	–8	7
50—60	6	55	–10	–1	–6	13
60—70	10	65	0	0	0	23
70—80	6	75	+10	+1	6	29
80—90	4	85	+20	+2	8	33
90—100	3	95	+30	+3	9	36
	$N = 36$				$\Sigma fd' = 0$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 65 + \frac{0}{36} \times 10 = 65$$

$$\text{Median} = \text{Size of} \left(\frac{N}{2}\right) \text{th item} = \frac{36}{2} = 18\text{th item.}$$

Median lies in the class 60—70

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 60 + \frac{18-13}{10} \times 10 = 60 + 5 = 65$$

By inspection mode lies in the class 60—70.

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 60 + \frac{10-6}{20-6-6} \times 10 = 60 + \frac{40}{8} = 60 + 5 = 65$$

Thus, Mean = 65, Median = 65 and Mode = 65.

IMPORTANT TYPICAL EXAMPLES ON MEAN, MEDIAN AND MODE

Example 80. Calculate mean, median and mode from the following data:

Marks:	10—20	10—30	10—40	10—50	10—60	10—70	10—80	10—90
No. of students:	4	16	56	97	124	137	146	150

Solution: The given data is in cumulative frequency form. It should first be converted into simple frequency.

Marks	f	M.V. (m)	d	d'	fd'	$c.f.$
10—20	4	15	–30	–3	–12	4
20—30	16 – 4 = 12	25	–20	–2	–24	16
30—40	56 – 16 = 40	35	–10	–1	–40	56
40—50	97 – 56 = 41	45	0	0	0	97
50—60	124 – 97 = 27	55	+10	+1	+27	124
60—70	137 – 24 =13	65	+20	+2	+26	137
70—80	146 – 137 = 9	75	+30	+3	+27	146
80—90	150 – 146 = 4	85	+40	+4	+16	150
	$N = 150$				$\Sigma fd' = 20$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 45 + \frac{20}{150} \times 10 = 45 + 1.33 = 46.33$$

$$\text{Median} = \text{Size of} \left(\frac{N}{2}\right) \text{th item} = \frac{150}{2} = 75\text{th item.}$$

Median lies in the class 40—50

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 40 + \frac{75-56}{41} \times 10 = 40 + \frac{19 \times 10}{41} = 40 + \frac{190}{41}$$

$$= 40 + 4.63 = 44.63$$

By inspection mode lies in the class 40—50

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 40 + \frac{41-40}{82-40-27} \times 10 = 40 + \frac{1 \times 10}{15} = 40 + 0.67 = 40.67$$

Thus, $\overline{X} = 46.33$, $M = 44.63$, $Z = 40.67$

Example 81. Calculate mean, median and mode from the following data:

X:	2	3	4	5—7	7—10	10—15	15—20	20—25
f:	1	2	2	3	5	10	8	4

Solution: First rewrite the given data in the form of equal classes and then calculate $\overline{X}$, Z and M

Class Intervals	f	M.V. (m)	d	d'	fd'	c.f.
0—5	1+2+2=5	2.5	–10	–2	–10	5
5—10	3+5=8	7.5	–5	–1	–8	13
10—15	10	12.5	0	0	0	23
15—20	8	17.5	+5	+1	8	31
20—25	4	22.5	+10	+2	8	35
	$\Sigma f = 35$				$\Sigma fd' = -2$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

$$= 12.5 + \frac{(-2)}{35} \times 5 = 12.5 - 0.285 = 12.215$$

By inspection, modal class is 10—15

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 10 + \frac{10-8}{2\times 10 - 8 - 8} \times 5$$

$$= 10 + \frac{2 \times 5}{4} = 10 + 2.5 = 12.5$$

$$\text{Median item} = \frac{N}{2} = \frac{35}{2} = 17.5\text{th item.}$$

Median class is 10—15

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 10 + \frac{17.5 - 13}{10} \times 5 = 10 + \frac{4.5}{10} \times 5$$

$$= 10 + 2.25 = 12.25$$

Thus, $\overline{X} = 12.21$, $Z = 12.5$, $M = 12.25$

Example 82. The arithmetic mean, the mode and the median of a group of 75 observations were calculated to be 27, 34 and 29 respectively. It was later discovered that one observation was wrongly read as 43 instead of the correct value 53. Examine to what extent the calculated values of the three averages will be affected by the error.

Solution: We know $\overline{X} = \frac{\Sigma X}{N}$ or $\Sigma X = N\overline{X}$

$$N = 75, \overline{X} = 27$$

$$\therefore \quad \Sigma X = 75 \times 27 = 2025$$

$$\text{Corrected } \Sigma X = 2025 - 43 + 53 = 2035$$

$$\text{Corrected } \overline{X} = \frac{2035}{75} = 27.13$$

Thus, the value of mean is affected. But the values of median and mode will not be affected by this error because these are positional averages and the median value is 29 which is far away from the values 43 and 53. Similarly, the value of mode would not be affected by the error since the modal value 34 is far away from the value 43 and 53.

EXERCISE 5.8

1. Calculate the arithmetic mean, median and mode for the data given below:

Weekly earnings (in Rs.):	50—53	53—56	56—59	59—62	62—65	65—68	68—71	71—74	74—77
No. of persons:	3	8	14	30	36	82	16	10	5

[**Ans.** $\overline{X}$=62.47, M=65.36, Z=66.23]

2. An incomplete distribution is given below:

X:	10—20	20—30	30—40	40—50	50—60	60—70	70—80
f:	12	30	?	65	?	25	18

Given that median value is 46 and $N = 229$

Calculate: (*i*) Missing frequencies

(*ii*) Mean, Mode, Quartiles, P_{54} and P_{85} from the completed table.

[**Ans.** $f_1 = 33.5 \simeq 34, f_2 = 45, \overline{X} = 45.82, Z = 46.07,$
$Q_1 = 34.48, Q_3 = 56.72, P_{54} = 47.33, P_{85} = 63.46$]

3. Compute $\overline{X}$, M and Z from the following data:

Variable:	10—15	10—20	10—25	10—30	10—35	10—40
Frequency:	5	12	25	35	43	50

[**Ans.** $\overline{X}$=25.5, M=25, Z=23.3]

4. Calculate mean, median and mode from the following data:

X:	0.5	1.0	3.0	5—7	7—10	10—15	15—16	16—20	20—25
f:	2	3	2	4	4	12	6	2	5

[**Hint:** Make equal intervals]

[**Ans.** $\overline{X}$=12, M=12.08, Z=12.5]

GEOMETRIC MEAN

Geometric mean is a mathematical average. The geometric mean is defined as the nth root of the product of all the n values of the variable. If there are two items in a series, we take the square root. If there are three items, we take the cube root and so on. Symbolically,

$$G.M. = \sqrt[n]{X_1 . X_2 . X_3 \ldots X_n}$$

Where, $X_1, X_2, X_3, \ldots$, etc., are the various values of the series, n-number of items and G.M. the geometric mean.

For example, if there are two items in a series and their values are 4 and 9, then their geometric mean is the square root of the product of the two items.

Thus, $G.M. = \sqrt{4 \times 9} = \sqrt{36} = 6$

Similarly, if there are three items is a series and their values are 2, 4, 8 respectively, their geometric mean will be the cube root of the product of three items. Thus,

$$G.M. = \sqrt[3]{2 \times 4 \times 8} = \sqrt[3]{64} = 4$$

However, if the number of items in a series is very large, it would be difficult to calculate the geometric mean by using the above formula. In order to facilitate the computation of geometric mean, logarithmic (log) values are used. In logarithmic form, geometric mean can be calculated by using the following formula:

$$G.M. = (X_1 . X_2 . X_3 \ldots X_n)^{\frac{1}{n}}$$

Taking logarithms on both sides,

$$\log G.M. = \frac{1}{n}[\log X_1 + \log X_2 + \ldots \log X_n]$$

$$\log G.M. = \frac{\Sigma \log X}{n}$$

$$\therefore \quad G.M. = \text{Antilog}\left[\frac{\Sigma \log X}{n}\right]$$

Note: The values of logs and antilog can be obtained from the log table and antilog tables.

Calculation of Geometric Mean

Individual Series

The formula used for calculating geometric mean in individual series is:

$$G.M. = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right]$$

Steps for Calculation

(i) Find the logarithms (logs) of the given values.

(ii) Find the sum total of logs, *i.e.*, $\Sigma \log X$

(iii) Divide $\Sigma \log X$ by the number of items (N) and calculate its antilog.

The result will give the value of geometric mean.

Example 83. Calculate geometric mean of the following series:

180, 190, 240, 386, 492, 662

Solution:

Calculation of Geometric Mean

X	$\log X$
180	2.2553
190	2.2788
240	2.3802
386	2.5866
492	2.6920
662	2.8209
$N = 6$	$\Sigma \log X = 15.0138$

$$G.M. = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left[\frac{15.0138}{6}\right]$$

$$\therefore \quad G.M. = \text{Antilog}\ [2.5023]$$

$$= 317.9$$

Example 84. Calculate geometric mean of the following series:

2574, 475, 75, 5, .8, .08, .005, .0009.

Solution:

Calculation of Geometric Mean

X	$\log X$
2574	3.4106
475	2.6767
75	1.8751
5	0.6990
0.8	$\bar{1}.9031$
0.08	$\bar{2}.9031$
0.005	$\bar{3}.6990$
0.0009	$\bar{4}.9542$
$N = 8$	$\Sigma \log X = 2.1208$

$$G.M. = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left(\frac{2.1208}{8}\right)$$

$$= \text{Antilog } 0.2651 = 1.841$$

Hence, G.M. = 1.841

Example 85. Calculate geometric mean of the following series:
0.9841, 0.3154, 0.0252, 0.0068, 0.0200, 0.0002, 0.5444, 0.4010

Solution:

Calculation of Geometric Mean

X	$\log X$
0.9841	$\bar{1}.9931$
0.3154	$\bar{1}.4989$
0.0252	$\bar{2}.4014$
0.0068	$\bar{3}.8325$
0.0200	$\bar{2}.3010$
0.0002	$\bar{4}.3010$
0.5444	$\bar{1}.7359$
0.4010	$\bar{1}.6031$
	$\Sigma \log X = \overline{11}.6669$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left[\frac{\overline{11}.6669}{8}\right] = \text{Antilog}\left[\frac{\overline{16}+5.6669}{8}\right]$$

$$= \text{Antilog}[\bar{2}.7084] = 0.0511$$

Discrete Series

The formula used for calculating geometric mean in discrete series is:

$$\textbf{G.M.} = \textbf{Antilog}\left[\frac{\Sigma f \log X}{N}\right]$$

Steps for Calculation

(1) Find the logarithms (logs) of the given values, *i.e.*, log X.

(2) Multiply the frequency of each item with the log $x$and obtain the total $\Sigma f \log X$.

(3) Divide $\Sigma f \log X$ by the total number of frequency (*i.e.*, $\Sigma f = N$) and calculate its antilog. The result will give the value of geometric mean.

Example 86. From the following data, calculate geometric mean:

Size of items (X):	6	7	8	9	10	11	12
Frequency (f):	8	12	18	26	16	12	8

Solution:

Calculation of Geometric Mean

X	f	$\log X$	$f \log X$
6	8	0.7782	6.2256
7	12	0.8451	10.1412
8	18	0.9031	16.2558
9	26	0.9542	24.8092
10	16	1.0000	16.0000
11	12	1.0414	12.4968
12	8	1.0792	8.6336
	$n = 100$		$\Sigma f \log X = 94.5622$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma f \log X}{N}\right] = \text{Antilog}\left[\frac{94.5622}{100}\right]$$

$$= \text{Antilog}\,(0.945622) = 8.822$$

$$\therefore \quad \text{G.M.} = 8.822$$

Continuous Series

The formula used for calculating geometric mean in continuous series is:

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma f \log m}{N}\right]$$

Where, m = mid-value of various classes, N = Total frequency.

Steps for Calculation

(1) Find the logarithms (logs) of the mid-values of the various classes.

(2) Multiply the frequency of each class with log m and obtain the total $\Sigma f \log m$.

(3) Divide $\Sigma f \log m$ by total number of frequency (*i.e.*, $\Sigma f = N$) and calculate its antilog.

Example 87. Calculate geometric mean of the data given below:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	3	4	6	3	2

Solution:

Calculation of Geometric Mean

Marks	M.V. (m)	f	$\log m$	$f \log m$
0—10	5	3	0.6990	2.0970
10—20	15	4	1.1761	4.7044
20—30	25	6	1.3979	8.3874
30—40	35	3	1.5441	4.6323
40—50	45	2	1.6532	3.3064
		$N = 18$		$\Sigma f \log m = 23.1275$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma f \log m}{N}\right] = \text{Antilog}\left[\frac{23.1275}{18}\right]$$

$$= \text{Antilog}\,[1.28486] = 19.27$$

Combined Geometric Mean

If G_1 and G_2 are the G.M. of two groups having N_1 and N_2 items, then the G.M. of the combined group is given by:

$$G_{12} = \textbf{Antilog}\left[\frac{N_1 \log G_1 + N_2 \log G_2}{N_1 + N_2}\right]$$

Note: The formula can be extended to three or more groups.

Example 88. The G.M. of a sample of 10 items was found to be 20 and that of a sample of 20 item was found to be 15. Find the combined G.M.

Solution: Here, $N_1 = 10,\ G_1 = 20$

$N_2 = 20,\ G_2 = 15$

$$G_{12} = \text{Antilog}\left[\frac{N_1 \log G_1 + N_2 \log G_2}{N_1 + N_2}\right]$$

$$= \text{Antilog}\left[\frac{10 \times \log 20 + 20 \times \log 15}{10 + 20}\right]$$

$$= \text{Antilog}\left[\frac{10(1.3010) + 20(1.1761)}{30}\right]$$

$$= \text{Antilog}\left[\frac{13.01 + 23.52}{30}\right] = \text{Antilog}\left[\frac{36.53}{30}\right]$$

$$= \text{Antilog}\,[1.2177] = 16.51$$

IMPORTANT TYPICAL EXAMPLES

Example 89. The G.M. of 10 observations on a certain variable was calculated as 16.2. It was la discovered that one of the observations was wrongly recorded as 12.9, in fact, it w 21.9. Apply appropriate correction and calculate the correct G.M.

