

ગુજરાત રાજ્યના શિક્ષણવિભાગના પત્ર-ક્રમાંક
મશબ/1215/12-22/છ, તા. 1-3-2016 -થી મંજૂર

STATISTICS

Standard 11



PLEDGE

India is my country.

All Indians are my brothers and sisters.

I love my country and I am proud of its rich and varied heritage.

I shall always strive to be worthy of it.

I shall respect my parents, teachers and all my elders and treat everyone with courtesy.

I pledge my devotion to my country and its people.

My happiness lies in their well-being and prosperity.

Price : ₹ 74.00



Gujarat State Board of School Textbooks
'Vidyayan', Sector 10-A, Gandhinagar-382010

© Gujarat State Board of School Textbooks, Gandhinagar

Copyright of this book is reserved by Gujarat State Board of School Textbooks.

No reproduction of this book in whole or in part, in any form is permitted without written permission of the Director, Gujarat State Board of School Textbooks.

Subject Adviser

Dr R. T. Ratani

Authors

Dr M. N. Patel (Convener) Prof. Shubha A. Lagvankar
Dr Chirag J. Trivedi Dr Kunjal H. Shah
Dr Parag B. Shah Shri Maheshbhai A. Patel
Dr Yatin A. Parikh

Translators

Dr M. N. Patel Prof. Shubha A. Lagvankar
Dr Chirag J. Trivedi Dr Kunjal H. Shah
Dr Parag B. Shah

Reviewers

Smt. Lalita K. Patel Shri Dilip S. Gandhi
Shri Girish C. Nayak Smt. Saroj M. Dave
Shri Unnikumaran K. M. Smt. Aparna S. Pathak
Smt. Beena Joseph Shri Nirav B. Shah
Shri Chirag S. Shah Smt. Deepa K. Lalwani

Language Reviewers

Dr Chetan Mewada

Co-ordinator

Dr Chirag N. Shah
(Subject Co-ordinator : Commerce)

Preparation and Planning

Shri Haren Shah
(Dy. Director : Academic)

Lay-out and Planning

Shri Haresh S. Limbachiya
(Dy. Director : Production)

PREFACE

The Gujarat Secondary and Higher Secondary Board has prepared new syllabi in accordance with the syllabi at the national level. These syllabi are approved by the Government of Gujarat.

The Gujarat State Board of School Textbooks takes pleasure in presenting this textbook to the students. It is prepared according to the new syllabus of **Statistics for Standard 11**.

This textbook is written and reviewed by expert teachers and professors. This textbook is published after incorporating the necessary changes suggested by the reviewers.

The Board has taken ample care to make this textbook interesting, useful and free of errors. However, suggestions are welcome to improve the quality of this book from persons taking interest in education.

P. Bharathi (IAS)

Director

Executive President

Date : 10-10-2019

Gandhinagar

First Edition : 2016, Reprint : 2017, 2018, 2019, 2020

Published by : P. Bharathi, Director, on behalf of Gujarat State Board of School Textbooks, 'Vidhyayn', Sector 10-A, Gandhinagar.

Printed by :

FUNDAMENTAL DUTIES

It shall be the duty of every citizen of India* :

- (a) to abide by the Constitution and respect its ideals and institutions, the National Flag and the National Anthem;
- (b) to cherish and follow the noble ideals which inspired our national struggle for freedom;
- (c) to uphold and protect the sovereignty, unity and integrity of India;
- (d) to defend the country and render national service when called upon to do so;
- (e) to promote harmony and the spirit of common brotherhood amongst all the people of India transcending religious, linguistic and regional or sectional diversities; to renounce practices derogatory to the dignity of women;
- (f) to value and preserve the rich heritage of our composite culture;
- (g) to protect and improve the natural environment including forests, lakes, rivers and wild life, and to have compassion for living creatures;
- (h) to develop scientific temper, humanism and the spirit of inquiry and reform;
- (i) to safeguard public property and to abjure violence;
- (j) to strive towards excellence in all spheres of individual and collective activity so that the nation constantly rises to higher levels of endeavour and achievement and
- (k) to provide opportunities for education by the parent, the guardian, to his child, or a ward between the age of 6-14 years as the case may be.

INDEX

1.	Collection of Data	1
2.	Presentation of Data	14
3.	Measures of Central Tendency	67
4.	Measures of Dispersion	121
5.	Skewness of Frequency Distribution	171
6.	Permutations, Combinations and Binomial Expansion	214
7.	Sampling Methods	248
8.	Function	265
9.	Geometric Progression	277
•	Answers	300



“Statistics is a Science that concerns itself with experimentation and the collection, description and analysis of data...Statistical methods are tools for examining data.”

– R. A. Hultquist

1

Collection of Data

Content :

1.1 Origin and Growth of Statistics

1.2 Quantitative and Qualitative Data: Meaning and Differences

1.3 Primary and secondary Data: Meaning and Differences

1.4 Methods of Collecting Primary Data

1.4.1 Direct Inquiry: Meaning, Advantages and Disadvantages

1.4.2 Indirect Inquiry: Meaning, Advantages and Disadvantages

1.4.3 Method of Questionnaire: Meaning, Types, Advantages and Disadvantages

1.4.3.1 Characteristics of an Ideal Questionnaire

1.4.3.2 Questionnaire by Post: Meaning, Advantages and Disadvantages

1.4.3.3 Questionnaire by Enumerators: Meaning, Advantages and Disadvantages

1.5 Secondary Data

1.5.1 Sources of Secondary Data

1.5.2 Precautions while using Secondary Data

1.1 Origin and Growth of Statistics

Indian contribution in the field of statistics has been quite significant from the time of Mauryan Empire (321-296 BC). The book 'Ain-I-Akbari' written by Abu Fazal during the time of Mughal Emperor, Akbar (1596-1597 AD), also mentions a highly developed statistical system for administrative and revenue services of the Mughal empire.

The German word 'Statistik' was first used by Gottfried Achen Wall in 1749 for analysis of data about the state. The contributions to statistics prior to 1750 were based mainly on the examples of data analysis without any use of explicit probabilistic ideas.

By 18th century, the term statistics was used for systematic collection of data by states. Statistics was formally introduced in Encyclopedia Britanica in 1797. The initial results on probability theory were found in 17th and 18th centuries, whose two giants were Laplace (1749-1827) and Gauss (1772-1855). The first statistical body "The Royal Statistical Society" was founded in 1834 in London.

The advanced field of statistics was developed in the late 19th and early 20th century. Galton and Karl Pearson used mathematical statistics in Science, Industry and Politics. Karl Pearson was the founder of mathematical statistics. He founded the advanced statistical laboratory in England and in 1911, the world's first University statistics department at University College, London. In 1910 and 1920 Gosset and Fisher initiated to develop modern statistical science for the data having small samples. Fisher applied statistics to a variety of diversified fields such as genetics, biometry, psychology, education, agriculture, etc. He is well known as father of statistics. During 1930, the refinement and expansion of earlier development was found in the collaborative work between E. Pearson and J. Neyman. Advanced methods of statistics were developed day by day after that.

Growth of Statistics in India

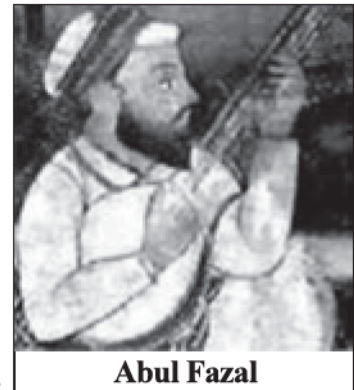
The Indian Statistical Institute (ISI) was founded by the well known Indian Statistician P. C. Mahalanobis in 1931 at Calcutta. He started the first post graduate course in Statistics at Calcutta University in 1941. The Government had approved National Sample Survey proposed by Mahalanobis in 1950 and the first round of data collection took place in October, 1950. Indian Agriculture Statistics Research Institute (IASRI) is another institute which contributed a lot in the development of statistics in India. In 1935 W. F. Willcox listed over a hundred definitions of statistics based on the meaning, scope and limitations of the subject. The definition of Statistics according to Croxton and Cowden is as under.

"Statistics is the Science which deals with the collection, analysis and interpretation of numerical data."

Nowadays, statistics is not only useful for quantitative data but also for qualitative data and considered as an independent part of scientific methods. Industrial statistics and a branch of statistics Operations Research (O.R.) was used in the military projects during the second world war. In India, O.R. came into existence in 1949 with the opening of an O.R. unit at the Regional Research Laboratory in Hyderabad. For national planning and survey in 1953 at the ISI, Calcutta, P. C. Mahalanobis formulated the second five year plan with the help of O.R. techniques.



Ain – I – Akbari

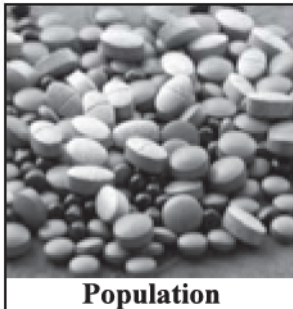


Abul Fazal

O.R. techniques can also be useful for government to maximize the per capita income with minimum resources, in the industry to decide optimum allocation of various resources such as men, machines etc., in marketing to decide the size of the stock to meet the future demand.

Thus, considering the use of Statistics in recent times, its importance cannot be ignored. The constructive development in statistical studies has considerably increased its scope and importance.

1.2 Quantitative Data and Qualitative Data



Meaning and Differences

In statistics, a group of all the units under study is called a **population**. For example, to study the standard of living of workers of a factory, a group of all the workers of the factory becomes a population for this study. The total number of units contained in the population is called the **size of the population**. If the total number of units of the population is finite, say N , we call it a **finite population**. For example, if 700 workers work in a factory then for this population $N = 700$ and we shall call such a population as a finite population. A set or group of units selected from the population on the basis of some definite criterion is called sample and the number of units in the sample is termed as the sample size (n). For example, if we select 150 workers from the above population of workers by any statistical method then the group of selected workers is called sample and sample size $n = 150$. A unit selected in the sample is called a sample unit. The selected workers here are considered as sample units. Inspection of all the units of the population or of the sample is done for a statistical study of a problem covering different aspects. The set of all the observations obtained by such inspection is called **data**.



The units of the population or a sample possess different types of characteristics. It is possible that the measurement of a characteristic varies from unit to unit, such characteristic of a unit is called variable characteristic. It can be either numerical or non numerical.

If the variable characteristic is non numeric then it is called **qualitative variable**. We shall call such qualitative characteristic as an attribute. The collection of observations on the attribute is called **qualitative data**. If the variable characteristic is numerical then it is called **numeric variable** and the collection of observations on the numerical variable is called **quantitative data**. In the above example, the data related to attributes like gender of workers, their education level is called the qualitative data and the data related to numeric variables like monthly income of workers, their age is called the quantitative data.