Solution: Let $x_1\, x_2 \ldots x_{10}$ be ten items.

$$\therefore \quad G.M. = \sqrt[10]{x_1 x_2 \ldots x_{10}} = (x_1 x_2 \ldots x_{10})^{\frac{1}{10}}$$

$$\Rightarrow \quad (16.2)^{10} = x_1 x_2 \ldots x_{10}$$

$$\Rightarrow \quad x_1 x_2 \ldots x_{10} = (16.2)^{10}$$

$$\text{Corrected } x_1 x_2 \ldots x_{10} = (16.2)^{10} \times \frac{\text{Correct Item}}{\text{Incorrect Item}} = (16.2)^{10} \times \frac{21.9}{12.9}$$

$$\text{G.M. (Corrected)} = [\text{Corrected } (x_1 x_2 \ldots x_{10})]^{\frac{1}{10}} = \left[(16.2)^{10} \times \frac{21.9}{12.9}\right]^{\frac{1}{10}}$$

$$\text{G.M. log (Corrected)} = \frac{1}{10}[10 \log 16.2 + \log 21.9 - \log 12.9]$$

$$= \frac{1}{10}[12.09 + 1.3404 - 1.1105]$$

$$= 1.23249$$

$$\therefore \quad \text{Corrected G.M.} = \text{Antilog}\,[1.23249]$$

$$= 17.08$$

Special Uses of Geometric Mean

Geometric mean is specially useful in the following cases:

(1) Geometric mean is generally used to find the average percentage increase in population, prices, sales, production or national income, etc.

(2) As geometric mean measures relative change, it finds special application in index numbers.

(3) This average is most suitable when large weights have to be given to small items and small weights to large items.

The following examples illustrate the uses of geometric mean:

Example 90. Find the average rate of increase in population which in the first decade has increased by 20%, in the second decade by 30% and in the third decade by 40%.

Solution: Suppose the population at the beginning of the decade is 100.

Calculation of Geometric Mean

Decade	Per cent rise	Population at the end of decade (X)	log x
1st	20	120	2.0792
2nd	30	130	2.1139
3rd	40	140	2.1461
$N = 3$			$\Sigma \log x = 6.3392$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left[\frac{6.3392}{3}\right]$$

$$= \text{Antilog}\,(2.1131) = 129.7$$

The average rate of increase in population is $129.7 - 100 = 29.7$ per cent per decade.

Example 91. The annual rates of growth of an economy over the last five years were: 1.5, 2.7, 3.0, 4.5 and 6.2 per cent respectively. What is the compound rate of growth per annum for the 5 year period?

Solution: Suppose the price at the beginning of the year is 100. Apply the geometric mean:

Years	Annual rate of growth	Growth rate at the end of the year (X)	log X
1	1.5	101.5	2.0064
2	2.7	102.7	2.0116
3	3.0	103.0	2.0128
4	4.5	104.5	2.0191
5	6.2	106.2	2.0261
$N = 5$			$\Sigma \log X = 10.076$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{n}\right] = \text{Antilog}\left[\frac{10.076}{5}\right]$$

$$= \text{Antilog}\,(2.0152) = 103.5$$

The compound rate of growth per annum $= 103.5 - 100 = 3.5$

Example 92. A machinery depreciates by 40% in the first year, 25% in the second and 10% per annum during the next three years. What is the average rate of depreciation during the whole period?

Solution: Suppose the value of machine in the initial year is 100

Year	Per cent decline	Diminishing value at the end of the year (X)	log X
I	40	100 – 40 = 60	1.7782
II	25	100 – 25 = 75	1.8751
III	10	100 – 10 = 90	1.9542
IV	10	100 – 10 = 90	1.9542
V	10	100 – 10 = 90	1.9542
N = 5			Σ log X = 9.5159

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left[\frac{9.5159}{5}\right]$$

$$= \text{Antilog}\,[1.9032] = 80.02 = 80$$

Average rate of depreciation = 100 – 80 = 20% per year.

Average Rate of Growth of Population

Geometric mean is also used to compute the average annual per cent increase in population, prices when the values of the variables at the beginning of the first and at the end of the nth period are given. The average annual per cent increase may be computed by applying the formula:

$$P_n = P_0\,(1+r)^n$$

Where, r = the average rate of growth

n = Number of years

P_0 = Value at the beginning of the period

P_n = Value at the end of the period

Using logarithms, r can be calculated as:

$$r = \sqrt[n]{\frac{P_n}{P_0}} - 1$$

or

$$r = \text{Antilog}\left[\frac{\log P_n - \log P_0}{n}\right] - 1$$

Note : The expression $P_n = P_0(1+n)^n$ is called **compound interest formula.**

Example 93. The population of a country has increased from 84 million in 1983 to 108 million in 1993. Find the annual rate of growth of population.

Solution: Given, P_0 = 84 million, P_n = 108 million.

Let r be the rate of growth of population. Applying the formula,

$$r = \sqrt[n]{\frac{P_n}{P_0}} - 1 = \sqrt[10]{\frac{108}{84}} - 1$$

$$= \text{Antilog}\left[\frac{\log 108 - \log 84}{10}\right] - 1$$

$$= \text{Antilog}\left[\frac{2.0334 - 1.9243}{10}\right] - 1$$

$$= \text{Antilog}\left[\frac{0.1091}{10}\right] - 1$$

$$= \text{Antilog}\,[0.0109] - 1$$

$$= 1.026 - 1 = .026$$

$$= 2.6\%$$

Thus, the annual growth rate of population is 2.6%.

Example 94. The population of a town was 10,000. At first it increased at the rate of 3% per annum for the first 3 years and then it decreased at the rate of 2% per annum for the next 2 years. What will be the population of the town after 5 years?

Solution: Given, $P_0 = 10{,}000, r_1 = 0.03, r_2 = 0.02, n_1 = 3, n_2 = 2, P_n = ?$

$$P_n = P_0(1+r_1)^3(1-r_2)^2 = 10{,}000(1+0.03)^3(1-0.02)^2$$

$$= 10{,}000(1.03)^3(0.98)^2 = 10{,}000(1.0927)(0.9604)$$

$$= 10{,}000 \times 1.0494 = 10{,}494$$

Example 95. The population of a city was 1,00,000 in 1980 and 1,44,000 in 1990. Estimate the population at the middle of 1980-1990.

Solution: $P_n = P_0(1+r)^n$

Here, $n = 10,\ P_0 = 1{,}00{,}000,\ P_{10} = 1{,}44{,}000$

$$144{,}000 = 1{,}00{,}000(1+r)^{10}$$

$$(1+r)^{10} = \frac{144}{100} = \left(\frac{12}{10}\right)^2$$

$$\Rightarrow \quad \left(\frac{12}{10}\right)^2 = [(1+r)^5]^2$$

$$\Rightarrow \quad (1+r)^5 = \frac{12}{10} \qquad \text{...(i)}$$

Now, $P_5 = P_0(1+r)^5$

$$P_5 = 1{,}00{,}000(1+r)^5 \qquad \text{...(ii)}$$

Put the value of $(1+r)^5$ from (i) in (ii), we get

$$P_5 = 1{,}00{,}000 \left(\frac{12}{10}\right)$$

$$P_5 = 1{,}20{,}000$$

Aliter: Population at middle of the decade as

$$\Rightarrow \quad \text{G.M.} = \sqrt{1{,}44{,}000 \times 1{,}00{,}000}$$

$$= \sqrt{14400000000} = 120{,}000$$

Example 96. The population of a capital of a State was 20 lakh in 1981. It went up to 25 lakhs in 1991. What is the average percentage increase per year? What will be the size of population in 2006 if it continued to increase at the same rate?

Solution:

$$P_n = P_0(1+r)^n$$

Here, $n = 10$, $P_0 = 20$, $P_{10} = 25$

$$25 = 20(1+r)^{10}$$

Taking log on both sides,

$$10 \log(1+r) = \log 1.25$$

$$\log(1+r) = \frac{0.0969}{10} = 0.00969$$

$$1 + r = \text{Antilog}\,[0.00969] = 1.023$$

$$r = 1.023 - 1 = 0.023$$

or $\quad r = 0.023 \times 100 = 2.3\%$

Population for 2006, $P_0 = 20, r = .023, N = 25$

$$P_n = P_0(1+r)^{25}$$

$$= 20\,(1 + 0.23)^{25}$$

Taking log on both sides,

$$\log P_n = \log 20 + 25 \log 1.023$$

$$\log P_n = 1.3010 + 25 \times 0.0099$$

$$= 1.3010 + 0.2475 = 1.5485$$

$$P_n = \text{Antilog}\,[1.5485]$$

$$= 35.36 \text{ lakhs}$$

Weighted Geometric Mean

Like weighted arithmetic mean, we can also calculate weighted geometric mean by using the following formula:

$$\textbf{Weighted G.M.} = \textbf{Antilog}\left[\frac{\Sigma W \log X}{\Sigma W}\right]$$

Where, $W_1, W_2 \ldots W_n$ are the weights assigned to different values $X_1, X_2 \ldots X_n$.

Example 97. From the following data, calculate weighted geometric mean:

Items	Index No.	Weights
Food	120	7
Rent	110	5
Clothing	125	3
Fuel	105	2
Others	140	3

Solution:

Calculation of Weighted G.M.

Items	X	W	$\log X$	$W \cdot \log X$
Food	120	7	2.0792	14.554
Rent	110	5	2.0414	10.2070
Clothing	125	3	2.0969	6.2907
Fuel	105	2	2.0212	4.0424
Others	140	3	2.1461	6.4383
		$\Sigma W = 20$		$\Sigma W \log X = 41.5328$

$$\text{Weighted G.M.} = \text{Antilog}\left[\frac{\Sigma W \log X}{\Sigma W}\right] = \text{Antilog}\left[\frac{41.5328}{20}\right]$$

$$= \text{Antilog}\,(2.0766) = 119.3$$

Merits and Demerits of Geometric Mean

Merits:

(1) It is rigidly defined.

(2) It is based on all the items of the series.

(3) It is suitable for further algebraic treatment.

(4) It gives less weight to large items.

(5) It possesses the merits of sampling stability.

(6) It is the best measure of ratios of change.

(7) It is least affected by extreme items.

Demerits:

(1) It is difficult to calculate as one has to make use of logs and antilogs to calculate geometric mean.

(2) Geometric mean cannot be calculated unless the values of all the items are known.

(3) It may be a value which does not exit in the series.

(4) If value of any one item is zero, then geometric mean will also become zero. Again if one item has negative value, G.M. becomes indeterminate.

EXERCISE 5.9

1. Calculate the geometric mean from the following series:
   85, 70, 15, 75, 500, 8, 45, 250, 40, 36 **[Ans.** G.M.=58.03]
2. Calculate the geometric mean from the following series:

X:	8	10	12	14	16	18
f:	6	10	20	8	5	1

[Ans. G.M.=11.71]

3. Find the geometric mean for the data given below:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of persons:	4	8	10	6	7

[Ans. G.M.=22.06]

4. The price of a commodity increased by 5% from 1985 to 1986, 8% from 1986 to 1987 pand 77% from 1987 to 1988. The average increase from 1985 to 1988 is quoted as 26% and not 30%. Explain and verify the result.
5. A machine was purchased for Rs. 10 lakh in 2000. An income tax assessee depreciated the machinery of his factory by 20 per cent in each of the first two years and 40 per cent in the third year. How much average depreciation relief should he claim from the taxation department? **[Ans.** 27.32%]
6. The geometric mean of 10 observations is 28.6. It was later discovered that one of the observation was recorded 23.4 instead of 32.4. Apply appropriate correction and calculate the correct Geometric mean.
   [Hint:See Example 89] **[Ans.** 29.54]
7. In 20 years, the population of a town increased from 10,000 to 20,000. Find the average annual rate of growth of population. **[Ans.** 3.5%]
8. The price of a commodity increased by 20% in the first year, decreased by 30% in the second year and increased by 40% in the third year. Find the average increase for these three years. **[Ans.** 5.56 %]
9. If the price of a commodity doubled in a period of 5 years, what is the average percentage increase per year? **[Ans.** 14.8%]
10. The number of bacteria in a certain culture was found to be 4×10^6 at noon of one day. At noon the next day the number was 9×10^6. If the number increased at a constant rate per hour, how many bacteria were there at the intervening mid-night?
    [Hint: $\sqrt{4\times10^6\times9\times10^6}$**]** **[Ans.** 6×10^6]
11. In 1950 and 1960 the population of US was 151.3 mw and 179.3 mw respectively (*i*) What was the average percentage increase per year (*ii*) Estimate the population in 1954 (*iii*) If the average percentage increase of population per year from 1960 to 1970 is the same as in (*ii*), what would be the population in 1970.

[Ans. (*i*) 1.7% (*ii*) 161.74 mw (*iii*) 212.71]

HARMONIC MEAN

The harmonic mean is a mathematical average. It is based on the reciprocal of the items. The harmonic mean of a series is defined as the reciprocal of the arithmetic average of the reciprocal of the values of its various items. Symbolically,

$$\text{H.M.} = \text{Reciprocal of } \frac{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots+\frac{1}{X_n}}{N}$$

$$= \frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots+\frac{1}{X_n}} = \frac{N}{\Sigma\left(\frac{1}{X}\right)}$$

Calculation of Harmonic Mean

Individual Series

In individual series, harmonic mean is calculated by using the following formula:

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

Where, $X_1, X_2, X_3, \ldots$, etc. refer to the various items of the variable.

Steps for Calculation

(1) Find out the reciprocals of the values of the series and add to these gets $\Sigma\left(\frac{1}{X}\right)$.

(2) Divide the sum total of reciprocals by the number of items, and find out its reciprocals. This gives the value of harmonic mean.

Example 98. Calculate the H.M. of the following:

2, 4, 7, 12, 19

Solution: **Calculation of H.M.**

X	Reciprocal ($1/X$)
2	0.5000
4	0.2500
7	0.1429
12	0.0833
19	0.0526
N = 5	$\Sigma\left(\frac{1}{X}\right)$=1.0288

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{1}{X}\right)} = \frac{5}{1.0288} = 4.86$$

Discrete Series

The formula used for calculating harmonic mean in discrete series is:

$$\text{H.M.} = \frac{N}{\Sigma f \times \frac{1}{X}}$$

Steps for Calculation

(1) Divide each frequency by the respective values of the variable.

(2) Obtain the total $\Sigma\left(f \times \frac{1}{X}\right)$

(3) Substitute the value of N and $\Sigma\left(f \times \frac{1}{X}\right)$ in the above formula.

Example 99. The following table gives the marks obtained by students in a class. Calculate the H.M.

Marks:	18	21	30	45
No. of students:	6	12	9	2

Solution:

Calculation of Harmonic Mean

Marks X	f	f/X
18	6	6/18 = 0.333
21	12	12/21 = 0.571
30	9	9/30 = 0.3000
45	2	2/45 0.0444
	N = 29	$\Sigma(f/X) = 1.248$

$$\text{H.M.} = \frac{N}{\Sigma(f/X)} = \frac{29}{1.248} = 23.237$$

Continuous Series

For calculating harmonic mean in continuous series, the following formula is used:

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

Where, m = mid-value of various classes, N = the total frequency.

Steps for Calculation

(1) Obtain the mid-value of each class and denote it by m.

(2) Divide each frequency by the respective mid-value of the class.

(3) Obtain the total $\Sigma\left(f \times \frac{1}{m}\right)$.

(4) Substitute the value of N and $\Sigma\left(f \times \frac{1}{m}\right)$ in the above formula.

Example 100. Calculate the H.M. for the following:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	4	7	28	12	9

Solution:

Calculation of Harmonic Mean

Class	f	Mid-value (m)	f/m
0—10	4	5	4/5 = 0.8000
10—20	7	15	7/15 = 0.4667
20—30	28	25	28/25 = 1.1200
30—40	12	35	12/35 = 0.3429
40—50	9	45	9/45 = 0.2000
	$N = 60$		$\Sigma\left(f \times \frac{1}{m}\right) = 2.9296$

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{f}{m}\right)} = \frac{60}{2.9296} = 20.48$$

Weighted Harmonic Mean

If all the values of the variable are not of equal importance, *i.e.*, are of varying importance, then we calculate weighted H.M. which is given as:

$$\text{Weighted H.M.} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$$

Where, $W_1, W_2, \ldots W_n$ are the weights of the corresponding values $X_1, X_2, \ldots X_n$ of the variables.

Example 101. Find the weighted H.M. of the items 4, 7, 12, 19, 25 with weights 1, 2, 1, 1, 1 respectively.

Solution:

Calculation of Weighted H.M.

X	W	W/X
4	1	0.2500
7	2	0.2857
12	1	0.0833
19	1	0.0526
25	1	0.0400
	$\Sigma W = 6$	$\Sigma\left(\frac{W}{X}\right) = 0.7116$

$$\text{Weighted H.M.} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)} = \frac{6}{0.7116} = 8.4317$$

Uses of Harmonic Mean

H.M. is of very limited use. It is useful in finding averages involving speed, time, price and ratios. The following examples illustrate the uses of H.M.:

Example 102. An aeroplane covers the four sides of a field at speeds of 1000, 2000, 3000 and 4000 km per hour respectively. What is the average speed of the plane in its flight around the field?

Solution: In this question simple H.M. is used which is calculated as:

$$\text{Average Speed} = \frac{4}{\frac{1}{1000}+\frac{1}{2000}+\frac{1}{3000}+\frac{1}{4000}}$$

$$= \frac{4}{\frac{12+6+4+3}{12000}} = \frac{4\times 12000}{25} = 1920 \text{ kms/hrs}$$

Example 103. A cyclist covers first three kms at an average speed of 8 km per hour, another 2 kms at a speed of 3 km per hour and the last two kms at a speed of 2 km per hour. Find the average speed for the entire journey and verify your answer.

Solution: In this case weighted harmonic mean is used with distance covered as weight. Denoting the speed by X and the distance covered by W.

Speed km/hr (X)	Distance (W)	$\frac{W}{X}$
8	3	3/8 = 0.375
3	2	2/3 = 0.667
2	2	2/2 = 1.000
	$\Sigma W = 7$	$\Sigma W/X = 2.042$

$$\text{Average speed} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)} = \frac{7}{2.042} = 3.42 \text{ km/hr.}$$

Thus, the average speed for the entire journey is 3.42 km/hr.

Verification:

Distance (km)	Speed (km. per hr.)	Time = $\frac{Distance}{Speed}$
3	8	3/8 = 0.375 hrs.
2	3	2/3 = 0.667 hrs.
2	2	2/2 = 1 hrs.
$\Sigma D = 7$		$\Sigma T = 2\cdot 042$

$$\text{Average speed} = \frac{\text{Total Distance Travelled}}{\text{Total Time Taken}} = \frac{7}{2\cdot 042} = 3.42 \text{ km/hr.}$$

Hence, verified.

Example 104. A bus runs 20 km at an average speed of 30 km per hour and then runs for 30 minutes at a speed of 60 km per hour and finally runs for 15 minutes at a speed of 20 km per hour. Find the average speed of the bus over the entire journey.

Solution: In 60 minutes bus travels = 60 km

In 1 minute, bus travels $= \frac{60}{60}$

In 30 minutes, bus travels $= \frac{60}{60} \times 30 = 30$ km

Similarly, in 15 minutes, distance travelled will be $= \frac{20}{60} \times 15 = 5$ km

Since, the distance travelled is unequal, so the required average speed will be weighted H.M..

X (Speed in km)	W (Distance)	$1/X$	$W \times \frac{1}{X}$
30	20	1/30	20/30 = 2/3
60	30	1/60	30/60 = 1/2
20	5	1/20	5/20 = 1/4
	$\Sigma W = 55$		$\Sigma\left(\frac{W}{X}\right) = 17/12$

$$\text{H.M.} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$$

$$= \frac{55}{\frac{17}{12}} = \frac{55}{1} \times \frac{12}{17} = 38.82 \text{ km / hr.}$$

Example 105. A man travelled by car for 3 days. He covered 480 kms each day. On the first day he drove for 10 hours at 48 kms an hour, on the second day he drove for 12 hours at 40 kms an hour and on last day he drove for 15 hours at 32 kms. What was his average speed?

Solution: Since, the distance travelled is constant, *i.e.*, 480 kms each day, the appropriate average is the Harmonic Mean.

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

$$= \frac{3}{\frac{1}{48}+\frac{1}{40}+\frac{1}{32}} = \frac{3}{\frac{37}{480}} = \frac{3\times 480}{37}$$

= 38.92 km/hr.

Example 106. In a certain factory, a unit of work is completed by A in 4 minutes, by B in 5 minutes, by C in 6 minutes, by D in 10 minutes and by E in 12 minutes. What is their average rate of working? What is the average number of units of work completed per minute? At this rate how many units will they complete in a six hour day?

Solution: (*i*) Average rate of working

$$\text{H.M.} = \frac{5}{\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{10}+\frac{1}{12}} = \frac{5}{\frac{15+12+10+6+5}{60}}$$

$$= 5 \times \frac{60}{48} = 6.25 \text{ minutes per unit.}$$

(*ii*) Average number of unit of work completed per minute $= \frac{1}{6.25} = 0.16$ units.

(*iii*) Total units of work completed within 6 hours (360 minutes) by 5 workers

$$= 360 \times 0.16 \times 5 = 288 \text{ units.}$$

Example 107. Typist A can type a letter in 5 minutes, B in 10 minutes and C in 15 minutes. What is the average number of letters typed per hour per typist?

Solution: Average time to type one letter

$$\text{H.M.} = \frac{3}{\frac{1}{5}+\frac{1}{10}+\frac{1}{15}} = \frac{3}{\frac{6+3+2}{30}} = \frac{3\times 30}{11} = \frac{90}{11}$$

$$= 8.18 \text{ minutes per letter.}$$

Number of letters typed in 60 minutes $= \frac{60}{8.18} = 7.33$

◘ CHOICE BETWEEN HARMONIC MEAN AND ARITHMETIC MEAN

The harmonic mean, like arithmetic mean, is also used in calculating average price. To explain the method of choosing an appropriate average, consider the following example:

Let the price of a commodity be Rs. 3, 4, and 5 per unit in three successive years. If we take A.M. of these prices, *i.e.*, $\frac{3+4+5}{3} = 4$, then it will denote average price when equal quantities of the commodity are purchased in each year. To verify this, let us assume that 10 units of commodity are purchased in each year.

∴ Total Expenditure on the commodity in 3 years $= 10 \times 3 + 10 \times 4 + 10 \times 5$

Also, average price $= \frac{\text{Total Expenditure}}{\text{Total Quantity Purchased}} = \frac{10\times 3+10\times 4+10\times 5}{10+10+10} = \frac{3+4+5}{3} =$ Rs. 4

which is arithmetic mean of the prices in three years.

Further, if we take harmonic mean of the given prices, *i.e.*, $\frac{3}{\frac{1}{3}+\frac{1}{4}+\frac{1}{5}} =$ Rs. 3.27, it will denote the average price when equal amounts of money are spent on the commodity in three years. To verify this let us assume that Rs. 100 is spent in each year on the purchase of the commodity.

$$\text{Average Price} = \frac{\text{Total Expenditure}}{\text{Total Quantity Purchased}} = \frac{300}{\frac{100}{3}+\frac{100}{4}+\frac{100}{5}} = \frac{3}{\frac{1}{3}+\frac{1}{4}+\frac{1}{5}} = \text{Rs. } 3.27$$

Next, we consider a situation where different quantities are purchased in the three years. Let us assume that 10, 15 and 20 units of the commodity are purchased at price of Rs. 3, 4, and 5 respectively.

$$\text{Average Price} = \frac{\text{Total Expenditure}}{\text{Total quantity purchased}} = \frac{3\times 10+4\times 15+5\times 20}{10+15+20}$$

which is the weighted arithmetic mean (A.M.) of the prices taking respective quantities as weights. Further, if Rs. 150, 200 and 250 are spent on the purchase of the commodity at price of Rs. 3, 4 and 5 respectively, then,

Average Price $= \frac{150+200+250}{\frac{150}{3}+\frac{200}{4}+\frac{250}{5}}$, where, $\frac{150}{3}, \frac{200}{4}$ and $\frac{250}{5}$ are the quantities purchased in respective situations.

The above average price is equal to the weighted harmonic mean of prices taking money spent as weights.

From the above example, it is clear that arithmetic means is appropriate for calculating average price $\left(\frac{\text{Money}}{\text{Quantity}}\right)$ where quantities purchased in different situations are given. Similarly harmonic mean will be appropriate when sums of money spent in different situations are given. Further, if the equal quantity is purchased or equal sums of money is spent in different situations, we use simple A.M. or simple H.M. otherwise we use weighted A.M. or H.M.

Example 108. A scooterist purchased petrol at the rate of Rs. 24, Rs. 29.50 and Rs. 36.85 per litre during three successive years. Calculate the average price of petrol:

(a) If he purchased 150, 180 and 195 litres of petrol in the respective years.

(b) If he spent Rs. 3,850, Rs. 4,675 and Rs. 5,825 in three years.

Solution: Average Price $= \frac{\text{Money}}{\text{Quantity}}$

(a) Since, the condition is given in terms of different litres of petrol in three years, therefore, weighted A.M. will be appropriate to calculate the average price.

Price Per Unit (X)	Quantity (W)	WX
24.00	150	3600
29.50	180	5310
36.85	195	7185.75
Total	525	16,095.75

$$\text{Average price} = \frac{\text{Money Spent}}{\text{Quantity}} = \frac{\Sigma WX}{\Sigma W}$$

$$= \frac{3600+5310+7185.75}{525} = \frac{16095.75}{525} = \text{Rs. } 30.65$$

(b) Since, the condition is given in terms of different sums of money spent in three years, therefore, weighted H.M. will be appropriate to calculate the average price.

Price Per Unit (X)	Money Spent (W)	$\frac{W}{X}$
24.00	3850	$\frac{3850}{24}$
29.50	4675	$\frac{4675}{29.5}$
36.85	5825	$\frac{5825}{36.85}$

$$\text{Average price} = \frac{\text{Money Spent}}{\text{Quantity}} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$$

$$= \frac{3850+4675+5825}{\frac{3850}{24}+\frac{4675}{29.5}+\frac{5825}{36.85}} = \frac{14350}{160.41+158.47+158.07}$$

$$= \frac{14350}{476.96} = \text{Rs. } 30.09$$

Example 109. Kapil purchases oranges from one shop @ Rs. 2 per kg, from second shop @ Rs. 2.50 per kg, from third shop @ Rs. 3 per kg and from the fourth shop @ Rs. 3.50 per kg. Find the average price per kg if

(a) He purchases 5 kg oranges from each shop.

(b) He purchases oranges of Rs. 50 from each shop.

Solution: $\text{Average price} = \frac{\text{Money}}{\text{Quantity}}$

(a) Since, the condition is given the terms of same quantity from different shops, therefore, simple A.M. will be appropriate to be calculate average price.

Price Per Unit (X)	Quantity (W)
2.00	5
2.50	5
3.00	5
3.50	5

$$\text{Average price} = \frac{2+2.50+3+3.50}{4}$$

$$= \text{Rs. } 2.75$$

(b) Since, the condition is given in terms of equal sum of money spent, therefore, simple H.M. will be appropriate to calculate the average price.

Price per unit (X)	Money spent (W)
2.00	50
2.50	50
3.00	50
3.50	50

$$\text{Average Price} = \frac{4}{\frac{1}{2}+\frac{1}{2.50}+\frac{1}{3}+\frac{1}{3.50}} = \text{Rs. } 2.64$$

Example 110. An individual purchases three qualities of pencil. The relevant data are given below:

Quality	Price per pencil (Rs.)	Money spent
A	1.00	50
B	1.50	30
C	2.00	20

Calculate the average price per pencil.

Solution: Since, different sums of money spent in various situations are given, we shall calculate weighted H.M. to calculate average price.

Price per unit (X)	Money spent (W)
1.00	50
1.50	30
2.00	20
	$\Sigma W = 100$

$$\text{Weighted H.M.} = \frac{50+30+20}{\frac{50}{1.00}+\frac{30}{1.50}+\frac{20}{2.00}} = \frac{50+30+20}{50+20+10} = \text{Rs. } 1.25$$

Merits and Demerits of Harmonic Mean

Merits:
(1) It is rigidly defined.
(2) It is based on all the observations in the given data.
(3) It is capable of further algebraic treatment.
(4) In problems relating to time and rates, it gives better results than other averages.