Thus, we shall say that the qualitative characteristic simply can be observed but cannot be measured on a scale, while the quantitative characteristic can be measured on a scale, e.g. income in rupees, age in years etc. Other examples could be the information related to the characteristics like religion, vegetarian or non-vegetarian, education level, etc are called qualitative data whereas the information related to characteristics like profit of a company, cost of advertising etc are called quantitative data.

1.3 Primary and Secondary Data

Meaning

The data obtained by any statistical inquiry must be correct and proper; otherwise no useful and valid conclusions can be drawn.

The data can be primary data or secondary data.

Primary Data

The data originally collected by any authorized agency or investigators for the first time are called **primary data**. For example, the data collected by NSSO (National Sample Survey Organization), the data of the population census of India, the data relating to death rates and birth rates in India published by office of the Registrar General of India, New Delhi are called **primary data**.

Secondary Data

When an investigator or agency uses the data already collected by any other agency or investigator, then such data becomes secondary data for the users. For example, the data regarding birth rate and death rate in India published by office of the Registrar General of India, New Delhi are reproduced by UNO (United Nations Organization) in the UN Statistical Abstract. For UNO, the data becomes secondary data. The data collected by an agency regarding some industries of Gujarat, which are later used by a research student for statistical analysis, become secondary data for the research student.

Difference Between Primary and Secondary Data

(1) Primary data are obtained for the first time and it is original, whereas secondary data are not original but it is the collected data reused by others.

(2) Primary data are collected from the units under inquiry, whereas secondary data are obtained from the primary data.

(3) Primary data are usually in raw form. Hence, they are required to be classified and tabulated, whereas secondary data are usually in classified and tabulated form.

(4) Collection of primary data requires large amount of money, time and energy, whereas secondary data require comparatively very less amount of money, time and energy.

(5) Primary data are almost reliable; whereas secondary data are obtained by individuals or institutions for their own interest and hence may not be reliable or relevant from the point of view of user's interest or purpose.

1.4 Methods of Collecting Primary Data

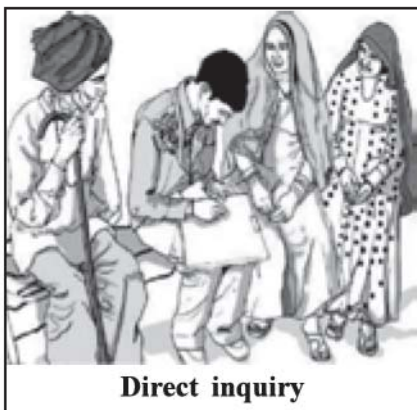
There are many popular methods for the collection of primary data. Commonly used methods are direct inquiry, indirect inquiry, inquiry by telephone or E-mail, method of questionnaire, inquiry by local news reporters.

We shall study the following three methods of collecting primary data :

(1) Direct inquiry (2) Indirect inquiry (3) Method of questionnaire.

1.4.1 Direct Inquiry

Meaning :



In this method an investigator himself or an enumerator appointed by him visits personally to the field and collects the necessary information. For example, if a person wants to collect data about the condition of farmers of Sanad taluka of Ahmedabad district, he must go to Sanand taluka, contact the farmers and obtain the data regarding their condition. The information obtained by this method is called data collected by direct inquiry. Similarly, a person interested in knowing the health status of the students of municipality schools of Ahmedabad city, personally meets the students of the municipality schools and asks questions pertaining to their health and collects the necessary information given by them. This is also called direct inquiry.

Advantages:

- (1) The data obtained by this method are accurate and reliable.
- (2) Due to face to face interviews with respondents, enumerators can solve any question or doubt or any confusion by giving proper explanation.
- (3) By personal interview enumerator sometimes gets supplementary information of the respondents which may be useful at the time of interpretation of the results.
- (4) This method is much better when the field of inquiry is limited.
- (5) Sometimes when respondents are unwilling to give particular type of information, the enumerator can get such information indirectly by asking supplementary questions and giving assurance against the misuse of the information provided by them.

Disadvantages:

- (1) This method requires more time, more number of enumerators when the field of inquiry is very wide.
- (2) This method is more expensive.
- (3) The information collected by this method may be less reliable if the enumerators are untrained.

Activity

Collect the primary data to study any five health characteristics of the students of your school. Identify the qualitative data and quantitative data from this collected data.

1.4.2 Indirect Inquiry**Meaning :**

In any of the following situations, the investigator contacts an organization or a person capable of conducting inquiries and gets relevant information from such organization or person.

- (1) When direct source does not exist.
- (2) When the field of direct inquiry is very large and there is a lack of time for personal contact with the respondents.
- (3) When the information to be obtained is complex.
- (4) When the cost associated with direct inquiry is unbearable.

The method of collecting information in such a way is called indirect inquiry. In this method, the information is obtained with the help of the third party instead of enumerators. For example, in the earlier illustration, a person interested in getting the information about the conditions of farmers of Sanand taluka, approaches the talati of Sanand taluka who has the entire information of all the farmers in Sanand taluka, and gets information from him, instead of personally visiting each farmer.

A person interested in information of the educational status of the workers of an industry, having 2000 workers, approaches the manager of the industry who has complete records of the workers of the industry, instead of making personal visit to each worker. Similarly, advocate can get information regarding murders, thefts from the police station when needed.

Thus, the method of indirect inquiry is quite popular in practice; however the reliability of the information obtained by this method depends on the ability, honesty and experience of the enumerators who collect the information.

Advantages :

- (1) This method is most suitable when the field of inquiry is very wide.
- (2) When a respondent associated with the inquiry is not willing to give information, this is the only useful method to collect the information.
- (3) It involves less cost, time and energy as compared to direct inquiry method.
- (4) This method is very much useful for various departments of government to collect various types of information.

Disadvantages :

- (1) When individual or organization associated with the task of providing information have any prejudices or a biased attitude, the information becomes less reliable.
- (2) The method becomes useless when the third party from whom the information is obtained, is dishonest, inefficient or unable to provide correct information.

1.4.3 Method of Questionnaire

A list of logically arranged questions relevant to the object of the study is made. The space between the questions is provided for the answers. A list of questions prepared in such a way is called questionnaire. The method of obtaining information using such type of questionnaire is called a method of questionnaire.

The questions in the questionnaire should be short and simple so that respondent can understand the questions and easily answer them. This method is quite useful when the field or area from which the information is to be collected is very wide. Since there is considerable saving of time and cost in this method, it is the most economical method of inquiry. Questionnaire may be used in direct inquiry as well as indirect inquiry.

There are two ways of collecting information by questionnaire : (1) By post (2) By enumerators. Before discussing the above two methods in the next section, we shall discuss the points which should be kept in mind for drafting an ideal questionnaire.

1.4.3.1 Characteristics of an Ideal Questionnaire

The success of collecting data depends mainly on the design of questionnaire. A well designed questionnaire is known as an ideal questionnaire. The following points should be kept in mind while drafting an ideal questionnaire:

(1) Every questionnaire must have a covering letter or an appropriate title so that the reader gets clear idea about the purpose of the study.

(2) Number of questions should be as minimum as possible. The fewer number of questions have greater chance of getting better response from the respondent. Long questions discourage the respondent to provide accurate answer particularly towards the end of questionnaire. There is no specific rule for exact number of questions. If the number of questions is large, say 25 or more, it is advisable to divide the questionnaire into various parts to ensure clarity of questions.

(3) The number of questions in the questionnaire should be consistent with the purpose of the inquiry.

(4) Questions should be simple, short and understandable. They should not be ambiguous and dual in meaning. For example, if the question is ‘ Are you educated?’, the respondent cannot understand the meaning of ‘educated’. Does it mean the education up to secondary or higher secondary or graduate? Instead, the question should be ‘What is your education level ?’ Following five options can be considered for the answer to this question.

(a) Upto primary (b) Upto secondary (c) Upto higher secondary (d) Upto graduate (e) Uneducated.

(5) The arrangement of the questions should be logical. The order of the questions should be from general to specific. For example, (a) What is your opinion about the 'smart phone'? (b) Do you use 'smart phone'? This is the improper arrangement of the questions. The correct order is question (b) followed by question (a).

(6) Questions of sensitive nature or personal life should be avoided. For example, questions relating to marital status, other sources of income of respondent, etc. If such questions are necessary in the survey, an assurance should be given to respondent that the information provided by him/her will be kept confidential and will not be used for any other purpose.

(7) Answers to the questions should be free from any calculations. For example, avoid the question like 'What is the Average income of the earning members of your family ?'

(8) The questions having answers YES/ NO or multiple alternatives should be asked. But the questions whose answers do not have specific options should be avoided. For example, what is your opinion about the semester system against the yearly system in the education? Since there are a lot of variations in the answers like: (a) Strongly in favour (b) In favour (c) Neutral (d) Against (e) Strongly against. They are difficult to tabulate or interpret. Hence if necessary, there must be minimum number of such questions in the questionnaire.

Let us see an example of a questionnaire.

A specimen questionnaire for collecting information of the study habits of 12-th standard students of Gujarat state :

1. Name of the student : _____
2. Sex : Female Male
3. Place : _____ Village/ city : _____ District : _____
4. School : Not taking government grant Taking government grant
5. Stream : Arts Commerce Science Others
6. Medium of study : English Gujarati Others
7. School Time : _____
8. How much time do you spend for your daily school home work ? _____
9. Average time spent daily for reading and preparation : _____
10. Mode of entertainment :
 Sports Film or TV Music Spending time with family members
11. How much time do you spend for entertainment ? _____
12. Average hours of daily sleeping time : _____

Activity

Prepare a suitable questionnaire to collect the data regarding the popularity of different types of chocolates among the students of your school.

1.4.3.2 Questionnaire by Post

Meaning :

After preparing an ideal questionnaire, it is dispatched by post to respondents from whom the information is to be collected along with a request to complete and return it in a given period of time. Therefore, it is necessary to send a blank envelope along with the questionnaire to all the respondents bearing the address of investigator and a proper postal stamp. Since the answers to the questions are to be given by respondents themselves, it is essential that the questions in questionnaire are clear, short, simple, relevant and self explanatory. Now a days, questionnaire may be sent through computer (E-mail) and mobile phone also.

Advantages :

- (1) This method of collecting data is quite useful when the field of investigation is very vast.
- (2) This method is simple and it provides large amount of information with less expense.
- (3) By this method, investigator can get information from respondents of those areas where it might be difficult to reach personally or by telephone.

Disadvantages :

- (1) When respondents are illiterate or non cooperative, this method becomes useless.
- (2) Sometimes information provided by the respondents may not be correct or they may return incomplete or blank questionnaire to the investigator.
- (3) There may be a loss of questionnaire or the delay in getting the information because of the laziness of the respondent.
- (4) There is a lack of assistance or a person who clarifies the instructions or gives explanation of questions that may arise. Hence, there is a possibility of misinterpretation of questions.