Demerits:
(1) It is difficult to understand and calculate.
(2) It gives more weight to small items. This is not a desirable feature.
(3) Its value cannot be computed when there are both positive and negative items in a series or when one or more items are zero.

EXERCISE 5.10

1. Find the harmonic mean of the following series:
   3, 5, 6, 6, 7, 10, 12 [Ans. 5.9]
2. The following table gives marks obtained by a group of students in a test. Calculate the harmonic mean of the series:

Marks obtained:	20	21	22	23	24	25
No. of students:	4	2	7	1	3	1

[Ans. 21.9]

3. Calculate the H.M. for the following:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	4	7	28	12	9

[Ans. 20.48]

4. A cyclist covers first 30 km at an average speed of 80 km/hr, another 20 km at 30 km/hr and the last 20 km at 20 km/hr. Find the average speed for the entire journey and verify your answer. [Ans. 34.3 km/hr]
5. A cyclist covers successive quarters of a mile at the speed of 12, 10, 8 and 7 km/hr respectively. Find the average speed. [Ans. 8.8 km/hr]
6. A train runs 25 km at an average speed of 30 km/hr, another 50 km at a speed of 40 km/hr, then due to repair of the track travels for 6 minutes at a speed of 10 km/hr and finally covers the remaining distance of 24 km at a speed of 24 km/hr. What is the average speed in km/hr? [Ans. 31.41km/hr]
7. A man starts from rest and travels successive quarters of miles at an average speeds of 12, 16, 24 and 48 kms per hour. The average speed over the whole distance is 19.2 km/hr. and not 25 km./hr. Explain and show how you can verify the arithmetic.
8. Four persons A, B, C and D are working in a factory. A produces a unit of production in 12 minutes, B in 16 minutes, C in 24 and D in 48 minutes. Find the average rate of working. What is the average number of units completed per minute. At this rate how many units will they complete in 6 hours a day. [Ans. (*i*) 19.2 minutes per unit (*ii*) 0.052 units (*iii*) 74.88 units]
9. The interest paid on each of the three different sums of money yielding 10%, 12% and 15% simple interest p.a. is the same. What is the average yield per cent on the sum invested? **[Hint:** Use H.M.] [Ans. 12%]
10. A man having to drive 90 km wishes to achieve an average speed of 30 km per hour. For the first half of the journey, he averages only 20 km per hour. What must be his average speed for the second half of the journey if his overall average speed is to be 30 km per hour? [Ans. X_2 = 60 m.p.h]
11. A man buys mangoes from one shop at the rate of Rs. 20 per kg, from the second shop at the rate of Rs. 25 per kg, from the third shop at the of Rs. 30 per kg and from the fourth shop at the rate of Rs. 35 per kg. Find the average rate in rupees per kg if he buys mangoes of Rs. 10 from each shop. [Ans. Rs. 26.33 per Kg.]
12. A person buys kerosene at Rs. 0.80, Rs. 1.20, Rs. 1.80 and Rs. 2.80 per litre for four successive years. What was the average cost of oil when he spends Rs. 1000 each year and when he buys 1,000 litres every year? [Ans. Rs. 1.335 per litre, Rs. 1.65 per litre]

RELATIONSHIP BETWEEN A.M., G.M. AND H.M.

(*i*) **For any two positive numbers, G.M.** $= \sqrt{\textbf{A.M.} \times \textbf{H.M.}}$. This result can verified as

Let a and b be two positive numbers.

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

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

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

Hence, the result is proved.

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

$$\text{A.M.} > \text{G.M.} > \text{H.M.}$$

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

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

Example 111. If the A.M. of two numbers is 10 and their G.M. is 8, find their harmonic mean and two numbers.

Solution: Given: A.M. = 10, G.M. = 8

We also know that

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

$$\sqrt{(10)(\text{H.M.})} = (8)$$

Squaring both means, we get

$$\Rightarrow \quad \text{H.M.} = \frac{64}{10} = 6.4$$

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

$$\text{A.M.} = \frac{X_1 + X_2}{2} = 10 \quad \text{and} \quad \text{G.M.} = \sqrt{X_1.X_2} = 8$$

$$\Rightarrow \quad X_1 + X_2 = 20 \;...(i) \qquad \Rightarrow \quad X_1.X_2 = 64 \qquad ...(ii)$$

We can write,

$$(X_1 - X_2)^2 = (X_1 + X_2)^2 - 4X_1.X_2$$
$$= 400 - 256 = 144$$

$$\Rightarrow \quad X_1 - X_2 = 12 \qquad ...(iii)$$

Adding (*i*) and (*iii*), we get

$$2X_1 = 32 \quad \therefore \quad X_1 = 16$$

Also $X_2 = 4$

Example 112. Using the values 2, 4 and 8, verify that

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

Solution:

$$\text{A.M.} = \frac{2+4+8}{3} = 4.67 \text{ approx.}$$

$$\text{G.M.} = 3\sqrt{2\times4\times8} = (64)^{1/3} = 4$$

$$\text{H.M.} = \frac{3}{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}} = \frac{3}{\frac{7}{8}} = \frac{3\times8}{7} = \frac{24}{7} = 3.43$$

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

Example 113. Comment on the following:

A.M.= 25, G.M.= 20 and H.M.= 21

Solution: The statement is wrong because H.M. cannot be greater than G.M.

RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE

The relationship between $\bar{X}$, M and Z depends on the shape of the frequency distribution which is discussed below:

(a) In a perfectly symmetrical distribution, mean, median and mode are all equal, *i.e.*,

$$\bar{X} = M = Z$$

(b) In a moderately asymmetrical distribution, mean, mode and median are not equal.

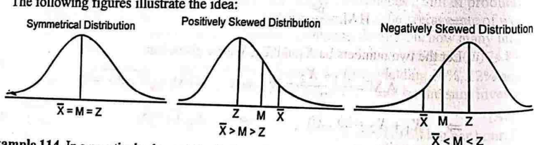
(*i*) when the distribution is positively skewed, *i.e.*, skewed to right, then

$$\bar{X} > M > Z$$

(*ii*) when the distribution is negatively skewed, *i.e.*, skewed to the left, then

$$\bar{X} < M < Z$$

The following figures illustrate the idea:



Example 114. In a negatively skewed distribution, mean, median and mode are calculated respectively as:

$$\bar{X} = 25, M = 28, Z = 22$$

Do you agree? Comment.

Solution: Given that in a nagatively skewed distribution,

$$\bar{X} = 25, M = 28, Z = 22$$

But we do not agree with the statement because in a negatively skewed distribution,

$$\bar{X} < M < Z$$

MISCELLANEOUS SOLVED EXAMPLES

Example 115. A firm declared bonus according to respective salary group as given below:

Salary group	Rate of Bonus	No.of Employees
60—75	60	3
75—90	70	4
90—105	80	5
105—120	90	5
120—135	100	7
135—150	110	6

Calculate the Arithmetic Mean of salary and Geometric Mean of the bonus to the employees.

Solution: **Calculation of Arithmetic Mean of Salary**

Salary group	Mid-value	No. of employees (f)	$A = 112.5$ $d = X - A$	d'	fd'
60—75	67.5	3	–45	–3	–9
75—90	82.5	4	–30	–2	–8
90—105	97.5	5	–15	–1	–5
105—120	112.5	5	0	0	0
120—135	127.5	7	+15	1	7
135—150	142.5	6	+30	2	12
		30			–3

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

$$= 112.5 + \frac{(-3)}{30} \times 15 = 112.5 - 1.5$$

$= 111$ rupees.

Rate of bonus (X)	Frequency (f)	$\log X$	$f \log X$
60	3	1.7782	5.3346
70	4	1.8451	7.3804
80	5	1.9031	9.5155
90	5	1.9542	9.7710
100	7	2.0000	14.0000
110	6	2.0414	12.2484
			58.2499

$$\log G = \frac{\Sigma f \log X}{\Sigma f} = \frac{58.2499}{30} = 1.9416$$

$$\text{G.M.} = \text{Antilog}\ [1.9416]$$
$$= 87.42$$

Example 116. The rate of certain commodity in the first week of January 1987 is 0.4 kg per rupee; it is 0.6 kg per week in the second week and 0.5 kg per rupee in the third week. Therefore is it correct to say that the average price is 0.5 kg per rupee? Verify.

Solution: The answer given is not correct as it is based on arithmetic mean whereas in case appropriate average is harmonic mean.

$$\text{H.M.} = \frac{N}{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}} = \frac{3}{\frac{1}{0.4}+\frac{1}{0.6}+\frac{1}{0.5}}$$

$$= \frac{3}{\frac{0.3+0.2+0.24}{0.12}} = \frac{3}{\frac{0.74}{0.12}} = \frac{3\times 0.12}{0.74}$$

$$= 0.486$$

Hence, the average price is 0.486 kg per rupee and not 0.5 kg per rupee.

Example 117. In a class of 50 students 10 have failed and their average of marks is 2.5. The total marks secured by the entire class were 281. Find the average marks those who have passed.

Solution: $N = 50$, failed = 10

Mean marks of those who failed = 2.5

Total marks of 10 students who failed = $2.5 \times 10 = 25$

Total marks secured by entire class = 281

Total marks obtained by those who passed = $281 - 25 = 256$

Average marks obtained by those who passed = $\frac{256}{40} = 6.4$

Example 118. B.Com. (Pass) III year has three Sections A, B and C with 50, 40, 60 students respectively. The mean marks for the three sections were determined as 85, 60 and 65 respectively. However, marks of a student of Section A were wrongly recorded as 50 instead of zero. Determine the mean marks of all the three sections put together.

Solution:

Section	A	B	C
Number (N)	50	40	60
Mean ($\overline{X}$)	85	60	65
ΣX	4250	2400	3900
Correct ΣX	(4250 – 50)	2400	3900

Correct mean of all the three sections = $\frac{4200+2400+3900}{150} = 70$

Example 119. An incomplete distribution is given below:

Variable:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	Total
Frequency:	10	20	—	40	—	25	15	=170

You are given that the median value is 35. Find out missing frequency.

Solution: Let the missing frequencies be denoted by x and y.

Variable	f	c.f.
0—10	10	10
10—20	20	30
20—30	x	30+x
30—40	40	70+x
40—50	y	70+x+y
50—60	25	95+x+y
60—70	15	110+x+y
	N=170	

Now, $110 + x + y = 170$

$\therefore$ $x + y = 60$

or $y = 60 - x$

Median = 35

$\therefore$ Median class is 30—40

$$\text{Median item} = \frac{N}{2} = \frac{170}{2} = 85 \text{ th item.}$$

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$35 = 30 + \frac{85-(30+x)}{40} \times 10$$

$$5 = \frac{85-30-x}{4} \times 10 = \frac{55-x}{4}$$

$\therefore$ $20 = 55 - x$

$\therefore$ $x = 35$

$y = 60 - x = 60 - 35 = 25$

Hence, the missing frequency of the class 20—30 is 35 and the missing frequency of the class 40—50 is 25.

Example 120. In 500 small-scale industrial units the return on investment ranged from 0 to 30 per cent, no unit sustaining any loss. 5 per cent of the units had returns ranging from 0 per cent upto (and including) 5 per cent and 15 per cent of the units earned returns exceeding 5 per cent but not exceeding 10 per cent. The median rate of return was

15 per cent and the upper quartile 20 per cent. The uppermost layer of the returns exceeding 25 per cent was earned by 50 units.

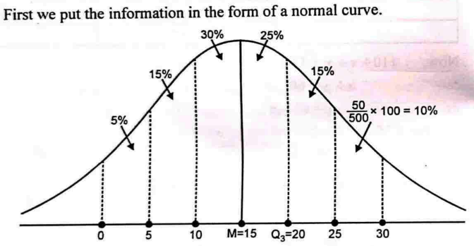
Present this information in the form of a frequency table with intervals of 5 per cent as follows:

Exceeding 0 per cent but not exceeding 5 per cent.
Exceeding 5 per cent but not exceeding 10 per cent.
Exceeding 10 per cent but not exceeding 15 per cent.
Exceeding 15 per cent but not exceeding 20 per cent.
Exceeding 20 per cent but not exceeding 25 per cent.
Exceeding 25 per cent but not exceeding 30 per cent.

Use $\frac{N}{4}, \frac{2N}{4}, \frac{3N}{4}$ as the ranks of the lower, middle and upper quartiles respectively.

Find rate of return round which there is maximum concentration of the units.

Solution: First we put the information in the form of a normal curve.



The information given above can be summarised as follows:

Rate of Return on Investment	Firms 5% of total	Number of firms (f)
Exceeding 0 but not exceeding 5	5	25
" 5 " " " 10	15	$75-f_0$
" 10 " " " 15	30	$150-f_1$
" 15 " " " 20	25	$125-f_2$
" 20 " " " 25	15 (Balance)	75
" 25 " " " 30	10	50
	100	500

The rate of return around which there is maximum concentration is the modal class. By inspection, mode lies in the classs 10—15.

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 10 + \frac{150-75}{300-75-125} \times 5$$

$$= 10 + \frac{75}{100} \times 5 = 10 + 3.75$$

$$= 13.75$$

Hence, the rate of return around which there is maximum concentration of the units is 13.75%.

Example 121. Find the missing frequency of the group 20—30 when the Median is 28.

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

Solution:

Calculation of Missing Frequency

X	f	c.f.
0—10	5	5
10—20	8	13
20—30	f_1	$13+f_1$
30—40	16	$29+f_1$
40—50	6	$35+f_1$
	$N=35+f$	

Since, Median = 28, it lies in the class 20—30

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$28 = 20 + \frac{\frac{35+f_1}{2} - 13}{f_1} \times 10$$

$$8 = \frac{\frac{35+f_1}{2} - 13}{f_1} \times 10$$

$\Rightarrow$ $$8f_1 = \left(\frac{35+f_1}{2} - \frac{13}{1}\right)10 = \left(\frac{35+f_1-26}{2}\right)10$$

$$16f_1 = 350 + 10f_1 - 260$$

$\Rightarrow$ $$6f_1 = 90$$

$\therefore$ $$f_1 = \frac{90}{6} = 15$$

Thus, the missing frequency is 15.

Example 122. The following table gives the frequency distribution of the marks of 400 candidate in an examination.