1.4.3.3 Questionnaire by Enumerator

Meaning :

In this method, enumerators themselves contact the respondents and get response to the questions. Enumerator himself fills the questionnaire. Thus, the difference between the earlier method and this method is that in earlier method the questionnaire is sent to the respondents by post, whereas in this method the enumerators carry the questionnaire and personally meet the respondents. Here the enumerators should be enthusiastic, polite, honest and efficient in their work so that they can extract correct answers to the questions by providing relevant and supplementary information to the respondents. The enumerators are also instructed to create friendly atmosphere without entering into any controversy or showing any disrespect towards the respondents.

Advantages :

- (1) The complete, correct and more relevant information can be obtained from the respondents by selecting proper enumerators.
- (2) Enumerator can get proper information by giving proper explanations to the respondents who are illiterate or not cooperative.
- (3) In this method, there is no problem of loss of questionnaire or incomplete answers to the questions.
- (4) The information obtained is more reliable.

Disadvantages :

- (1) Sometimes it is quite difficult to get expert enumerators in sufficient number.
- (2) Due to the large number of enumerators, the remuneration paid to enumerators increases the total cost of inquiry.

- (3) It is a very difficult task to train the untrained enumerators for the inquiry. Even after training, it is not proper to believe that they will work skillfully, honestly or efficiently. Hence, sometimes the information collected by them may not be reliable.
- (4) This method becomes unsuitable when the respondents are spread over a wide area. In these circumstances, the cost and time of inquiry increases.
- (5) Enumerators have to adjust with the convenient time of respondents or sometimes enumerators have to visit two to three times to the same respondent. As a result, it becomes difficult to finish the inquiry within the stipulated time.

Among the three methods of collecting primary data discussed earlier, there is no definite rule for using a particular method. The selection of a method of inquiry mainly depends on the following aspects:

- (1) Spread of the area under inquiry
- (2) The purpose of inquiry
- (3) The monetary provision
- (4) The time frame
- (5) Possibility of getting expert enumerators and
- (6) The standard of accuracy.

1.5 Secondary Data

1.5.1 Sources of Secondary Data

There are two main sources of secondary data. viz: published and unpublished.

Secondary data from published sources :

(1) Government publications : There are number of central and state government organizations which collect statistical data and publish their findings in the interest of the public. Some government publishers provide data on a periodical and regular basis. For example, Central Statistical Organization (CSO), National Sample Survey Organization (NSSO), Office of the Registrar General and Census Commissioners of India, Indian council of Agricultural Research (ICAR), Statistical Abstract of Gujarat State, Statistical Outline; Gujarat State, Socio-Economic Review; Gujarat State, etc.

From the published data of such government bodies the information about wholesale price index numbers, import-export, vital statistics, agricultural statistics, results of population census can be obtained.

(2) Semi-Government Publications: : Semi-Government organizations such as Life Insurance Corporation of India, State Electricity Boards, etc. regularly publish various types of important data.

(3) International Publications : International organizations like United Nations Organizations (UNO), International Monetary Fund (IMF), International Labour Organization (ILO) publish their important data.

(4) Reports of Research Organizations : Research institutes like Ahmedabad Textile Industry's Research Association (ATIRA), Physical Research Laboratory (PRL); Ahmedabad, Salt and Marine Research Laboratory; Bhavnagar, Institute of Economic Growth; Delhi, National Council of Applied Economic Research; New Delhi provide data in their publications.

(5) Local Self Government Institutions and Autonomous Educational Institutions, Municipal Corporation, Jilla panchayats and agricultural universities publish their annual reports.

(6) Publications of Business and Commerce Organizations : Federation of Indian Chambers of Commerce and Industry publishes the journal ‘Economic Trends’, Institute of Chartered Accountants publishes the journal ‘The Chartered Accountant’ and The Institute of Foreign Trade publishes the journal ‘Foreign Trade Review.’

(7) News papers and Periodicals : The data regarding economics, commerce, business, sports, etc. can be collected from the different news papers and periodicals like Economic and Political Weekly (EPW), Commerce, Business Today, Financial Express etc. These are also important sources of getting secondary data.

Secondary Data from Unpublished Sources :

Some of the statistical data may not be published. Sometimes the data may be drawn on request from the unpublished internal records of private and public organizations which are prepared for their reference. For example, salary of employees, their length of service, their educational level, data regarding investment of mutual funds in public and private sector companies, Ph.D. theses of various universities, etc.

1.5.2 Precautions while using Secondary Data :

A careful scrutiny must be done before using secondary data. The reason is that such data may be erroneous of inadequate sample size or may not be suitable with the purpose of the inquiry. Such secondary data may not be useful for statistical analysis, drawing conclusions and inferences from the analysis. Therefore, before using such data, the following precautions should be taken :

(1) Before using the secondary data, it should be verified as to who has collected the data and from where the data is collected. The data obtained from the records of private organizations may be less reliable as private organizations collect data according to their own ideology and prejudices.

(2) The purpose of collecting data must be relevant to the purpose of the study; otherwise the data will be of no use.

(3) The collected data must not be too old. It should be relevant with the current time period. For example, price of grains, gold, petrol etc.

(4) Before using the secondary data, the matters regarding scope of the data, region for which the data are collected and definition of the terms used in the data should be ascertained.

(5) Direct use of the estimated data should be avoided. The estimates given in the data may be wrongly calculated.

(6) The method of collection of the data should be known so that investigator becomes familiar with its advantages and disadvantages. To get proper benefits of secondary data, it should be used with the above precautions.

Summary

- Data is a set of observations expressed in quantitative or qualitative form.
- Data can be obtained through primary source or secondary source.
- When the data is collected by the investigator himself, it is called primary data.
- When the data has been collected by others and used by investigator then it is known as secondary data.
- The most important method for collecting primary data is by questionnaire method.
- A questionnaire refers to a tool used to get answers of questions from the respondent.

EXERCISE 1

Section A

For the following multiple choice questions choose the correct option.

- Who used the German word 'Statistik' for the first time ?
(a) John Graunt (b) Willian Patty (c) Gottfried Adren Wall (d) Gauss
- Who was one of the giants of initial results of probability theory among the following ?
(a) John Graunt (b) Lapalce (c) Fisher (d) J. Neyman
- Who was the founder of mathematical statistics ?
(a) Karl Pearson (b) Laplace (c) Mahalanobis (d) Gosset
- Out of the following, which one is an example of primary data ?
(a) Data collected from the records of Municipality.
(b) Data Collected from a published journal of an industry.
(c) Data collected from website.
(d) Data collected by NSSO.
- Which one of the following is an example of qualitative data ?
(a) Income category (b) Production (in tons)
(c) Age of workers (in year) (d) Height of persons (in meter)
- Which one of the following is true for secondary data ?
(a) Should never be used.
(b) Use after careful verification.
(c) It is not necessary to check while using it.
(d) Secondary data itself is a primary data.
- Which one of the following is true for primary data ?
(a) Primary data is always more reliable as compared to secondary data.
(b) Primary data is less reliable as compared to secondary data.
(c) Primary data depends on whether the data is collected carefully or not.
(d) Primary data can be obtained from the government publications.
- Which of the following statements is true ?
(a) The data collected by direct inquiry may be more accurate.
(b) The data collected by direct inquiry may be less accurate.
(c) The data collected by direct inquiry may not be reliable.
(d) The data obtained through e-mail is known as the data obtained by direct inquiry.
- Which of the following is a proper method of getting supplementary information about the personal characteristics of the respondents ?
(a) Questionnaire by post (b) Direct inquiry (c) Indirect inquiry (d) From the news papers
- Which method will be costly when the number of respondents are more and spread over the large area ?
(a) Questionnaire by post (b) Indirect inquiry (c) Direct inquiry (d) By telephone

Section B

Answer the following questions in one sentence :

1. Who was the founder of Indian Statistical Institute?
2. Define population.
3. Define sample.
4. Define qualitative data.
5. Define quantitative data.
6. Define primary data.
7. Define secondary data.
8. State the methods of collecting primary data.

Section C

Answer the following questions.

1. State the definition of statistics given by Croxton and Cowden.
2. What is data ?
3. What is questionnaire ?
4. What is unpublished data ?
5. What is a variable characteristic ?
6. What is an attribute ?

Section D

Answer the following questions.

1. What is the role of P. C. Mahalanobis in development of statistics in India?
2. State the difference between qualitative and quantitative data.
3. Give some examples of primary data.
4. Discuss the method of questionnaire.
5. Discuss questionnaire by post.
6. Discuss questionnaire by enumerators.
7. Describe the method of collecting secondary data from unpublished sources.
8. Discuss some applications of statistics.

Section E

Answer the following questions.

1. State the difference between primary and secondary data.
2. Discuss the method of collecting primary data by direct inquiry.
3. Discuss the method of collecting primary data by indirect inquiry.
4. Discuss the method of collecting secondary data.

5. Discuss origin and growth of statistics.
6. Discuss advantages and disadvantages of direct inquiry.
7. Discuss advantages and disadvantages of indirect inquiry.
8. Discuss the characteristics of an ideal questionnaire.
9. Discuss advantages and disadvantages of questionnaire by post.
10. Discuss advantages and disadvantages of questionnaire by enumerators.
11. Discuss what types of precautions should be taken while using secondary data.

● **Historical Note** ●

The term 'Statistics' is derived from the Latin word "Statisticum Collegium" (means council of State) and the Italian word "Statista" (means Statesman or politician). John Graunt and William Petty developed statistical and census methods in 1662 to analyze the bills of mortality.

Some Indian Statisticians who have made a significant contribution in the development of Statistics are : Prof. C. R. Rao, Prof. R. R. Bahadur, Prof. D. Basu, Prof. D. Lahiri, Prof. K. R. Nair, Prof. P. V. Sukhatme, Prof. S. K. Mitra, Prof. R. C. Bose, Prof. S. N. Roy, Prof. N. M. Bhatt, Prof. C. G. Khatri, etc.



P. C. Mahalanobis
(1893 - 1972)

Prasanta Chandra Mahalanobis, popularly known as P. C. Mahalanobis, was an Indian statistician. He devised the Mahalanobis distance, a measure of distance between two populations. It is a fundamental concept in multivariate analysis. He was instrumental in formulating India's strategy for industrialization in the Second Five-Year Plan (1956–61). He founded the Indian Statistical Institute in Kolkata on December 17, 1931.

With the objective of providing comprehensive socioeconomic statistics, Mahalanobis became the pioneer of the establishment of the National Sample Survey in 1950 and also of the Central Statistical Organization to coordinate statistical activities in India. He served as the chairman of the United Nations' Sub-Commission on sampling from 1947 to 1951 and was appointed as the honorary statistical advisor to the Government of India in 1949.

For the pioneering work, he was awarded the Padma Vibhushan, one of India's highest honours, by the Indian government in 1968. A postage stamp was issued by Government of India with his picture.



“Statistics are measurements, enumerations or estimates of natural or social phenomena, usually systematically arranged, analysed and presented as to exhibit important inter–relationships among them.”

– A. M. Tuttle

2

Presentation of Data

Contents :

- 2.1 Classification: Meaning and necessity**
- 2.2 Types of classification:**
 - 2.2.1 Classification of quantitative data
 - 2.2.1.1 Discrete frequency distribution
 - 2.2.1.2 Continuous frequency distribution
 - 2.2.1.3 Cumulative frequency distribution
 - 2.2.1.4 Points for preparing continuous frequency distribution
 - 2.2.2 Classification of qualitative data
 - 2.2.2.1 Simple classification
 - 2.2.2.2 Manifold classification
- 2.3 Tabulation-types and uses**
 - 2.3.1 Guiding rules for tabulation
- 2.4 Diagrams: Importance and limitations of diagrams in statistics**
 - 2.4.1 Types of diagrams
 - 2.4.2 One dimensional diagram
 - 2.4.2.1 Bar diagram
 - 2.4.2.2 Multiple bar diagram
 - 2.4.2.3 Simple divided bar diagram
 - 2.4.2.4 Percentage divided bar diagrams
 - 2.4.3 Two dimensional diagrams
 - 2.4.3.1 Circle diagrams
 - 2.4.3.2 Pie diagrams
 - 2.4.4 Pictorial diagrams

2.1 Classification

Meaning and necessity

In the previous chapter, we have seen that the statistical information consists of two types of data : quantitative data and qualitative data. Quantitative data are based on quantitative variable where as qualitative data are based on qualitative variable. Quantitative variable is of two types (i) discrete variable and (ii) continuous variable. If a variable can assume definite or countable values within the specified range, then it is called discrete variable. For example, number of children per family, number of accidents on road. If a variable x assumes the specific values 1.2, 1.3, 1.5, etc. then it is also called discrete variable. If a variable can assume any value within the specified range, then it is called continuous variable. For example, height of person, maximum temperature on a day, etc. are the examples of continuous variable. In practice, if we have to count something for obtaining the value of variable, then it is called discrete variable. For example, for obtaining the value of number of children per family, we have to count the children, for obtaining the value of number of accident on road, we have to count the accidents, so they are examples of discrete variables. If we have to measure the observation for obtaining the value of variable or if the value of the variable is expressed along with the measuring unit, then it is called continuous variable. Height of person is measured in cm. or inches or feet, maximum temperature is measured in celsius so they are the examples of continuous variables. The data based on discrete variable are called discrete data and the data based on continuous variable are called continuous data.

The data obtained at the end of sample inquiry or population inquiry are called raw data or ungrouped data. These data are in a haphazard form so the statistical analysis becomes difficult. Hence it is necessary to arrange the data in systematic and short form. A process of arranging ungrouped or raw data into proper form is called classification of data and the data thus obtained are called classified data. For example, it is known that the daily demands of a certain commodity during a week are 12, 16, 8, 12, 8, 8, and 10 respectively. From these raw data, it is clear that there are three days during which demand of commodity is 8 units, one day during which the demand is 10 units, two days during which the demand is 12 units and one day during which the demand is 16 units. This information can be presented in the following classified table :

Table showing daily demand of commodity during a week

Demand of commodity	8	10	12	16	Total
Number of days	3	1	2	1	7

Thus, the method of representing raw data into systematic and short form is called classification. In statistical study, the main reasons to classify the data are as follows :

- (1) To represent large data into simple, short and attractive manner.
- (2) For easy comparison between the various characteristics of the data. (In classification data are distributed in different groups according to the similarities of characteristics. Hence the comparisons become simple.)
- (3) To save time, money and labour. (Analysis based on raw or ungrouped data requires more time, money and labour.)
- (4) To obtain information easily, regarding various characteristics of the area under study.

2.2 Types of Classification

There are two types of classifications : (i) classification of quantitative data (ii) classification of qualitative data. Let us understand them by considering the following examples:

Suppose a sample of 100 families is selected from a region and the information regarding the 'number of children per family' is obtained. 100 observations regarding 'number of children per family' are collected and these data are called ungrouped or raw data. Now, on the basis of the study, it is concluded that there are 10 families having no children, 35 families having one child, 40 families having 2 children and 15 families having three children. This is called quantitative classification and concisely it can be represented as under :

Classification of number of children in 100 families

Number of children	0	1	2	3	Total
Number of families	10	35	40	15	100

In the above example, 'number of children per family' is a numeric variable and hence it is called numerical classification or frequency distribution. In the above example, instead of 'number of children per family', if the information regarding 'core occupation of family' is collected and on the basis of it, if it is concluded that core occupation of 30 families is farming, core occupation of 25 is business, core occupation 25 is service and core occupation of remaining 20 families is labour then this is also called classification and can be represented as under :

Tabulation of 100 families with reference to their core occupation

Core occupation of family	Farming	Business	Service	Labour	Total
Number of families	30	25	25	20	100

In the above example, 'core occupation of family' is qualitative variable, so it is called classification of qualitative data or tabulation of data.

Thus, the raw or ungrouped data are mainly classified as (i) classification of quantitative data (ii) qualitative classification of data.

2.2.1 Classification of Quantitative Data

There are two types of quantitative variable (i) discrete variable and (ii) continuous variable. Classification of discrete variable is called discrete frequency distribution and classification of continuous variable is called continuous frequency distribution.

2.2.1.1 Discrete Frequency Distribution

A numeric value showing the repetition of value of an observation is called frequency (f) of that observation. A table showing various possible values of discrete variable with their respective frequencies is called discrete frequency distribution. For understanding, let us consider the following illustration:

During the month of May, information regarding number of accidents per day on a particular road of a city is as under :

0, 1, 3, 2, 0, 3, 4, 5, 0, 1, 0, 2, 0, 1, 3
3, 0, 2, 1, 2, 4, 5, 0, 1, 0, 2, 2, 0, 1, 2, 1

We have to classify these data.

‘The number of accidents per day’ given here is a discrete variable (x) and we prepare its frequency distribution as under:

The minimum value of variable x is 0 and its maximum value is 5. Therefore, the possible values of variable x are 0, 1, 2, 3, 4, and 5. We read the given ungrouped data in sequence and put a tally mark (‘|’) against the value of the variable that we read. If four tally marks are put against any value of the variable then the fifth tally mark is put across to have a group of five tally marks (||||). The purpose of doing this is only to simplify the counting. Continue reading of the raw data till all the vlues are exhausted. Then tally marks are counted against each value of the variable and the respective frequency (f) is obtained. The sum of all frequencies against the value of observations must be the total number of values of the variable i.e. in this illustration $n = \sum f = 31$. The frequency distribution thus obtained is as under :

A discrete frequency distribution of ‘number of road accidents per day’ during the month of May

Number of accidents x	Tally marks	Number of days (f)
0	,	9
1	,	7
2	,	7
3		4
4		2
5		2
	Total	31

The difference between maximum and minimum value of a variable of raw data is called range (R).
i.e. Range $R = \text{Maximum value} - \text{minimum value}$

For the above illustration $R = 5 - 0 = 5$.

Note : When the range of discrete variable is too large then the discrete frequency distribution of that data is not advisable. In such case, inclusive continuous frequency distribution is appropriate, which we will study with continuous frequency distribution.

Illustration 1 : In a television manufacturing company, 500 television sets are produced during a week. A sample of 50 television sets is drawn and each television set is examined. The number of defects per set is given below. Prepare an appropriate frequency distribution.

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

‘Number of defects in a television set’ is a discrete variable and its maximum value is 5 and minimum value is 0. Hence, the range of the given data is $R = 5 - 0 = 5$. Therefore, the discrete frequency distribution of number of defects per television set can be obtained as shown on page no. 18.

A discrete frequency distribution of ‘number of defects per television set’ in a sample of 50 units

Number of defects per television set (x)	Tally marks	Number of television sets (f)
0	 	18
1	 	12
2	 	9
3	 	6
4		3
5		2
	Total	50

Activity

Collect the information regarding the number of family members for 30 families residing around your residence and prepare its frequency distribution.

2.2.1.2 Continuous Frequency Distribution

A continuous frequency distribution is carried out when the variable of raw data is continuous or range of the data is large. For this, first of all the number of classes ‘ K ’ or class length ‘ C ’ is decided for the raw data. Generally, depending on the raw data, any number from 6 to 15 is decided for the number of classes. It should be noted here that this is accepted as a universal trend which may not be true for every study. Under special circumstances, the number of classes may be less than 6 or more than 15. After deciding number of classes, the class length ‘ C ’ is determined for each class. For this purpose, range R of the raw data is divided by the number of classes ‘ K ’. Practically, the value of number of classes K and the value of class length C are decided in such a way that they are positive integers and their product value is at least the value of range R . In notations, $C \cdot K \geq R$. If the class length for different classes are equal then it is called continuous frequency distribution with equal class length. Whenever the range of raw data is too large then depending on the given information, the class length of different classes may be different. It is called frequency distribution with unequal class length. After deciding the class length, class limits are decided for each class. On the basis of class limits, continuous frequency distribution can be categorized in two ways : exclusive frequency distribution and inclusive frequency distribution. If the upper limit of any class and the lower limit of its succeeding class are same then it is called exclusive class. In exclusive class, the observation having value equal to the lower limit is included in that class but the observation having value equal to the upper limit is included in the next class. e.g. for an exclusive class 30 – 35, observation having value 30 is included in this class whereas another observation having value 35 is included in the succeeding class. If the upper limit of any class and lower limit of its succeeding class are not same then it is called inclusive class. In inclusive class,

observation having value equal to the lower limit is included in that class and the observation having value equal to the upper limit is also included in the same class, e.g. for an inclusive class 30 - 35, observation having value 35 is included in this class.

Usually, exclusive continuous frequency distribution is carried out for the continuous raw data and inclusive continuous frequency distribution is carried out for the discrete raw data having large value of range. For converting inclusive continuous frequency distribution into exclusive continuous frequency distribution, each class limit of inclusive continuous frequency distribution is replaced by class boundary point. Lower boundary point and upper boundary point of any class are calculated by using following formulae :

$$\text{Lower boundary point of a class} = \frac{\text{value of lower limit of that class} + \text{value of upper limit of previous class}}{2}$$

$$\text{Upper boundary point of a class} = \frac{\text{value of upper limit of that class} + \text{value of lower limit of succeeding class}}{2}$$

OR

$$\text{Upper boundary point of a class} = \text{lower boundary point} + \text{class length}$$

Thus, the lower boundary point of a class is an average of lower limit of that class and the upper limit of previous class. Similarly, the upper boundary point of a class is an average of the upper limit of that class and the lower limit of succeeding class. It should be noted here that for exclusive classes, their lower limit is the lower boundary point and the upper limit is the upper boundary point. The difference between the upper boundary point and lower boundary point of a class is called class length of that class.

$$\text{Class length} = \text{upper boundary point} - \text{lower boundary point}$$

An average of class limits is called mid point or mid value of that class.

$$\text{Mid value of a class} = \frac{\text{value of upper limit} + \text{value of lower limit}}{2}$$

When mid value of a class and class lengths are known then the class boundary Points of that class can determined by using the following formula :

$$\text{Lower boundary point} = \text{mid value} - \frac{1}{2} (\text{Class length})$$

$$\text{Upper boundary point} = \text{mid value} + \frac{1}{2} (\text{Class length})$$

Let us consider the following illustration for classifying ungrouped continuous data into continuous frequency distribution :

Following are the ages (in years) of employees working in a school.