Marks	No. of Candidates	Marks	No. of candidates
0—10	5	50—60	65
10—20	20	60—70	50
20—30	40	70—80	35
30—40	70	80—90	20
40—50	85	90—100	10

(*i*) If the minimum marks required for passing are 35, what percentage of candidates pass the examination?

(*ii*) It is decided to allow 80% of the candidates to pass, what should be minimum marks for passing?

Solution:

Marks	f	$c.f.$
0—10	5	5
10—20	20	25
20—30	40	65
30—40	70	135
40—50	85	220
50—60	65	285
60—70	50	335
70—80	35	370
80—90	20	390
90—100	10	400
	$N = 400$	

(*i*) Let us assume that marks are given in whole numbers, the students getting marks upto 34 will fail.

∴ The percentage of candidates passing the examination

$$= \left[\left(\frac{5}{10} \times 70\right) + 85 + 65 + 50 + 35 + 20 + 10\right] \times \frac{100}{400}$$

$$= [(35) + 265] \times \frac{100}{400} = 300 \times \frac{100}{400} = 75\%$$

(*ii*) Since 80% candidates are passed and 20% have failed. The minimum pass marks are given by P_{20}.

$$P_{20} = \text{Size of } \frac{20(N)}{100}\text{th item} = \frac{20(400)}{100} = 80\text{th item}$$

It lies in class 30—40.

Applying the formula

$$P_{20} = 30 + \frac{80 - 65}{70} \times 10 = 30 + 2.1 = 32.1 \approx 32$$

Example 123. During a period of decline in stock market price, a stock sold at Rs. 50 per share on one day, Rs.40 on the next day and Rs. 25 on the third.

(*i*) If an investor bought 100, 120 and 180 shares on the respective three days, find the average price paid per share.

(*ii*) If an investor bought Rs. 1000 worth of shares on each of three days, find the average price paid per share.

Solution: (*i*) Since different quantities of shares are bought, we use weighted A.M.

Price (M) (X)	Quantity (W)	WX
50	100	5,000
40	120	4,800
25	180	4,500
	$\Sigma W = 400$	$\Sigma WX = 14{,}300$

$$\text{Average price} = \frac{\Sigma WX}{\Sigma W} = \frac{14{,}300}{400} = 35.75$$

(*ii*) Since equal sum of money is spent, we use simple H.M.

Price (X)	Money spent (Rs.)
50	1000
40	1000
25	1000

$$\text{Average price} = \frac{3}{\frac{1}{50} + \frac{1}{40} + \frac{1}{25}} = 35.29$$

Example 124. A man gets three annual increment in salary. At the end of the first year, he gets an increase of 4%, at the end of second an increase of 6% on his salary and at the end of the third year an increase of 9% on his salary. What is the average percentage increase?

Solution: Average rate of increase in the salary would be obtained by applying geometric mean. Suppose the salary at the beginning of the year is 100.

Year	Annual increase	Salary at the end of year (X)	log X
I	4%	104	2.0170
II	6%	106	2.0253
III	9%	109	2.0374
$N = 3$			$\Sigma \log X = 6.0797$

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

$$= \text{Antilog}\left(\frac{6.0797}{3}\right)$$

$$= \text{Antilog}\,[2.0265]$$

$$= 106.3$$

The average percentage increase $= 106.3 - 100 = 6.3\%$

Example 125. Find the missing information in the following table:

	A	B	C	Combined
Number	10	8	–	24
Mean	20	–	6	15
Geometric Mean	10	7	–	8.397

Solution: Finding missing information:

Number: For C missing information shall be

$$24 - (10 + 8) = 6$$

Mean: Let x be the mean of B

Then, $(20 \times 10) + (8 \times x) + (6 \times 6) = (15 \times 24)$

$$200 + 8x + 36 = 360$$

$\Rightarrow$ $8x = 360 - 200 - 36 = 124$

$\therefore$ $x = \dfrac{124}{8}$

$$= 15.5$$

Hence, mean of B = 15.5.

Geometric Mean : Let x be the geometric mean of C.

$$(10)^{10} \times (7)^{8} \times x^{6} = (8.397)^{24}$$

$$10 \log 10 + 8 \log 7 + 6 \log x = 24 \log 8.397$$

$$10 + (8 \times 0.8451) + 6 \log x = 24(0.9241)$$

$$10 + 6.7608 + 6 \log x = 22.1784$$

$\Rightarrow$ $6 \log x = 22.1784 - 10 - 6.7608 = 5.4176$

$$\log x = 0.9029$$

$$x = \text{Antilog}\; 0.9029$$

$$= 7.997$$

Hence, geometric mean of C is 7.997.

IMPORTANT FORMULAE

1. Arithmetic Mean

For Individual Series:

(*i*) $\overline{X} = \dfrac{\Sigma x}{N}$ — Direct Method

(*ii*) $\overline{X} = A + \dfrac{\Sigma d}{N}$ — Short-cut Method

(*iii*) $\overline{X} = A + \dfrac{\Sigma d'}{N} \times i$ — Step Deviation Method

For Discrete and Continuous Series:

(*i*) $\overline{X} = \dfrac{\Sigma fx}{N}$ *or* $\dfrac{\Sigma fm}{N}$ — Direct Method

(*ii*) $\overline{X} = A + \dfrac{\Sigma fd}{N}$ — Short-cut Method

(*iii*) $\overline{X} = A + \dfrac{\Sigma fd'}{N} \times i$ — Step Deviation Method

2. Weighted Arithmetic Mean

$$\overline{X}_w = \frac{\Sigma WX}{\Sigma W}$$

3. Combined Mean

$$\overline{X}_{123} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3}{N_1 + N_2 + N_3}$$

4. Median

For Individual and Discrete Series:

$$M = \text{Size of}\left(\frac{N+1}{2}\right)\text{th item}$$

For Continuous Series:

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

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

5. Mode

For Continuous Series:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

6. Emprical Mode

Mode (Z) = 3 Median − 2 Arithmetic Mean

or $Z = 3M - 2\bar{X}$

7. Geometric Mean

For Individual Series

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right]$$

For Discrete and Continuous Series

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma f \log X}{N}\right]$$

Compound Interest Formula

$$P_n = P_0 (1 + r)^n$$

or $$r = \text{Antilog}\left[\frac{\log P_n - \log P_0}{n}\right] - 1$$

8. Harmonic Mean

For Individual Series

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{1}{X}\right)}$$

For Discrete and Continuous Series

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{f}{X}\right)}$$

Weighted Harmonic Mean

$$\text{Weighted H.M.} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$$

QUESTIONS

1. What is meant by central tendency? State important measures of central tendency.
2. What is a statistical average? What are the properties of an ideal average?
3. Explain the relative merits and demerits of arithmetic mean, median and mode as measures of central tendency.

   Or

   Define arithmetic mean, median and mode and discuss their relative merits and demerits.
4. Discuss in brief merits and demerits of various measures of central tendency.
5. Give different measures of central tendency with their formulae. Also state the situations where these measures can be used.
6. Give relationship between A.M., G.M. and H.M.
7. Give the empirical relationship between $\bar{X}$, M and Z for a perfectly symmetrical distribution and a moderately skewed distribution.
8. What are the essentials of a good average?

   Or

   What are the characteristics of a good measure of central tendency?
9. Explain the mathematical properties of Arithmetic Mean. What is the relationship among Mean, Median and Mode?
10. What are the desirable properties of an average? Which of the average you know possess most of them?

6 Measures of Dispersion

INTRODUCTION

The various measures of central tendency or averages discussed in the previous chapter give us only one single figure that represents the entire set of data. But the average alone cannot describe the set of observations fully. It does not reveal the degree of spread out or extent of variability of individual observations in a series. There can be a number of series whose average may be the same but still there can be wide disparities in the formation of the series. In such a case, it becomes necessary to study the variability or dispersion of the observations. Measures of dispersion help us to study variability of the items, *i.e.*, the extent to which the items vary from one another and from the central value.

MEANING OF DISPERSION

The term dispersion is generally used in two senses: (1) Firstly, dispersion refers to the variations of the items among themselves. If the value of all the items of a series is the same, there will be no variation among the various items and the dispersion will be zero. On the other hand, the greater the variation among different items of a series, the more will be the dispersion. (2) Secondly, dispersion refers to the variation of the items around an average. If the difference between the value of items and the average is large, the dispersion will be high and on the other hand if the difference between the value of the items and average is small, the dispersion will be low. Thus, dispersion is defined as scatterness or spreadness of the individual items in a given series.

DEFINITION OF DISPERSION

Some important definitions of dispersion are given below:

1. *Dispersion is a measure of the variations of the items.* —**Bowley**
2. *Dispersion is a measure of the extent to which the individual items vary.* —**Connor**
3. *The degree to which numerical data tend to spread about average is called variation or dispersion of data.* —**Spiegel**

The above definitions make it clear that dispersion refers to the extent to whch the items vary from one another and from the central value.

OBJECTIVES OF MEASURING DISPERSION

The measures of dispersion are helpful in statistical investigation. Some of the main objectives of dispersion are as under:

(1) To determine the reliability of an average: The measures of dispersion help in determining the reliability of an average. It points out as to how far an average is representative of a statistical

series. If the dispersion or variation is small, the average will closely represent the individual values and it is highly representative. On the other hand, if the dispersion or variation is large, the average will be quite unreliable.

(2) To compare the variability of two or more series: The measures of dispersion helps in comparing the variability of two or more series. It is also useful to determine the uniformity or consistency of two or more series. A high degree of variation would mean less consistency or less uniformity as compared to the data having less variation.

(3) For facilitating the use of other statistical measures: Measures of dispersion serves the basis of many other statistical measures such as correlation, regression, testing of hypothesis, etc. These measures are based on measures of variation of one kind or another.

(4) Basis of Statistical Quality Control: The measures of dispersion is the basis of statistical quality control. The extent of the dispersion gives indication to the management as to whether the variation in the quality of the product is due to random factors or there is some defect in the manufacturing process. On the basis of this analysis, the management may take suitable measures to control the cause of variation in the quality of the product.

PROPERTIES OF A GOOD MEASURE OF DISPERSION

A good measure of dispersion should possess the following properties:

(1) It should be easy to understand.

(2) It should be simple to calculate.

(3) It should be uniquely defined.

(4) It should be based on all observations.

(5) It should not be unduly affected by the extreme items.

(6) It should be capable of further algebraic treatment.

ABSOLUTE AND RELATIVE MEASURES OF DISPERSION

Measures of dispersion may be either **absolute** or **relative**.

Absolute Measure of Dispersion: Absolute measure of dispersion is expressed in the same unit in which data of the series are expressed. They are expressed in same statistical unit, *e.g.*, rupees, kilogram, tons, years, centimeters, etc.

Relative Measure of Dispersion: Relative measure of dispersion refers to the variability stated in the form of ratio or percentage. Thus, relative measure of dispersion is independent of unit of measurement. It is also called **coefficient of dispersion**. These measures are used to compare two series expressed in different units.

METHODS OF MEASURING DISPERSION

The following are the main methods of measuring dispersion:

(1) Range

(2) Interquartile Range and Quartile Deviation

(3) Mean Deviation

(4) **Standard Deviation**
(5) **Coefficient of Variation**
(6) **Lorenz Curve.**

(1) RANGE

It is the simplest measures of dispersion. It is defined as the difference between the largest and smallest values in the series. Its formula is:

$$R = L - S$$

Where, R = Range, L = Largest value in the series, S = Smallest value in the series.

The relative measure of range, also called **coefficient of range**, is defined as:

$$\textbf{Coefficient of Range} = \frac{L-S}{L+S}$$

The following examples illustrate the calculation of range:

Calculation of Range

Individual Series

Example 1. Five students obtained the following marks in statistics:

20, 35, 25, 30, 15

Find the range and coefficient of Range.

Solution: Here, L = 35, and S = 15

Range = L − S

∴ Range = 35 − 15 = 20

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

$$= \frac{35-15}{35+15} = \frac{20}{50} = \frac{2}{5} = +0.40$$

Discrete Series

Example 2. Find the range and coefficient of range from the following data:

Marks:	10	20	30	40	50	60	70
No. of students:	15	18	25	30	16	10	9

Solution: Here, L = 70, and S = 10

Range = L − S

∴ Range = 70 − 10 = 60

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{70-10}{70+10} = \frac{60}{80} = \frac{3}{4} = +0.75$$

Continuous Series

Example 3. Find out range and the coefficient of range of the following series:

Size:	5—10	10—15	15—20	20—25	25—30
Frequency:	4	9	15	30	40

Solution: Range = L − S

Here, L = Upper limit of the largest class = 30, S = Lower limit of the smallest class = 5

∴ Range = 30 − 5 = 25

$$\text{Coefficient of Range} = \frac{30-5}{30+5} = \frac{25}{35} = \frac{5}{7}$$

Note: Since the maximum and minimum of the observations are not identifiable for a continuous series, the range is defined as the difference between the upper limit of the largest class and the lower limit of the smallest class.

Example 4. Find out range and coefficient of range of the following series:

Marks:	20—29	30—39	40—49	50—59	60—69
No. of students:	8	12	20	7	3

Solution: This is an inclusive series. For the calculation of range, the series must be converted into exclusive series as:

Marks	No. of students
19.5—29.5	8
29.5—39.5	12
39.5—49.5	20
49.5—59.5	7
59.5—69.5	3

Here, L = 69.5, S = 19.5

∴ Range = 69.5 − 19.5 = 50

$$\text{Coefficient of Range} = \frac{69.5-19.5}{69.5+19.5} = \frac{50}{89} = 0.562$$

Merits and Demerits of Range

Merits:

(i) It is simple to understand.
(ii) It is easy to calculate.
(iii) It is widely used in statistical quality control. Range charts are useful in controlling the quality of the product.

Demerits:

(i) It cannot be calculated in case of open ended distribution.
(ii) It is not based on all observations of the series.
(iii) It is affected by sampling fluctuations.
(iv) It is affected by extreme values in the series.