32	34	48	31	34	27	57	36	49	51
45	29	36	46	46	49	51	47	50	30
35	41	36	47	30	35	48	53	37	47
45	30	50	44	49	43	42	46	28	48
52	36	43	38	39	50	49	34	36	50

Suppose we have to form a continuous frequency distribution of 7 classes from the above data.

Raw data on age of 50 employees of a school are given here and ‘age of employee’ is a continuous variable. We have to prepare its frequency distribution.

The youngest employee is 27 years old and the eldest one is 57 years old.

$$\begin{aligned} \text{Therefore the range } R &= 57 - 27 \\ &= 30 \text{ years} \end{aligned}$$

The data are to be divided into 7 classes $\therefore K = 7$

$$\begin{aligned} \therefore \text{ class length } C &= \frac{\text{Range}}{\text{number of classes}} = \frac{R}{K} \\ &= \frac{30}{7} \\ &= 4.29 \end{aligned}$$

The value of class length is usually taken as a positive integer. Hence, we can take $C = 4$ or $C = 5$. But if we take $C = 4$, then $C \cdot K = 4 \times 7 = 28$, which is less than the value of range. ($\because C \cdot K \geq R$) Therefore, $C = 4$ is not possible. Now if $C = 5$, then $C \cdot K = 5 \times 7 = 35$ which is greater than the value of range. So, $C = 5$.

After deciding $C = 5$ and $K = 7$, the lower class limits for the first class is decided in such a way that it contains the smallest value 27. So, we can take the lower limit as 25 and if we add class length 5 to it then the upper limit is 30. Therefore the first class is 25 – 30, next class is 30 – 35, and so on. Now the last class must contain the maximum value of the observation. Hence, it will be 55 – 60. It should be noted here that for the given data, frequency distribution other than this can also be made.

Now the given ungrouped data can be distributed into different classes by putting tally marks as under:

An exclusive continuous frequency distribution of ‘age of 50 employees of a school’

Age of employee (years) (Exclusive classes)	Tally marks	Number of employees Frequency (<i>f</i>)
25 - 30	III	3
30 - 35	IIII III	8
35 - 40	IIII II	10
40 - 45	IIII	5
45 - 50	IIII III	15
50 - 55	IIII III	8
55 - 60	I	1
	Total	50

Illustration 2 : Figures regarding the sales (in thousand rupees) of different items in a super mall during four weeks are as follows:

228	125	100	90	115	125	230
220	130	80	95	160	180	200
200	128	120	85	185	140	265
230	135	127	100	145	150	210

Construct a frequency distribution by classifying these data into 8 classes.

The variable here is 'sales (in thousand rupees)' which is a continuous variable and the number of classes $K = 8$ is given.

$$\begin{aligned} \text{Range of given data } R &= \text{Maximum value} - \text{Minimum value} \\ &= 265 - 80 \\ &= 185 \end{aligned}$$

$$\begin{aligned} \therefore \text{Class length } C &= \frac{\text{Range}}{\text{Number of class}} = \frac{R}{K} \\ &= \frac{185}{8} \\ &= 23.125 \end{aligned}$$

\therefore For the convenience, class length should be taken as 25, i.e. $C = 25$ ($C \cdot K \geq R$)

By taking the lower limit of the first class as 75 and the upper limit as 100, the minimum value 80 is included in the first class 75 – 100. Similarly the last class 250 – 275 includes the maximum value 265.

An exclusive continuous frequency distribution of 'sales (in thousand rupees)' during for weeks

Sales of items (thousand rupees) classes	Tally marks	Number of days Frequency (f)
75 - 100		4
100 - 125		4
125 - 150		8
150 - 175		2
175 - 200		2
200 - 225		4
225 - 250		3
250 - 275		1
	Total	28

Illustration 3: During a season of 50 days, the data regarding the numbers of roses grown daily on the different plants in a garden are given below. Prepare a frequency distribution from it having one class as 30 – 39.

34 35 37 39 39 54 52 69 71 75 74 76 84 96 23 33 51 39
 26 46 65 65 53 53 72 71 84 94 34 24 99 19 18 27 17 38
 45 55 57 66 82 85 35 19 18 28 47 52 64 75

The variable 'number of roses on different rose plants of a garden' is a discrete variable.

$$\begin{aligned} \text{Range of the given data } R &= 99 - 17 \\ &= 82 \end{aligned}$$

The range of this discrete variable is large, so it is advisable to prepare inclusive continuous frequency distribution. According to the given inclusive class 30 – 39, the class including the lowest value of observation 17 is 10 – 19 and the class including the highest value of observation 99 is 90 – 99.

An inclusive continuous frequency distribution of ‘number of roses on different rose plants of a garden’

Number of roses (Inclusive classes)	Tally marks	Number of days Frequency (f)
10 - 19		5
20 - 29		5
30 - 39		10
40 - 49		3
50 - 59		8
60 - 69		5
70 - 79		7
80 - 89		4
90 - 99		3
Total		50

Illustration 4 : By considering the inclusive frequency distribution obtained in illustration 3, find
(i) Exclusive frequency distribution and (ii) Frequency distribution showing mid value of each class.

(i) In illustration 3, upper limit of the first class is 19 and the lower limit of the second class is 20 and its class length is 10. Lower boundary point of the second class is $\frac{20+19}{2} = 19.5$ and its upper boundary point is $19.5 + 10 = 29.5$. Thus, the class boundaries for the second class are 19.5 – 29.5 and class boundaries for the first class are 9.5 – 19.5.

The alternate method for calculating these is as under :

Upper limit for the first class is 19 and the lower limit for the next class is 20. The difference of these two limits ($20 - 19 = 1$) is divided by 2, which gives 0.5. Now subtract 0.5 from the lower limit and add 0.5 to the upper limit of a each class. We have the lower boundary point and upper boundary point of each class as under :

Number of roses (exclusive classes)	Number of days Frequency (f)
9.5 – 19.5	5
19.5 – 29.5	5
29.5 – 39.5	10
39.5 – 49.5	3
49.5 – 59.5	8
59.5 – 69.5	5
69.5 – 79.5	7
79.5 – 89.5	4
89.5 – 99.5	3
Total	50

(ii) Mid value of a class is obtained by taking average of the upper limit and the lower limit of that class. Hence the frequency distribution showing the mid value of each class is as under :

Number of roses Inclusive classes	Mid value $= \frac{\text{upper limit} + \text{lower limit}}{2}$	Number of days Frequency (<i>f</i>)
10 - 19	$\frac{10+19}{2} = 14.5$	5
20 - 29	$\frac{20+29}{2} = 24.5$	5
30 - 39	$\frac{30+39}{2} = 34.5$	10
40 - 49	$\frac{40+49}{2} = 44.5$	3
50 - 59	$\frac{50+59}{2} = 54.5$	8
60 - 69	$\frac{60+69}{2} = 64.5$	5
70 - 79	$\frac{70+79}{2} = 74.5$	7
80 - 89	$\frac{80+89}{2} = 84.5$	4
90 - 99	$\frac{90+99}{2} = 94.5$	3
Total		50

Illustration 5: The following frequency distribution of unequal class length is obtained for some data. Prepare a frequency distribution stating class length and mid value of each class.

classes	0 - 20	20 - 50	50 - 70	70 - 90	90 - 100	Total
Frequency	20	30	30	15	5	100

A frequency distribution showing mid value and class length of each class is as under :

Classes	Class length	Mid value	Frequency
0 - 20	$20 - 0 = 20$	$\frac{0+20}{2} = 10$	20
20 - 50	$50 - 20 = 30$	$\frac{20+50}{2} = 35$	30
50 - 70	$70 - 50 = 20$	$\frac{50+70}{2} = 60$	30
70 - 90	$90 - 70 = 20$	$\frac{70+90}{2} = 80$	15
90 - 100	$100 - 90 = 10$	$\frac{90+100}{2} = 95$	5
Total			100

Activity

For 30 families residing around your residence, collect information regarding the height of the eldest person of each family and prepare its frequency distribution.

2.2.1.3 Cumulative Frequency Distribution

Sum of the frequencies upto the value of an observation or a class is called cumulative frequency (cf) and its distribution is called cumulative frequency distribution.

The sum of the frequencies upto the specified value of the observation or specified upper boundary point of a class is called 'less than' type cumulative frequency of that specified value and the distribution is called 'less than' type cumulative frequency distribution.

The sum of the all frequencies of the specified value of the observation or the lower limit of the specified class and the values or classes succeeding to it is called as 'more than' type cumulative frequency. Its distribution is called 'more than' type cumulative frequency distribution.

When the cumulative frequency distribution is obtained by considering the value of a discrete variable, it is called discrete cumulative distribution, whereas the cumulative distribution obtained by considering the boundary points is called continuous cumulative frequency distribution.

Illustration 6 : A frequency distribution of number of children in 50 families of a region is as under :

Number of children (x)	0	1	2	3	Total
Number of families (f)	10	25	12	3	50

Obtain 'less than' type and 'more than' type cumulative frequency distributions for these data.

This distribution is for discrete variable. Therefore, 'less than' and 'more than' type cumulative frequency distributions can be obtained as under:

'less than' type discrete cumulative frequency distribution

Number of children (x)	Number of families (f)	Number of children less than or equal to x ($\leq x$)	Cumulative frequency (cf)
0	10	0	$10 = 10$
1	25	1	$10 + 25 = 35$
2	12	2	$10 + 25 + 12 = 47$
3	3	3	$10 + 25 + 12 + 3 = 50$
Total	50		

‘More than’ type discrete cumulative frequency distribution

Number of children (x)	Number of families (f)	Number of children x or above($\geq x$)	Cumulative frequency (cf)
0	10	0	$3 + 12 + 25 + 10 = 50$
1	25	1	$3 + 12 + 25 = 40$
2	12	2	$3 + 12 = 15$
3	3	3	$3 = 3$
Total	50		

Illustration 7: A frequency distribution of monthly income of 500 persons is as under. Obtain ‘less than’ type and ‘more than’ type cumulative frequency distributions :

Monthly Income (Thousand ₹)	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 55	55 - 60	Total
Number of persons (f)	30	80	100	50	150	80	10	500

An exclusive continuous frequency distribution of monthly income of 500 persons is given here and the following ‘less than’ type cumulative frequency distribution is obtained from it by considering upper boundary points. It is clear from the given data that no one is earning less than ₹ 25000. For indicating this, the lower boundary point of the first class is taken as upper boundary point of previous class with frequency zero.

‘less than’ type cumulative frequency distribution showing the monthly income of 500 persons

monthly income lesser than upper boundary point (Thousand rupees)	‘less than’ type cumulative frequency (cf)
25	$0 = 0$
30	$0 + 30 = 30$
35	$0 + 30 + 80 = 110$
40	$0 + 30 + 80 + 100 = 210$
45	$0 + 30 + 80 + 100 + 50 = 260$
50	$0 + 30 + 80 + 100 + 50 + 150 = 410$
55	$0 + 30 + 80 + 100 + 50 + 150 + 80 = 490$
60	$0 + 30 + 80 + 100 + 50 + 150 + 80 + 10 = 500$

Now, by considering the lower boundary point of each class, ‘more than’ type cumulative frequency distribution is obtained as under. It is clear from the given data that no one is earning more than ₹ 60,000. For indicating this, the upper boundary point of the last class is taken as the lower boundary point of the next class with frequency zero.

‘more than’ type cumulative frequency distribution showing the monthly income of 500 persons

Income more than or equal to lower boundary point (thousand ₹)	‘more than’ type cumulative frequency
25	$10 + 80 + 150 + 50 + 100 + 80 + 30 = 500$
30	$10 + 80 + 150 + 50 + 100 + 80 = 470$
35	$10 + 80 + 150 + 50 + 100 = 390$
40	$10 + 80 + 150 + 50 = 290$
45	$10 + 80 + 150 = 240$
50	$10 + 80 = 90$
55	$10 = 10$
60	$0 = 0$

Illustration 8: The frequency distribution of daily demand of rooms at an international hotel during 90 days is as under. Obtain ‘less than’ type and ‘more than type’ cumulative frequency distribution from it.

Demand of rooms	1 - 50	51 - 100	101 - 150	151 - 200	201 - 250	Total
Number of days	10	20	30	18	12	90

This inclusive continuous frequency distribution can be expressed as exclusive continuous frequency distribution as under :

Demand of rooms	No. of days
0.5 - 50.5	10
50.5 - 100.5	20
100.5 - 150.5	30
150.5 - 200.5	18
200.5 - 250.5	12
Total	90

From the above exclusive classes, the ‘less than’ and ‘more than’ cumulative frequency distributions can be obtained as under :

‘less than’ type cumulative frequency distribution showing the demand of rooms at the hotel during 90 days

Demand less than upper boundary	‘less than’ type cumulative frequency
0.5	$0 = 0$
50.5	$0 + 10 = 10$
100.5	$0 + 10 + 20 = 30$
150.5	$0 + 10 + 20 + 30 = 60$
200.5	$0 + 10 + 20 + 30 + 18 = 78$
250.5	$0 + 10 + 20 + 30 + 18 + 12 = 90$

**‘More than’ type cumulative frequency distribution showing
the demand of rooms at hotel during 90 days**

Demand more than or equal to lower boundary point	‘More than’ type cumulative frequency
0.5	$12 + 18 + 30 + 20 + 10 = 90$
50.5	$12 + 18 + 30 + 20 = 80$
100.5	$12 + 18 + 30 = 60$
150.5	$12 + 18 = 30$
200.5	$12 = 12$
250.5	$0 = 0$

Illustration 9 : ‘Less than’ type cumulative frequency distribution of weight (in kg.) of 50 persons is given in the following table.

less than upper boundary point Weight (kg.)	‘less than’ type cumulative frequency
30	0
35	7
40	15
45	30
50	38
55	44
60	47
65	49
70	50

- (i) How many persons have weight less than 45 kg. ?
(ii) How many persons have weight between 50 kg. and 65 kg. ?
(iii) Obtain the original frequency distribution.

(i) From the given table, it is clear that 30 persons have weight less than 45 kg.

(ii) Number of persons having weight less than 65 kg. = 49

Number of persons having weight less than 50 kg. = 38

∴ Number of persons having weight between 65 and 50 kg. = $49 - 38 = 11$

(iii) Original distribution of weights of 50 persons is as under :

Weight (kg.) Class	Number of persons (frequency)
30 - 35	$7 - 0 = 7$
35 - 40	$15 - 7 = 8$
40 - 45	$30 - 15 = 15$
45 - 50	$38 - 30 = 8$
50 - 55	$44 - 38 = 6$
55 - 60	$47 - 44 = 3$
60 - 65	$49 - 47 = 2$
65 - 70	$50 - 49 = 1$
Total	50

Illustration 10 : From the following 'more than' cumulative frequency distribution,

(i) Determine the number of the persons having age 40 or more. (ii) Determine the number of the persons having age less than 40 (iii) Determine original frequency distribution.

Lower boundary point or more than it	25	30	35	40	45	50	55	60	65	70
Cumulative frequency of persons	100	96	87	70	45	25	14	6	1	0

(i) The number of persons having age 40 or more = 70

(ii) Total frequency = 100

Number of persons having age less than 40

= Total number of persons – number of persons having age 40 or more.

= $100 - 70 = 30$

(iii) Original frequency distribution can be obtained as under :

Age	Number of Persons
25 - 30	$100 - 96 = 4$
30 - 35	$96 - 87 = 9$
35 - 40	$87 - 70 = 17$
40 - 45	$70 - 45 = 25$
45 - 50	$45 - 25 = 20$
50 - 55	$25 - 14 = 11$
55 - 60	$14 - 6 = 8$
60 - 65	$6 - 1 = 5$
65 - 70	$1 - 0 = 1$
Total	100

2.2.1.4 Points for preparing continuous frequency distribution :

Some important points to be considered for preparing continuous frequency distribution from ungrouped data are as under :

(1) Normally, continuous frequency distribution may consist minimum 6 and maximum 15 classes to represent the raw data. In exceptional situation, the number of classes may be other than these, depending upon the given data.

(2) Total number of classes is denoted by K .

(3) Range R is obtained for the given raw data.

$$\text{Range } R = \text{Maximum value of observations} - \text{Minimum value of observations.}$$

(4) The value of class length C is determined by $C = \frac{R}{K} = \frac{\text{Range}}{\text{No. of classes}}$ such that $C \cdot K \geq R$.

(5) Depending on the class length, class limits are decided in such a way that the first class must include the value of the minimum observation of the raw data and the last class must include the maximum value of observations of the raw data. Usually, frequency distribution with equal class length is preferred but when the range of the data is large then irregular class length is preferred.

(6) When mid values of the class and class lengths of a distribution are known, class boundary points can be determined by using the following formula.

$$\text{Lower boundary point} = \text{mid value} - \frac{1}{2} (\text{Class length})$$

$$\text{Upper boundary point} = \text{mid value} + \frac{1}{2} (\text{Class length})$$

(7) Usually, exclusive classes are to be prepared for the continuous data. Whereas, when the range of the discrete data is large then it is a normal practice to prepare inclusive classes.

(8) Inclusive classes should be converted into exclusive classes for preparing cumulative frequency distribution.

(9) In order to prepare 'less than' type cumulative frequency distribution from the continuous frequency distribution, the lower boundary point of the first class is taken as upper boundary point of previous class with respective frequency zero whereas in 'more than' type cumulative frequencies the upper boundary point of the last class is taken as the lower boundary point of the next class and its respective frequency is also taken as zero.

Note : By the classification of the raw data into continuous frequency distribution, the approximate values are used instead of original data. e.g. if the value of one observation is 8, then it is included in the class 0 – 10. Thus, the original value 8 is lost in such a classification. But on the basis of classification of the data, we can easily get idea about the spread and other characteristics of the data.

Illustration 11: Information regarding the ages of 250 drivers of a transport company is as under. Obtain continuous frequency distribution of equal class length from it.

Mid value of age of driver	22.5	27.5	32.5	37.5	42.5	47.5
Number of drivers (Frequency f)	25	30	50	80	50	15

'Age of driver' is a continuous variable. Hence, from the given data, it is advisable to prepare exclusive continuous frequency distribution.

Difference between two successive mid values is 5. So, the class length, $C = 5$

$$\begin{aligned} \text{Lower limit of the first class} &= \text{Mid value} - \frac{1}{2}(\text{Class length}) \\ &= 22.5 - \frac{1}{2}(5) \\ &= 22.5 - 2.5 \\ &= 20 \end{aligned}$$

$$\begin{aligned} \text{Upper limit of the first class} &= \text{Mid value} + \frac{1}{2}(\text{Class length}) \\ &= 22.5 + \frac{1}{2}(5) \\ &= 22.5 + 2.5 \\ &= 25 \end{aligned}$$

∴ the first class is 20–25 and similarly the class limits are obtained for the remaining classes.

An exclusive continuous classification showing the age of 250 drivers

Age of Driver	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	Total
Number of drivers	25	30	50	80	50	15	250

Illustration 12: Information regarding the number of errors per page in a book of 500 pages is as under. Find the inclusive frequency distribution from it.

Mid value of a class of number of errors per page	0.5	2.5	4.5	6.5	8.5
Number of pages (f)	380	100	12	6	2

Number of errors per page of a book is a discrete variable. Hence, for the given data, it is advisable to have inclusive continuous frequency distribution.

Difference between two successive mid value is 2. So, class length, $C = 2$

$$\begin{aligned} \text{Lower limit of the first class} &= \text{Mid value} - \frac{1}{2}(\text{Class length}) \\ &= 0.5 - \frac{1}{2}(2) \\ &= 0.5 - 1.0 \\ &= -0.5 \end{aligned}$$

$$\begin{aligned} \text{Upper limit of the first class} &= \text{Mid value} + \frac{1}{2}(\text{Class length}) \\ &= 0.5 + \frac{1}{2}(2) \\ &= 0.5 + 1 \\ &= 1.5 \end{aligned}$$

∴ the first class is –0.5 to 1.5 and similarly the class limits are obtained for the remaining classes.

Number of errors per page	0.5 - 1.5	1.5 - 3.5	3.5 - 5.5	5.5 - 7.5	7.5 - 9.5	Total
Number of pages(f)	380	100	12	6	2	500

The above exclusive continuous frequency distribution can be converted into inclusive continuous frequency distribution by adding 0.5 to the lower boundary point and subtracting 0.5 from the upper boundary point.