EXERCISE 6.1

1. 5 students obtained following marks in statistics:
   20, 35, 25, 30, 15
   Find out range and coefficient of range. [Ans. R=20, Coeff. of R = 0.4]
2. Calculate the range and its coefficient from the following data:

Marks:	10	20	30	40	50	60	70
No. of students:	15	18	25	30	16	10	9

[Ans. R=60, Coeff. of Range = 0.75]

3. Calculate the range and its coefficient from the following data:

Marks:	10-20	20-30	30-40	40-50	50-60
No. of students:	8	10	12	8	4

[Ans. R=50, Coeff. of R = 0.714]

4. Find the value of the smallest item of data if the coefficient of range is 0.6 and largest value is 60. [Ans. S=15]

(2) INTERQUARTILE RANGE AND QUARTILE DEVIATION

Interquartile range and quartile deviation are another measures of dispersion. The difference between the upper quartile (Q_3) and the lower quartile (Q_1) is called interquartile range. Symbolically,

$$\textbf{Interquartile Range} = Q_3 - Q_1$$

The interquartile ranges covers dispersion of middle 50% of the items of the series.

Quartile deviation, also called **Semi-interquartile Range** is half of the difference between the upper and lower quartiles, *i.e.*, half of the interquartile range. Its formula is as:

$$\textbf{Quartile Deviation (Q.D.)} = \frac{Q_3 - Q_1}{2}$$

The relative measure of quartile deviation also called the **coefficient of quartile deviation** is defined as:

$$\textbf{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

▶ Steps for Calculation

(*i*) Find out the lower quartile (Q_1).
(*ii*) Find out the upper quartile (Q_3).
(*iii*) Put the values of Q_1 and Q_3 in the formula of quartile deviation and coefficient of quartile deviation.

The following examples illustrate the calculation of quartile deviation:

Individual Series

Example 5. Find interquartile range, quartile deviation and coefficient of quartile deviation from the following data:

28, 18, 20, 24, 27, 30, 15

Solution: First arrange the data in ascending order:

15, 18, 20, 24, 27, 28, 30

$$Q_1 = \text{Size of}\left(\frac{N+1}{4}\right)\text{th item} = \text{Size of}\left(\frac{7+1}{4}\right)\text{th item}$$
$$= \text{Size of 2nd item} = 18 \text{ marks}$$
$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)\text{th item} = \text{Size of } 3\left(\frac{7+1}{4}\right)\text{th item}$$
$$= \text{Size of 6th item} = 28 \text{ marks}$$
$$\text{Interquartile Range} = Q_3 - Q_1 = 28 - 18 = 10$$
$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{28-18}{2} = \frac{10}{2} = 5$$
$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{28-18}{28+18} = \frac{10}{46} = 0.217$$

Discrete Series

Example 6. Calculate interquartile range, quartile deviation and the coefficient of quartile deviation from the following data:

Wages (Rs.):	10	20	30	40	50	60
No. of workers:	2	8	20	35	42	20

Solution:

Calculation of Q.D.

(X)	f	*c.f.*
10	2	2
20	8	10
30	20	30
40	35	65
50	42	107
60	20	127
	$N = 127$	

$$Q_1 = \text{Size of } \frac{N+1}{4} \text{ th item} = \text{Size of}\left(\frac{127+1}{4}\right)\text{th item}$$
$$= \text{Size of 32nd item} = 40$$
$$\therefore \quad Q_1 = 40$$

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)\text{th item} = \frac{3\times128}{4} = \text{Size of 96th item} = 50$$

$$\text{Interquartile Range } = Q_3 - Q_1 = 50 - 40 = 10$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{50-40}{2} = \frac{10}{2} = 5$$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{50-40}{50+40} = \frac{10}{90} = 0.11$$

Continuous Series

Example 7. Calculate interquartile range, quartile deviation and coefficient of quartile deviation from the following data:

Age (years):	0—20	20—40	40—60	60—80	80—100
Persons:	4	10	15	20	11

Solution: **Calculation of Q.D. and Coefficient of Q.D.**

Age (yrs.)	f	$c.f.$
0—20	4	4
20—40	10	14
40—60	15	29
60—80	20	49
80—100	11	60
	$N = 60$	

Calculation of Q_1

$\frac{N}{4} = \frac{60}{4} = 15$th item $\quad\therefore Q_1$ lies in the class 40—60

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 40 + \frac{15-14}{15} \times 20$$

$$= 40 + \frac{1}{15} \times 20 = 41.33$$

Calculation of Q_3

$\frac{3N}{4} = \frac{3}{4} \times 60 = 45$ th item. $\therefore Q_3$ lies in the class 60—80

$$Q_1 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i$$

$$= 60 + \frac{45-29}{20} \times 20$$

$$= 60 + 16 = 76$$

$$\text{Interquartile Range} = Q_3 - Q_1 = 76 - 41.33 = 34.67$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{76 - 41.33}{2} = 17.33$$

$$\text{Coefficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{76 - 41.33}{76 + 41.33} = \frac{34.67}{117.33} = 0.29$$

Example 8. Find the range which contains the middle 50% of the items and coefficient of quartile deviation from the following data:

X:	11—20	21—30	31—40	41—50	51—60
f:	4	8	20	12	6

Solution: Since, we are given inclusive series, we first convert it into exclusive one:

Calculation of Q.D. and Coefficient of Q.D.

X	f	$c.f.$
10.5—20.5	4	4
20.5—30.5	8	12
30.5—40.5	20	32
40.5—50.5	12	44
50.5—60.5	6	50
	$N = 50$	

Calculation of Q_1

$\frac{N}{4} = \frac{50}{4} = 12.5$th item. $\quad\therefore Q_1$ lies in the class 30.5 – 40.5

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 30.5 + \frac{12.5 - 12}{20} \times 10$$

$$= 30.5 + \frac{0.5}{20} \times 10 = 30.5 + 0.25 = 30.75$$

Calculation of Q_3

$\frac{3N}{4} = \frac{3 \times 50}{4} = \frac{150}{4} = 37.5$ th item. $\therefore Q_3$ lies in the class 40.5 – 50.5

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 40.5 + \frac{37.5 - 32}{12} \times 10$$

$$= 40.5 + 4.58 = 45.08$$

To find the middle 50% of the items, compute $Q_3 - Q_1$, *i.e.*, Interquartile range.

$$\text{Interquartile Range} = Q_3 - Q_1 = 45.08 - 30.75 = 14.33$$

$$\text{Coefficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{45.08 - 30.75}{45.08 + 30.75} = \frac{14.33}{75.83} = 0.188 = 0.19$$

Merits and Demerits of Quartile Deviation

Merits:

(i) It is easy to compute and simple to understand.

(ii) It is less affected by extreme values.

(iii) It can be computed in open ended classes.

(iv) It is superior and more reliable than the range.

(v) It is useful when dispersion of middle 50% items is to be calculated.

Demerits:

(i) It gives 50% of the items, *i.e.*, the first 25% and the last 25%.

(ii) It is not capable for further algebraic treatment.

(iii) It is affected by sampling fluctuations.

(iv) It is not a good measure of dispersion particularly for series in which variation is considerable.

EXERCISE 6.2

1. From the following data, compute Q.D. and Coefficient of Q.D.:

X:	4	8	10	7	11	15	18	14	12	16

[Ans. Q.D.=3.75, Coeff. of Q.D.= 0.326]

2. From the following data, calculate Q.D. and its coefficient:

Height:	58	59	60	61	62	63	64	65	66
No. of students:	15	20	32	35	33	22	20	10	8

[Ans. Q.D.=1.5, Coeff. of Q.D. = 0.024]

3. Calculate Quartile deviation and its coefficient from the following data:

Size:	less than 32	32—34	34—36	36—38	38—40	40—42	over 42
Frequency:	12	18	16	14	12	8	6

[Ans. Q.D. = 2.85, Coeff. of Q.D. = 0.079]

4. Calculate the range of marks obtained by middle 50% of the students from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	5	8	12	8	7

[Hint: Calculate $Q_3 - Q_1$]

[Ans. $IQR = 20$]

5. Calculate Q.D. and its coefficient from the following data:

X:	31—40	41—50	51—60	61—70	71—80
f:	40	60	20	40	25

[Ans. Q.D. = 7.087, Coeff. of Q.D. = 0.0556]

6. Calculate the range of marks obtained by middle 80% of the students:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60170
No. of students:	8	16	22	30	24	12	6

[**Hint:** Calculate $P_{90} - P_{10}$]

[Ans. Percentile Range = 4279]

7. If the first quartile is 142 and the semi-interquartile range is 18, find the median (assuming the distribution to be symmetrical).

[Ans. Q_2 (M) = 160]

(3) MEAN DEVIATION OR AVERAGE DEVIATION

Mean deviation is another measure of dispersion. It is also known as **average deviation**. Mean deviation is defined as the arithmetic average of the deviation of the various items of a series computed from some measures of central tendency say mean or median. In taking deviation of the various items, algebraic signs '+' and '–' are not taken into consideration. Although mean deviation can be computed either from the mean or median, but theoretically median is preferred because **the sum of the deviations of the items taken from median is minimum when signs are ignored.** The formulae of calculating mean deviation are:

$$\textbf{M.D. from Median} = \frac{\Sigma|X - M|}{N} \quad or \quad \frac{\Sigma|d_M|}{N}$$

$$\textbf{M.D. from Mean} = \frac{\Sigma|X - \bar{X}|}{N} \quad or \quad \frac{\Sigma|d_{\bar{x}}|}{N}$$

The relative measure of mean deviation, also called the **coefficient of mean deviation** is obtained by dividing mean deviation by the particular average used in computing mean deviation. Thus,

$$\textbf{Coefficient of } M.D._M = \frac{M.D._M}{Median}$$

$$\textbf{Coefficient of } M.D._{\bar{x}} = \frac{M.D._{\bar{x}}}{Mean}$$

Calculation of Mean Deviation

Individual Series

In case of individual series, the following formula is used for calculating mean deviation:

$$M.D._{\bar{X}} = \frac{\Sigma|d_{\bar{X}}|}{N}, \text{ from mean}$$

$$M.D._M = \frac{\Sigma|d_M|}{M}, \text{ from median.}$$

Steps for Calculation

(i) Calculate either the mean or median.

(ii) Find the deviations of the items from $\bar{X}$ or M ignoring signs and denote these deviations by $|d_{\bar{X}}|$ or $|d_m|$.

(iii) Find the total of the deviations, *i.e.*, $\Sigma|d_{\bar{X}}|$ or $\Sigma|d_M|$.

(iv) Divide the total obtained in step (iii) by the number of item (N). This gives us the mean deviation.

Coefficient of Mean Deviation

The formulae are:

$$\text{Coefficient of } M.D._{\bar{X}} = \frac{M.D._{\bar{X}}}{\bar{X}}$$

$$\text{Coefficient of } M.D._M = \frac{M.D._M}{M}$$

Example 9. Calculate the mean deviation from mean as well as from median and coefficient of mean deviation from the following data:

Marks: 20, 22, 25, 38, 40, 50, 65, 70, 75

Solution:

Calculation of Mean Deviation

Marks (X)	Deviation from Mean 45 $\lvert d_{\bar{X}} \rvert$	Deviations from Median 40 $\lvert d_M \rvert$
20	25	20
22	23	18
25	20	15
38	7	2
40	5	0
50	5	10
65	20	25
70	25	30
75	30	35
$N = 9$, $\Sigma X = 405$	$\Sigma \lvert d_{\bar{X}} \rvert = 160$	$\Sigma \lvert d_M \rvert = 155$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{405}{9} = 45$$

$$\text{M.D. from Mean} = \frac{\Sigma \lvert d_{\bar{X}} \rvert}{N} = \frac{160}{9} = 17.78$$

$$\text{Coeff. of } M.D._{\bar{X}} = \frac{M.D._{\bar{X}}}{\bar{X}} = \frac{17.78}{45} = 0.39$$

$$\text{Median} = \text{Size of}\left(\frac{N+1}{2}\right)\text{th item} = \text{Size of}\left(\frac{9+1}{2}\right)\text{th item}$$

$$= \text{Size of 5th item} = 40$$

$$\text{M.D. from Median} = \frac{\Sigma \lvert d_M \rvert}{N} = \frac{155}{9} = 17.22$$

$$\text{Coeff. of } M.D._M = \frac{M.D._M}{M} = \frac{17.22}{40} = 0.43$$

Discrete Series

In case of discrete series, the following formula is used for calculating mean deviation:

$$MD._{\bar{X}} = \frac{\Sigma f \lvert d_{\bar{X}} \rvert}{N}, \text{ from mean}$$

$$M.D._M = \frac{\Sigma f \lvert d_M \rvert}{M}, \text{ from median.}$$

Steps for Calculation

(*i*) Calculate the mean or median from the given series.

(*ii*) Find the deviations of the items from $\bar{X}$ ignoring '+' or '–' signs and denote these deviations by $\lvert d_{\bar{X}} \rvert$ or $\lvert d_M \rvert$.

(*iii*) Multiply these deviations by their respective frequencies and obtain the total $\Sigma f \lvert d \rvert$.

(*iv*) Finally, divide this total by the number of items. This will give the mean deviation.

(*v*) To calculate the coefficient of mean deviation, we divide the mean deviation by the particular average used in computing mean deviation.