An inclusive continuous frequency distribution of number of errors per page of a book

Number of errors per page	0 - 1	2 - 3	4 - 5	6 - 7	8 - 9	Total
Number of pages	380	100	12	6	2	500

Note : The constant added to the lower boundary point and subtracted from the upper boundary point for converting exclusive continuous frequency distribution into inclusive continuous frequency distribution is decided on basis of the given data.

Illustration 13 : The daily record of sales of electronic equipments sold by a shopkeeper of electronic items during the month of April in as under. Prepare an exclusive continuous frequency distribution for which one of the classes is 60 - 70. Hence,

- (i) Find the number of days during which the sales of equipment is maximum.
- (ii) State the number of equipments sold for the maximum number of days.

54 58 52 73 57 39 46 64 49 53 75 34 57 68 51
44 34 40 82 88 80 36 85 66 58 41 62 72 80 81

The variable ‘number of electronic equipments’ sold is a discrete variable but it is clearly given that we have to prepare an exclusive continuous frequency distribution for which one of the classes is 60 – 70.

First class = class including of the smallest value 34 = 30 – 40

Last class = class including the maximum value 88 = 80 – 90

An exclusive continuous frequency distribution of ‘number of electronic equipments’ sold

Number of equipments sold Exclusive classes	Tally marks	Number of days Frequency (<i>f</i>)
30 - 40	IIII	4
40 - 50	IIII	5
50 - 60	IIII III	8
60 - 70	IIII	4
70 - 80	III	3
80 - 90	IIII I	6
	Total	30

- (i) The maximum number of equipments sold is 80 to 90 and there are 6 such days.
- (ii) 50 to 60 equipments are sold on maximum 8 days.

Illustration 14: Out of 300 persons residing in a region, a sample of 30 persons is selected at random and the heights (in cm.) of these selected persons are as under :

163 148 151 162 145 152 149 158 153 149
150 152 145 141 162 168 148 158 149 141
146 155 159 150 161 153 162 160 154 165

- (i) Distribute these data into 6 classes and also find the mid value of each class.
- (ii) Obtain 'less than' type cumulative frequency distribution.
- (iii) What is the percentage of persons having height less than 155 cm.?
- (iv) Obtain 'more than' type cumulative frequency distribution.
- (v) How many persons have the height between 147 to 157 cm.?
- (i) 'Height of a person' is a continuous variable.

$$\begin{aligned} \text{Range of data } R &= 168-141 \\ &= 27 \end{aligned}$$

$$\text{Number of classes } K = 6$$

$$\text{Class interval } C = \frac{R}{K} = 4.5 \approx 5$$

An exclusive continuous frequency distribution of height of 30 persons selected at random from a group of 300 persons

Height (cm.) Exclusive classes	Tally marks	Number of persons Frequency (<i>f</i>)	Mid value
140 - 145		2	142.5
145 - 150	III	8	147.5
150 - 155	III	8	152.5
155 - 160		4	157.5
160 - 165	I	6	162.5
165 - 170		2	167.5
	Total	30	

(ii) 'less than' type cumulative frequency distribution of heights of 30 persons selected at random from a group of 300 persons is as under :

Upper boundary point (less than height)	'less than' type cumulative frequency (<i>cf</i>)
140	0 = 0
145	0 + 2 = 2
150	0 + 2 + 8 = 10
155	0 + 2 + 8 + 8 = 18
160	0 + 2 + 8 + 8 + 4 = 22
165	0 + 2 + 8 + 8 + 4 + 6 = 28
170	0 + 2 + 8 + 8 + 4 + 6 + 2 = 30

(iii) From the above table, it is clear that the number of persons having height less than 155 cm. = 18
 \therefore the percentage of persons = $\frac{18}{30} \times 100 = 60\%$

(iv) 'More than' type cumulative frequency distribution can be obtained as under :

Lower boundary point (minimum height)	'More than' type cumulative frequency
140	$2 + 6 + 4 + 8 + 8 + 2 = 30$
145	$2 + 6 + 4 + 8 + 8 = 28$
150	$2 + 6 + 4 + 8 = 20$
155	$2 + 6 + 4 = 12$
160	$2 + 6 = 8$
165	$2 = 2$
170	$0 = 0$

(v) To determine the number of persons having height 147 to 157 cm., we will use the original frequency distribution.

147 is included in the class 145 - 150, which has frequency 8.

Thus, when the class length is 5 (145 - 150) then frequency is 8.

When the class length is 3 (147 - 150) then the frequency is $\frac{3}{5} \times 8 = 4.8$

The frequency of the class 150 - 155 is 8 and the frequency for the class 155 - 157 can be calculated as follows:

When the class length is 5 (155 - 160) then frequency is 4

\therefore When the class length is 2 (155 - 157) then the frequency is $= \frac{2}{5} \times 4 = 1.6$

\therefore the total number of persons having height 147 to 157 cm. = $4.8 + 8 + 1.6 = 14.4 \approx 14$

Note : If we count the number of persons having height between 147 and 157 cm. from the given ungrouped data, we find 15 persons. But according to above calculation, it is 14. The obvious reason for this difference is, by classification into continuous frequency distribution, approximate values are used instead of the original raw data.

Illustration 15 : Changes in prices of shares during a day of 40 different companies registered at Bombay stock exchange are as follows. Find an inclusive continuous frequency distribution having mid value -1 for one of the classes and regular class length as 5.

- 8 8 7 16 8 22 6 10 - 7 5
 3 - 4 9 - 11 11 16 9 - 3 - 11 2
 5 - 6 10 - 6 13 - 5 3 - 7 12 0
 7 6 12 - 5 21 0 4 - 10 14 - 2

‘Changes in the price of share’ is a continuous variable.

The class limits of a class having mid value -1 and class length $C = 5$ are :

$$\begin{aligned} \text{Lower limit} &= \text{Mid value} - \frac{1}{2} C \\ &= -1 - \frac{1}{2} (5) \\ &= -3.5 \end{aligned}$$

$$\begin{aligned} \text{Upper limit} &= \text{Mid value} + \frac{1}{2} C \\ &= -1 + \frac{1}{2} (5) \\ &= 1.5 \end{aligned}$$

Hence that class is -3.5 to 1.5 which is exclusive class.

Converting it into the inclusive class, we get $-3.5 + 0.5$ to $1.5 - 0.5$
 $= -3$ to 1

In the given data, the minimum value is -11 and its maximum value is 22 . Hence the remaining classes should be made by including these values.

An inclusive continuous frequency distribution of change in price of shares of 40 compaines

Change in the price of shares	Tally marks	Number of companies (f)
-13 to -9		3
-8 to -4		8
-3 to 1		4
2 to 6		8
7 to 11		9
12 to 16		6
17 to 21		1
22 to 26		1
	Total	40

Illustration 16 : The monthly salaries (in rupees) of 24 employees working in a private company are given below. Prepare an appropriate frequency distribution from it :

3000, 3500, 4200, 5600, 7500, 9100, 10600, 16200, 18100, 24000, 30000, 36000
 3200, 3800, 5200, 7000, 8400, 9600, 12800, 17700, 22750, 24900, 34000, 40000

‘Salary of employee’ is considered as a continuous variable. Range $R = 40000 - 3000 = 37000$ is too large. Hence, we can form a frequency distribution with unequal class lengths. The class lengths of different classes will be decided by studying the given raw data.

A continuous frequency distribution with unequal class length of salary of 24 employees of a company

Salary of employee (₹) (classes)	Tally marks	Number of employees frequency (<i>f</i>)
3000 - 5000		5
5000 - 10000		7
10000 - 20000		5
20000 - 25000		3
25000 - 45000		4
	Total	24

Note : Here it is possible to have frequency distribution other than these classes.

Illustration 17 : Obtain the original frequency distribution from the following frequency distribution:

(i)

Class	24 - 29	24 - 34	24 - 39	24 - 44	24 - 49	24 - 54	24 - 59	24 - 64
Cumulative frequency	3	12	30	55	78	88	95	100

(ii)

Class	10 - 90	20 - 90	30 - 90	40 - 90	50 - 90	60 - 90	70 - 90	80 - 90
Cumulative frequency	200	180	140	90	55	30	8	3

(i) 'Less than' type cumulative frequency distribution is given here. The original frequency distribution can be obtained from it as under :

Classes	Cumulative frequency	Frequency (<i>f</i>)
24 - 29	3	3
29 - 34	12	12 - 3 = 9
34 - 39	30	30 - 12 = 18
39 - 44	55	55 - 30 = 25
44 - 49	78	78 - 55 = 23
49 - 54	88	88 - 78 = 10
54 - 59	95	95 - 88 = 7
59 - 64	100	100 - 95 = 5
	Total	100

(ii) 'More than' type cumulative frequency distribution is given here. The original frequency distribution can be obtained from it as under :

Classes	Cumulative frequency	Frequency (f)
10 - 20	200	$200 - 180 = 20$
20 - 30	180	$180 - 140 = 40$
30 - 40	140	$140 - 90 = 50$
40 - 50	90	$90 - 55 = 35$
50 - 60	55	$55 - 30 = 25$
60 - 70	30	$30 - 8 = 22$
70 - 80	8	$8 - 3 = 5$
80 - 90	3	$= 3$
	Total	200

EXERCISE 2.1

1. The data regarding the number of children of 50 families residing in a certain area are given below. Prepare appropriate frequency distribution :

1	1	2	1	1	1	1	2	1	0
0	2	2	0	3	3	2	1	2	1
2	1	3	1	1	2	2	2	1	2
3	0	3	0	2	1	2	2	2	2
0	1	2	2	2	2	3	3	2	1

2. The ages (in full years) of 60 employees working in an office are registered as follows. Prepare a frequency distribution by taking class length as 5 from this information.

32	42	48	35	23	58	52	38	36	44	48	39
24	27	29	32	34	41	45	51	30	47	45	44
52	38	41	31	25	38	36	34	37	51	25	56
32	39	32	35	42	26	46	42	57	28	43	33
31	42	43	53	43	39	27	54	21	47	26	40

3. The data regarding the number of mobile phones produced during last 60 days by a mobile phone manufacturing company is given below. Distribute it into 10 classes :

699	380	625	653	452	763	385	959	485	970
749	595	1029	500	499	453	525	621	465	565
103	785	286	1060	760	355	645	775	825	235
390	399	530	540	695	999	849	550	720	430
752	389	1075	701	875	552	351	265	199	370
1025	825	783	225	603	553	503	663	385	465

Obtain 'less than' and 'more than' type cumulative frequency distribution from it.

4. Rewrite the following frequency distribution by stating class length and mid value of each class.

Class	0 - 99	100 - 299	300 - 499	500 - 749	750 - 899	900 - 999
Frequency	10	12	14	16	8	10

5. From the following frequency distribution, obtain 'less than' and 'more than' type cumulative frequency distributions.