Example 10. Calculate the mean deviation from median and mean and their coefficients from the following table:

X:	20	30	40	50	60	70
f:	8	12	20	10	6	4

Solution: (*i*) **Calculation of Mean Deviation from Median**

X	f	c.f.	M = 40 $\lvert d_M \rvert = \lvert X - M \rvert$	$f \lvert d_M \rvert$
20	8	8	20	160
30	12	20	10	120
40	20	40	0	0
50	10	50	10	100
60	6	56	20	120
70	4	60	30	120
	$N = 60$			$\Sigma f \lvert d_M \rvert = 620$

$$M = \text{Size of}\left(\frac{N+1}{2}\right)\text{th item} = \text{Size of}\left(\frac{60+1}{2}\right)\text{th item}$$

$$= \text{Size of 30.5th item} = 40$$

$$\therefore \quad M = 40$$

$$\text{M.D. from Median } (M.D._M) = \frac{\Sigma f \lvert d_M \rvert}{N} = \frac{620}{60} = 10.33$$

$$\text{Coefficient of M.D. from Median} = \frac{M.D._M}{M} = \frac{10.33}{40} = 0.258$$

(ii)

Calculation of Mean Deviation from Mean

X	f	fX	$\bar{X}=41$ $\lvert d_{\bar{X}}\rvert = \lvert X-\bar{X}\rvert$	$f\lvert d_{\bar{X}}\rvert$
20	8	160	21	168
30	12	360	11	132
40	20	800	1	20
50	10	500	9	90
60	6	360	19	114
70	4	280	29	116
	$N=60$	$\Sigma fX=2460$		$\Sigma f\lvert d_{\bar{X}}\rvert=640$

$$\bar{X}=\frac{\Sigma fX}{N}=\frac{2460}{60}=41$$

$$\text{M.D. from Mean } (M.D._{\bar{X}})=\frac{\Sigma f|d_{\bar{X}}|}{N}=\frac{640}{60}=10.67$$

$$\text{Coefficient of M.D. from Mean}=\frac{M.D_{\bar{X}}}{\bar{X}}=\frac{10.67}{41}=0.26$$

Continuous Series

For calculating the mean deviation in continuous series, the procedure remains the same as discussed above. The only difference is that here we have to obtain the mid-values of the various classes and then take deviations from these values as before. The formulae are:

$$M.D._{\bar{X}}=\frac{\Sigma f|d_{\bar{X}}|}{N}, \text{ where } d_{\bar{X}}=m-\bar{X}$$

$$M.D_M=\frac{\Sigma f|d_M|}{N}, \text{ where } d_M=m-M$$

$$\text{Coefficient of } M.D_{\bar{X}}=\frac{M.D_{\bar{X}}}{\bar{X}}$$

$$\text{Coefficient of } M.D_M=\frac{M.D_M}{M}$$

Example 11. Calculate the mean deviation from mean and its coefficient from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	5	8	15	16	6

Solution:

Calculation of Mean Deviation from Mean

Marks	f	M.V. (m)	fm	$\bar{X}=27$ $\lvert d_{\bar{X}}\rvert=\lvert m-27\rvert$	$f\lvert d_{\bar{X}}\rvert$
0—10	5	5	25	22	110
10—20	8	15	120	12	96
20—30	15	25	375	2	30
30—40	16	35	560	8	128
40—50	6	45	270	18	108
	$N=50$		$\Sigma fm=1350$		$\Sigma f\lvert d_{\bar{X}}\rvert=472$

$$\bar{X}=\frac{\Sigma fm}{N}=\frac{1350}{50}=27$$

$$\text{M.D. from Mean}=\frac{\Sigma f|d_{\bar{X}}|}{N}=\frac{472}{50}=9.44$$

$$\text{Coefficient of } M.D._{\bar{X}}=\frac{M.D._{\bar{X}}}{\bar{X}}=\frac{9.44}{27}=0.349$$

Example 12. Calculate mean deviation from median and its coefficient from the following data:

Size:	100—120	120—140	140—160	160—180	180—200
Frequency:	4	6	10	8	5

Solution:

Calculation of Mean Deviation from Median

Size	f	$c.f.$	$M.V.$ (m)	M = 153 $\lvert d_M\rvert=\lvert m-M\rvert$	$f\lvert d_M\rvert$
100—120	4	4	110	43	172
120—140	6	10	130	23	138
140—160	10	20	150	3	30
160—180	8	28	170	17	136
180—200	5	33	190	37	185
	$N=33$				$\Sigma f\lvert d_M\rvert=661$

N/2 = 33/2 = 16.5th item ∴ Median lies in 140 – 160.

Applying the formula

$$M=l_1+\frac{\frac{N}{2}-c.f.}{f}\times i$$

$$=140+\left(\frac{16.5-10}{10}\right)\times 20=140+13=153$$

$$\text{Mean Deviation from Median}=\frac{\Sigma f|d_M|}{N}=\frac{661}{33}=20.03$$

$$\text{Coefficient of Mean Deviation from Median}=\frac{M.D._M}{M}=\frac{20.03}{153}=0.1309$$

Thus, $M.D._M=20.03$, and Coefficient of $M.D._M=0.1309$.

Short-cut Method for Mean Deviation

If value of the average comes out to be in fractions, the calculation of M.D. by $\frac{\Sigma f|d|}{N}$ would become quite tedious. In such a case, the following formula is used:

$$M.D.=\frac{(\Sigma fX)_A-(\Sigma fX)_B-[(\Sigma f)_A-(\Sigma f)_B]\ \bar{X} \text{ or } M}{N}$$

Where, $\bar{X}$ or M is the average about which M.D. is to be calculated. In this formula, suffixes A and B denote the sums corresponding to the values of $X>\bar{X}$ or M and $X<\bar{X}$ or M respectively.

This formula can also be used for an individual series, by taking 'f' equal to 1 for each X, in the series. In this case, the formula reduces to

$$M.D. = \frac{(\Sigma X)_A - (\Sigma X)_B - [(N)_A - (N)_B]\ \overline{X}\ or\ M}{N}$$

Where, $(N)_A$ and $(N)_B$ are the number of items whose values are greater than $\overline{X}$ or M and less than $\overline{X}$ or M respectively.

Note: If short-cut method is to be used to find M.D. $(\overline{X})$, then it is advisable to use *direct method* to find $\overline{X}$, because we would be needing $(\Sigma fX)_A$ and $(\Sigma fX)_B$ in the calculation of M.D. $(\overline{X})$.

Now, we shall take some examples to illustrate this method:

Example 13. Using short-cut method, calculate the mean deviations from mean and median from the following data:

7, 9, 13, 13, 15, 17, 19, 21, 23

Solution:

X	Taking $\overline{X}$	Taking M
7	$\Sigma X_B = 57$	$\Sigma X_B = 42$
9	$N_B = 5$	$N_B = 4$
13		
13		
15		M = 15
	$\overline{X} = 15.22$	
17	$\Sigma X_A = 80$	$\Sigma X_A = 80$
19	$N_A = 4$	$N_A = 4$
21		
23		
$\Sigma X = 137$ $N = 9$		

$\therefore$

$$\overline{X} = \frac{137}{9} = 15.22$$

$$M = \text{Size of} \left(\frac{N+1}{2}\right)\text{th} = \frac{9+1}{2} = \text{5th item} = 15$$

From Mean:

$$M.D_{\overline{X}} = \frac{\Sigma X_A - \Sigma X_B - (N_A - N_B)(\overline{X})}{N}$$

$$= \frac{80 - 57 - (4-5)(15.22)}{9}$$

$$= \frac{23 + 15.22}{9} = \frac{38.22}{9} = 4.25$$

From Median:

$$M.D_M = \frac{\Sigma X_A - \Sigma X_B - (N_A - N_B)(M)}{N}$$

$$= \frac{80 - 42 - (4-4)(15)}{9}$$

$$= \frac{38}{9} = 4.22$$

Example 14. Calculate the mean deviation and its coefficient from mean and median from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No.of students:	6	28	51	11	4

Use short-cut method.

Solution: (*i*) **Calculation of M.D. from Mean**

Marks	f		M.V. (X)	fX	
0—10	6	$(\Sigma f)_B = 34$	5	30	$(\Sigma fX)_B = 450$
10—20	28		15	420	
					$\overline{X} = 22.9$
20—30	51	$(\Sigma f)_A = 66$	25	1275	$(\Sigma fX)_A = 1840$
30—40	11		35	385	
40—50	4		45	180	
	$N = 100$			$\Sigma fX = 2290$	

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{2290}{100} = 22.9$$

$$M.D_{\overline{X}} = \frac{(\Sigma fX)_A - (\Sigma fX)_B - [(\Sigma f)_A - (\Sigma f)_B]\overline{X}}{N}$$

$$= \frac{1840 - 450 - [66 - 34][22.9]}{100}$$

$$= \frac{1390 - (32 \times 22.9)}{100}$$

$$= \frac{1390 - 732.8}{100}$$

$$= \frac{657.2}{100} = 6.572$$

$$\text{Coefficient of } M.D_{\overline{X}} = \frac{M.D_{\overline{X}}}{\overline{X}} = \frac{6.572}{22.9} = 0.287$$

(ii)

Calculation of M.D. from Median

Marks	f		c.f.	M.V. (X)	fX	
0—10	6	$(\Sigma f)_B = 34$	6	5	30	$(\Sigma fX)_B = 450$
10—20	28		34	15	420	
						M = 34.14
20—30	51	$(\Sigma f)_A = 66$	85	25	1275	$(\Sigma fX)_A = 1840$
30—40	11		96	35	385	
40—50	4		100	45	180	
	$N = 100$					

$$\frac{N}{2} = \frac{100}{2} = 50\text{th item}$$

Median lies in the class 20—30

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{50 - 34}{51} \times 10$$

$$= 20 + \frac{160}{51} = 20 + 3.14 = 23.14$$

$$M.D._M = \frac{(\Sigma fX)_A - (\Sigma fX)_B - [(\Sigma f)_A - (\Sigma f)_B](M)}{N}$$

$$= \frac{1840 - 450 - (66 - 34)(23.14)}{100}$$

$$= \frac{1390 - 32 \times 23.14}{100}$$

$$= \frac{1390 - 740.48}{100} = \frac{649.52}{100} = 6.4952$$

$$\text{Coefficient of } M.D._M = \frac{M.D._M}{M} = \frac{6.4952}{23.14} = 0.281$$

Merits and Demerits of Mean Deviation

Merits:

(i) It is simple to understand and easy to compute.
(ii) It is based on all the observations.
(iii) It is less affected by the extreme items.
(iv) It is very useful in various fields such as economics, commerce and social fields.
(v) Comparision about formation of different series can be easily made as deviations are taken from a central value.

Demerits:

(i) Ignoring '±' signs are not correct from mathematical point of view.
(ii) It is not an accurate method when it is calculated from mode.
(iii) It is difficult to compute when the value of mean or median comes in fractions.
(iv) It is not capable of further algebraic treatment.
(v) It is not used in statistical conclusion.

EXERCISE 6.3

1. The monthly salary of five families are given below:
852, 635, 792, 836, 750
Find the mean deviation from the mean. [Ans. $M.D._{\bar{X}} = 64.4$]

2. (i) Find mean deviation and its coefficient from the mean for the following data:

X:	5	10	15	20	25	30	35	40
f:	8	16	18	22	14	9	6	7

[Ans. $M.D._{\bar{X}} = 7.57$, Coefficient of $M.D._{\bar{X}} = 0.37$]

(ii) Calculate the average deviation from median and its coefficient for the following data:

X:	6	12	18	24	30	36	42
f:	4	7	9	18	15	10	5

[Ans. $M = 24$, $M.D._M = 7.94$, Coeff. of $M.D._M = 0.330$]

3. Calculate mean deviation from mean and median for the following data:

Marks:	140—150	150—160	160—170	170—180	180—190	190—200
No. of students:	4	6	10	18	9	3

[Ans. $M.D._{\bar{X}} = 10.56, M.D._M = 10.24$]

4. Calculate mean deviation from the median and its coefficient for the following data:

X:	10—19	20—29	30—39	40—49	50—59
f:	3	4	6	5	2

[Ans. M = 34.5, $M.D._M = 9.5$, Coefficient of $M.D_M = 0.275$]

5. Calculate the average deviation from mean and its coefficient from the following data:

Profit (Rs.):	0—10	10—20	20—30	30—40	40—50
No. of Shops:	5	10	15	20	25

[Ans. $\bar{X} = 31.66, M.D._{\bar{X}} = 10.67$, Coeff. of $M.D._{\bar{X}} = 0.34$]

6. Calculate the mean deviation from the median and its coefficient for the following data:

Class:	0—5	5—10	10—15	15—20	20—25	25—30
f:	7	10	16	32	24	18

Use short-cut method. [Ans. $M = 18.2, M.D. = 4.617$, Coeff. of $M.D. = 0.0308$]

7. Find out the mean deviation from the mean and its coefficient from the following data:

Class:	0—3	3—6	6—9	9—12	12—15	15—18	18—21
f:	2	7	10	12	9	6	4

Use short-cut method. [Ans. $\overline{X}=10.68, M.D.=3.8184$, Coeff. of $M.D.=0.3579$]

(4) STANDARD DEVIATION

Standard deviation is the most important and widely used measure of dispersion. It was first used by Karl Pearson in 1893. Standard deviation is also called as **root mean square deviation.** Standard deviation is defined as the square root of the arithmetic mean of the squares of the deviation of the values taken from the mean. Standard deviation is denoted by the small Greak letter σ(read as sigma) and is computed as follows:

$$\sigma=\sqrt{\frac{\Sigma(X-\overline{X})^2}{N}} \text{ or } \sqrt{\frac{\Sigma x^2}{N}} \quad \text{where, } x=X-\overline{X},\ \sigma=S.D.$$

The relative measure of standard deviation, called the **coefficient of S.D.** is obtained by dividing the standard deviation by the arithmetic average. Thus,

$$\text{Coefficient of S.D.}=\frac{\sigma}{\overline{X}}$$

Difference between Mean Deviation and Standard Deviation

Both these measures of dispersion are based on each and every item of the series. But they differ in the following respects:

(1) Algebraic signs of deviations (+ or –) are ignored while calculating mean deviation whereas in the calculation of standard deviation signs of deviations are not ignored, *i.e.*, they are taken into account.

(2) Mean deviation can be computed either from mean, median or mode. The standard deviation, on the other hand, is always computed from the mean because the sum of the squares of the deviations taken from the mean is minimum.

Calculation of Standard Deviation

Individual Series

In case of individual series, standard deviation can be computed by applying any of the three methods:

(1) Actual Mean Method

When deviations are taken from the actual mean, the following formula is used:

$$\sigma=\sqrt{\frac{\Sigma(X-\overline{X})^2}{N}} \text{ or } \sqrt{\frac{\Sigma x^2}{N}} \quad \text{where, } x=X-\overline{X}$$

Steps for Calculation

(*i*) Calculate the actual mean ($\overline{X}$) of the series.

(*ii*) Then take the deviations of the items from the mean, *i.e.*, find $X-\overline{X}$ and denote these deviations by x.

(*iii*) Square these deviations and obtain the total, *i.e.*, Σx^2.

(*iv*) Divide Σx^2 by the total number of items, *i.e.*, N and take the square root of it. The result will give the value of standard deviation.

Example 15. Calculate the standard deviation from the following data:

X: 16, 20, 18, 19, 20, 20, 28, 17, 22, 20

Solution: **Calculation of Standard Deviation**

X	$\overline{X}=20$ $x=X-\overline{X}$	x^2
16	–4	16
20	0	0
18	–2	4
19	–1	1
20	0	0
20	0	0
28	8	64
17	–3	9
22	2	4
20	0	0
$N=10$, $\Sigma X=200$	$\Sigma x=0$	$\Sigma x^2=98$

$$\overline{X}=\frac{\Sigma X}{N}=\frac{200}{10}=20$$

$$\sigma=\sqrt{\frac{\Sigma x^2}{N}}=\sqrt{\frac{98}{10}}=\sqrt{9.8}=3.13$$

(2) Assumed Mean Method

When the actual mean is not a whole number but in fraction, then it becomes difficult to take deviations from mean and then obtain the squares of these deviations. To save time and labour, we use assumed mean method or called **short-cut method.** When deviations are taken from assumed mean, the following formula is used:

$$\sigma=\sqrt{\frac{\Sigma d^2}{N}-\left(\frac{\Sigma d}{N}\right)^2} \quad \text{where, } d=X-A$$

Steps for Calculation

(i) Any one of items in the series is taken as assumed mean, A.

(ii) Take the deviations of the items from the assumed mean, i.e., X – A and denote these deviations by 'd'. Sum up these deviations to obtain Σd.

(iii) Then square these deviations taken from assumed mean and obtain the total, i.e., Σd^2.

(iv) Substitute the value of Σd^2, Σd and N in the above formula. The result will give the value of standard deviation.

Example 16. Calculate the standard deviation of the following series:

7, 10, 12, 13, 15, 20, 21, 28, 29, 35

Use assumed mean method.

Solution:

Calculation of Standard Deviation

X	$A = 20$ $d = X - A$	d^2
7	−13	169
10	−10	100
12	−8	64
13	−7	49
15	−5	25
20 = A	0	0
21	1	1
28	8	64
29	9	81
35	15	225
$N = 10$	$\Sigma d = -10$	$\Sigma d^2 = 778$

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

$$= \sqrt{\frac{778}{10} - \left(\frac{-10}{10}\right)^2} = \sqrt{77.8 - 1} = \sqrt{76.8} = 8.76$$

$$\therefore \quad \sigma = 8.76$$

(3) Method Based on Use of Actual Data

When number of observations are few, standard deviation can be calculated by using the actual data. When this method is used, the following formula is used:

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2} \quad \text{or} \quad \sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}$$

Steps for Calculation

(i) First, we find the sum of the items, i.e., ΣX.

(ii) Then, the values of X are squared up and added to get ΣX^2.

(iii) Substitute the values in the above formula. The result will give the value of standard deviation.

Example 17. Calculate the standard deviation from the following series:

X: 16, 20, 18, 19, 20, 20, 28, 17, 22, 20

Solution:

Calculation of Standard Deviation

X	X^2
16	256
20	400
18	324
19	361
20	400
20	400
28	784
17	289
22	484
20	400
$N = 10$, $\Sigma X = 200$	$\Sigma X^2 = 4098$

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2} = \sqrt{\frac{4098}{10} - \left(\frac{200}{10}\right)^2}$$

$$= \sqrt{409.8 - 400} = \sqrt{9.8} = 3.13$$

Aliter: σ can be calculated by using the formula:

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

Here, $\bar{X} = \frac{\Sigma X}{N} = \frac{200}{10} = 20$, $\Sigma X^2 = 4098$

$$\therefore \quad \sigma = \sqrt{\frac{4098}{10} - (20)^2} = \sqrt{409.8 - 400}$$

$$= \sqrt{9.8} = 3.13$$

Discrete Series

For calculating standard deviation in discrete series, the following three methods may be used:

(1) Actual Mean Method

(2) Assumed Mean Method (or Short-cut Method)

(3) Step Deviation Method.

(1) Actual Mean Method

Under this method, deviations of the items are taken from actual mean, *i.e.*, we find $X - \overline{X}$ and denote these deviations by x. Then these deviations are squared and multiplied by their respective frequencies. The following formula is used:

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} \qquad \text{where, } x = X - \overline{X}$$

However, this method is rarely used in practice because if the actual mean is in fraction, the calculations become tedious and time consuming.

(2) Assumed Mean Method

When this method is applied, the following formula is used:

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \qquad \text{where, } d = X - A$$

Steps for Calculation

(*i*) Take the deviations of the items from an assumed mean instead of actual mean, *i.e.*, $X - A$ and denote these deviations by d.

(*ii*) Then multiply these deviations by their respective frequencies and find Σfd.

(*iii*) Now we square up the deviations, *i.e.*, calculate d^2.

(*iv*) Multiply the squared deviations d^2 by the respective frequencies and find Σfd^2.

(*v*) Substitute the values of Σfd^2, Σfd and N in the above formula.

Example 18. Calculate the standard deviation from the data given below:

X:	3	4	5	6	7	8	9
f:	7	8	10	12	4	3	2

Solution:

Calculation of Standard Deviation

X	f	$A=6$ $d = X-6$	fd	fd^2
3	7	−3	−21	63
4	8	−2	−16	32
5	10	−1	−10	10
6	12	0	0	0
7	4	+1	+4	4
8	3	+2	+6	12
9	2	+3	+6	18
	$N = 46$		$\Sigma fd = -31$	$\Sigma fd^2 = 139$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{139}{46} - \left(\frac{-31}{46}\right)^2}$$

$$= \sqrt{3.0217 - 0.4541} = \sqrt{2.5676} = 1.602$$

(3) Step Deviation Method

This method is used to simplify calculations. Under it, we divide the deviations taken from assumed mean (d) by the common factor and get step deviation d', *i.e.*, $d' = \frac{d}{i}$. The remaining process remains as such mentioned in assumed mean method. The formula for calculating standard deviation is:

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i$$

Where, $d' = \frac{X - A}{i}$ and i = common factor.

Example 19. Calculate standard deviation from the following data:

X:	10	20	30	40	50	60	70
f:	3	5	7	9	8	5	3

Solution:

Calculation of Standard Deviation

X	f	$A = 40$ $d = X - 40$	$d' = \frac{d}{10}$	fd'	fd'^2
10	3	−30	−3	−9	27
20	5	−20	−2	−10	20
30	7	−10	−1	−7	7
40 = A	9	0	0	0	0
50	8	+10	+1	+8	8
60	5	+20	+2	+10	20
70	3	+30	+3	+9	27
	$N = 40$			$\Sigma fd' = +1$	$\Sigma fd'^2 = 109$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{109}{40} - \left(\frac{1}{40}\right)^2} \times 10$$

$$= \sqrt{2.725 - 0.000625} \times 10 = \sqrt{2.724375} \times 10$$

$$= 1.65 \times 10 = 16.5$$

Continuous Series

In continuous series, we can use any of the three methods discussed above for discrete series because the classes are represented by their mid-values. However, in practice, it is the **step deviation method** which is most commonly used. The formula for calculating standard deviation is

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i$$

Where, $d' = \frac{m - A}{i}$ and i = Size of class interval or common factor, m = mid-value of the classes.

▶ **Steps for Calculation**

(*i*) Find out the mid-point of the various classes.

(*ii*) Then take the deviations of these mid-points from assumed mean and denote these deviations by 'd'.

(*iii*) If the class intervals are equal, then divide with common factor (i) and get $d' = \frac{d}{i}$.

(*iv*) Find fd' and fd'^2.

(*v*) Obtain $\Sigma fd'$ and $\Sigma fd'^2$.

(*vi*) Substitute the values of $\Sigma fd'^2$, $\Sigma fd'$ and N in the above formula.

Thus, the only difference in the procedure in case of continuous series is to find the mid-values of the various classes.

Example 20. Calculate mean and standard deviation for the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80
No. of students:	5	10	20	40	30	20	10	4

Solution:

Calculation of Mean and Standard Deviation

Marks	f	M.V. (m)	$d = m - 35$	$d' = \frac{d}{10}$	fd'	fd'^2
0—10	5	5	–30	–3	–15	45
10—20	10	15	–20	–2	–20	40
20—30	20	25	–10	–1	–20	20
30—40	40	35 = A	0	0	0	0
40—50	30	45	+10	+1	+30	30
50—60	20	55	+20	+2	+40	80
60—70	10	65	+30	+3	+30	90
70—80	4	75	+40	+4	+16	64
	$N = 139$				$\Sigma fd' = 61$	$\Sigma fd'^2 = 369$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 35 + \frac{61}{139} \times 10$$

$$= 35 + 4.38 = 39.38$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{369}{139} - \left(\frac{61}{139}\right)^2} \times 10$$

$$= \sqrt{2.6546 - 0.1925} \times 10 = \sqrt{2.4621} \times 10$$

$$= 1.569 \times 10 = 15.69$$

Example 21. Find the mean and standard deviation for the following data:

Age (under):	10	20	30	40	50	60
No. of persons:	15	32	51	78	97	109

Solution: Since the cumulative frequencies are given, firstly we find the simple frequencies:

Age	f	M.V. (m)	d	$d' = \frac{d}{10}$	fd'	fd'^2
0—10	15	5	–30	–3	–45	135
10—20	17	15	–20	–2	–34	68
20—30	19	25	–10	–1	–19	19
30—40	27	35 A	0	0	0	0
40—50	19	45	+10	+1	19	19
50—60	12	55	+20	+2	24	48
	$N = 109$				$\Sigma fd' = -55$	$\Sigma fd'^2 = 289$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 35 - \frac{55}{109} \times 10 = 29.95$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{289}{109} - \left(\frac{-55}{109}\right)^2} \times 10$$

$$= \sqrt{2.6514 - 0.2546} \times 10 = \sqrt{2.3968} \times 10 = 1.548 \times 10 = 15.48$$

IMPORTANT TYPICAL EXAMPLE

Example 22. A welfare organisation introduced an educational scholarship scheme for the school going children of a backward village. The rates of scholarship were fixed as per table:

Age group (in completed years):	5—7	8—10	11—13	14—16	17—19
Amount of scholarship p.m. (in Rs.):	30	40	50	60	70

The ages of 30 school going children were noted as: 11, 8, 10, 5, 7, 12, 7, 17, 5, 13, 9, 8, 10, 15, 7, 12, 6, 7, 8, 11, 14, 18, 6, 13, 9, 10, 6, 15, 13, 5 years respectively. Calculate mean and S.D. of monthly scholarship.

Solution: Firstly we classify data into different groups:

Age group	Tally Bars	No. of children (f)	Amount of Scholarship (X)
5—7	𝍸 𝍸	10	30
8—10	𝍸 III	8	40
11—13	𝍸 II	7	50
14—16	III	3	60
17—19	II	2	70
		$N = 30$	

Calculation of Mean and Standard Deviation

Amt. of Scholarship (X)	No. of Children (f)	$A = 50$ $d = X - 50$	$d' = \frac{d}{10}$	fd'	fd'^2
30	10	–20	–2	–20	40
40	8	–10	–1	–8	8
50	7	0	0	0	0
60	3	+10	+1	3	3
70	2	+20	+2	4	8
	$N = 30$			$\Sigma fd' = -21$	$\Sigma fd'^2 = 59$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

$$= 50 - \frac{21}{30} \times 10 = 50 - \frac{210}{30} = 50 - 7 = 43$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i$$

$$= \sqrt{\frac{59}{30} - \left(\frac{-21}{30}\right)^2} \times 10 = \sqrt{1.966 - 0.49} \times 10$$

$$= \sqrt{1.476} \times 10 = 1.214 \times 10 = 12.14$$

Variance

Variance is another measure of dispersion. The term variance was first used by R.A. Fisher in 1918. **Variance is the square of the standard deviation.** Symbolically,

$$\textbf{Variance} = (S.D.)^2 = \sigma^2$$

Calculation of Variance

(*i*) Variance $= \dfrac{\Sigma f(X - \overline{X})^2}{N}$ (Actual Mean Method)

(*ii*) Variance $= \dfrac{\Sigma fd^2}{N} - \left(\dfrac{\Sigma fd}{N}\right)^2$ (Assumed Mean Method)

(*iii*) Variance $= \left[\dfrac{\Sigma fd'^2}{N} - \left(\dfrac{\Sigma fd'}{N}\right)^2\right] \times i^2$ (Step Deviation Method)

Example 23. Calculate the mean and variance from the data given below:

Daily wages:	0—10	10—20	20—30	30—40	40—50
No. of workers:	2	7	10	5	3

Solution:

Calculation of Mean and Variance

Daily wages	f	M.V. (m)	$A = 25$ $d = m - 25$	$d' = \frac{d}{10}$	fd'	fd'^2
0—10	2	5	–20	–2	–4	8
10—20	7	15	–10	–1	–7	7
20—30	10	25 = A	0	0	0	0
30—40	5	35	+10	+1	+5	5
40—50	3	45	+20	+2	+6	12
	$N = 27$				$\Sigma fd' = 0$	$\Sigma fd'^2 = 32$

$$\overline{X} = = A + \frac{\Sigma fd'}{N} \times i = 25 + \frac{0}{7} \times 10 = 25$$

$$\text{Variance} = \left[\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2\right] \times i^2 = \left[\frac{32}{27} - \left(\frac{0}{27}\right)^2\right] \times 10^2$$

$$\Rightarrow \quad \sigma^2 = 1.185 \times 100 = 118.51$$

$$\therefore \quad \overline{X} = 25,\ \sigma^2 = 118.51$$

EXERCISE 6.4

1. Calculate the standard deviation from the following data:

X:	63	67	64	59	61	67	68	66	63	61	68	61

[Ans. $\sigma = 3$]

2. Calculate the standard deviation from the following data using assumed mean method:

X:	48	75	54	60	63	69	72	51	57	56

[Ans. σ = 8.62]

3. Calculate the mean and standard deviation for the following data:

Size:	10	20	30	40	50	60	70
Frequency:	6	8	16	15	33	11	12

[Ans. $\overline{X} = 44.059$, $\sigma = 16.36$]

4. Calculate mean and standard deviation of the following series:

Daily wages:	0—10	10—20	20—30	30—40	40—50
No. of workers:	2	7	10	5	3

[Ans. $\overline{X} = 25, \sigma = 10.88$]

5. Calculate median and S.D. from the following data:

Variable:	21—25	26—30	31—35	36—40	41—45	46—50	51—55
Frequency:	5	15	28	42	15	12	3

[Ans. $M = 36.928$, $\sigma = 6.735$]