Number of errors per page	0	1	2	3
Number of pages	140	110	120	30

6. Obtain an inclusive continuous frequency distribution from the following data.

Lower boundary point or more than that	44.5	49.5	54.5	59.5	64.5	69.5	74.5	79.5
Cumulative frequency	500	470	390	290	240	90	10	0

7. Obtain an exclusive continuous frequency distribution from the following data.

Less than weight (kg.)	30	35	40	45	50	55	60	65	70
Cumulative frequency	0	17	25	40	48	54	57	59	60

8.

Mid values	25	105	230	400	650	900	Total
Frequency	10	30	40	60	80	30	250
Class length	50	110	140	200	300	200	

Obtain the original frequency distribution.

9. Information regarding the number of accidents in a city during a year is as under. Find the inclusive continuous frequency distribution from it.

No. of accidents (Mid values)	11.5	21.5	31.5	41.5	51.5	Total
No. of days	160	120	43	40	2	365

10. From the following data, obtain class boundary points from the class limits and write the frequency distribution.

Class	1 - 1.475	1.5 - 1.975	2 - 2.475	2.5 - 2.975	3 - 3.475	3.5 - 3.975	Total
Frequency	5	10	20	20	10	5	70

2.2.2 Classification of Qualitative Data

The process of systematic arrangement of qualitative or attribute raw data into rows and/or columns on the basis of standards of classification is called classification of qualitative data. Classification of qualitative data is done to represent qualitative ungrouped data into short and attractive form which is also called tabulation. Usually, there are two types of classification : (i) Simple classification or table and (ii) Manifold classification or table.

2.2.2.1 Simple Classification

A classification prepared on the basis of a single attribute is called simple classification or tabulation. e.g. The employees of a bank can be classified as a manager, a clerk, a peon, a security personnel, etc. based on their designation or position in the bank.

Illustration 18: On the basis of a study of different branches of a co-operative bank of Ahmedabad city, the following information is obtained. There are 20 security personnel, 30 peons, 40 clerks and 8 managers in the bank. Express this information in a table.

In the given data, 'designation of employee' is a qualitative characteristic.

A Table showing designation of the employees working at different branches of a co-operative bank of Ahmedabad city

Designation of employee	Number of employees
Security personal	20
Peon	30
Clerk	40
Manager	8
Total	98

2.2.2.2 Manifold Classification

A classification of raw data carried out by considering more than one attribute of the units under study is called manifold classification.

Illustration 19: On the basis of the study of different branches of a co-operative bank of Ahmedabad city, the following information is obtained. In this bank, out of 20 employees working as security personal, 6 are females, out of 30 peons, 10 are females, out of 40 clerks, 25 are females and out of 8 managers, 3 are females.

Express this information in a table.

In the given data, there are two attributes (i) 'designation of employee' and (ii) 'gender of employee'.

A Table showing designation and gender of the employees working at different branches of a co-operative bank of Ahmedabad city

Designation of employee	Gender		Total
	Male	Female	
Security person	14	6	20
Peon	20	10	30
Clerk	15	25	40
Manager	5	3	8
Total	54	44	98

Note : The figures shown in bold letters are the values given in the description of the illustration and the remaining values can be obtained by simple calculations.

Illustration 20 : On the basis of the study of different branches of a co-operative bank of Ahmedabad city, the following information is obtained. In this bank, out of 20 employees working as security persons, 7 males are married and out of 6 females, 4 are married. 12 peons out of 20 male peons are married and 10 female peons are unmarried. Out of 40 clerks, 25 are females and of them 12 are married whereas 7 male clerks were married and out of 8 managers, all 3 female managers are unmarried and all the male managers are married.

Express this information in a table.

In the given data, there are three attributes (i) 'designation of employee' and (ii) 'gender of employee' and (iii) 'marital status of employee'

A Table showing designation, gender and marital status of the employees working at different branches of a co-operative bank of Ahmedabad city

Designation of employee	Gender								
	Male			Female			Total		
	Unmarried	married	Total	Unmarried	married	Total	Unmarried	married	Total
Security person	7	7	14	2	4	6	9	11	20
Peon	8	12	20	10	0	10	18	12	30
Clerk	8	7	15	13	12	25	21	19	40
Manager	0	5	5	3	0	3	3	5	8
Total	23	31	54	28	16	44	51	47	98

Note : The figures shown in bold letters are the given values in the description of the problem and the remaining values can be obtained by simple calculations. e.g. there are 20 security persons, of which 6 are females. Therefore, number of males = $20 - 6 = 14$. Also, out of 6 security female employees, 4 are married. Therefore, unmarried female security employees = $6 - 4 = 2$. Out of 14 male security employees, 7 are married. Hence, the unmarried male security employees = $14 - 7 = 7$. Total unmarried employees working as security = $7 + 2 = 9$ and total number of married security employees = $7 + 4 = 11$ and the total of these is $9 + 11 = 20$ which is the total number of security persons in the bank. Similarly, the calculations for the remaining employees can be carried out.

2.3 Types of Tabulation and its uses

After deciding the standards of classification the quantitative data or the qualitative data are arranged in the form of classified table. Simple table of numerical data is also called frequency distribution and frequency distribution of bi-variate data is called bi-variate frequency distribution which is not included in our syllabus.

Uses of tabulation :

- (1) It represents the extensive data in simple, organized and precise manner.
- (2) The time, money and labour required for the study of data under consideration is saved as unnecessary information is removed from the data.
- (3) In tabulation, various characteristics to be compared are placed side by side. Hence, the comparison becomes easy.
- (4) Row and/or column totals are found in the table. Therefore, any error occurring due to oversight can be detected and eventually rectified easily.
- (5) The analysis of the data becomes simple and convenient if they are presented in table.

2.3.1 Guiding rules for Tabulation

In order to make the information more meaningful and to derive significant decisions easily from the table, some important guiding rules of tabulation are followed :

- (1) Appropriate title should be given to the table.
- (2) There should be clear and simple captions to the rows and columns.
- (3) If the figures are too large then it should be represented in hundreds, thousands, lakhs or crores.
- (4) The interrelated information should be shown adjacent to each other in such a way that the analysis can be done in a simple way the conclusions can be drawn easily.
- (5) In order to distinguish between the main characteristics of the data in the table, it should be separated by lines.
- (6) There should be a provision for indicating the totals of primary and subsidiary characteristics.
- (7) Source of the data must be mentioned at the end of the table.
- (8) Before preparing the final table, a rough table should be prepared.
- (9) It is better to present as much data as possible in a single table, so that the comparison, computation and analysis becomes easy. But for large volume of data, it is advisable to prepare different tables instead of a single table.

For representing qualitative data, different persons can prepare different tables, but the table which satisfies the objective of the classification is called the best table. In practice, there may be more than 3 attributes to be classified through tabulation, but we have classified the data having maximum 3 attributes as per the limitation of our syllabus.

Illustration 21 : In a university, out of total 50,000 students, 35% are in commerce faculty, 30% are in arts faculty, 20% are in science faculty, 10% are in engineering faculty and remaining 5% are in medical faculty. The ratio of number of boys and girls in commerce faculty is 4:3. In arts faculty, the number of girls is double than that of boys. In science and engineering faculty, there are 60% and 70% boys respectively and in medical faculty, boys and girls are in equal numbers.

Represent the above data in appropriate table.

The given attributes are (i) faculty of students (ii) gender of students.

$$\text{Number of students in commerce faculty} = 50000 \times \frac{35}{100} = 17500$$

$$\text{The number of boys} = \frac{4}{4+3} \times 17500 = 10000$$

$$\text{The number of girls} = \frac{3}{4+3} \times 17500 = 7500$$

$$\text{Number of students in arts faculty} = 50000 \times \frac{30}{100} = 15000$$

The number of girls are double than that of boys.

If number of boys = x then the number of girls = $2x$

$$\text{and } x + 2x = 15000$$

$$\therefore x = 5000$$

$$\therefore \text{Number of boys in arts faculty} = 5000 \text{ and number of girls} = 10000$$

$$\text{Number of students in science faculty} = 50000 \times \frac{20}{100} = 10000$$

$$\text{The number of boys} = 10000 \times \frac{60}{100} = 6000$$

$$\therefore \text{The number of girls} = 10000 - 6000 = 4000$$

$$\text{Number of students in engineering faculty} = 50000 \times \frac{10}{100} = 5000$$

$$\text{The number of boys} = 5000 \times \frac{70}{100} = 3500$$

$$\therefore \text{The number of girls} = 5000 \times \frac{30}{100} = 1500$$

$$\text{Number of students in medical faculty} = 50000 \times \frac{5}{100} = 2500$$

$$\text{The number of boys} = \frac{2500}{2} = 1250$$

$$\text{and the number of girls} = \frac{2500}{2} = 1250$$

A table showing classification of 50,000 students of a university according to their faculty and gender

Faculty	Gender of student		Total
	Boys	Girls	
Commerce	10,000	7500	17500
Arts	5000	10,000	15,000
Science	6000	4000	10,000
Engineering	3500	1500	5000
Medical	1250	1250	2500
Total	25,750	24,250	50,000

Illustration 22: 80 members participated in a picnic organized by a school and the average contribution was ₹ 300 as expenditure. There were 60 students and each of them contributed ₹ 325. Teachers contributed little more for the picnic. There was a support staff of 10 males and contribution was not collected from them. 20% of those participated in picnic were females and 2 of them were teachers. Represent the data in a table.

There are two attributes : (1) Participant (2) Gender.

$$\text{Total contribution for the picnic} = 80 \times 300 = 24000$$

$$\text{Contribution of the students} = 60 \times 325 = 19500$$

$$\text{Contribution of support staff} = 0$$

$$\text{Contribution of the teachers} = 24000 - 19500 = 4500$$

$$\therefore \text{Per head contribution of teacher} = \frac{4500}{10} = 450$$

Table showing category and gender wise contribution of the participants of school picnic

Category of participant	Gender		Total	Per head Contribution (₹)	Total contribution (₹)
	Male	Female			
Students	46	14	60	325	19500
Support staff	10	–	10	–	–
Teacher	8	2	10	450	4500
Total	64	16	80	–	24,000

Activity

For 40 families residing around your residence, collect information for the adults of the family regarding their gender, education and marital status and represent it in a table.

EXERCISE 2.2

- There were 1400 students studying in a commerce college. Among them, 855 were boys and of them, 225 boys were in the second year. In the second year, the number of boys and the number of girls are equal. Among the 550 students of the first year, the proportion of number of boys and girls is 3:2. In the third year, number of boys is three times the number of girls. Represent the above information in a table.
- 1600 employees are working in an office. Among these employees the number of men exceeded the number of women by 15% of the total number of employees. The number of unmarried employees are 800 less than the number of married employees. The number of unmarried women is 195. Represent the above data in a suitable table.
- Prepare a blank table by considering the following characteristics for the candidates who appear for the different jobs at a bank.
 - Designation: Manager, clerk, cashier, peon.
 - Marital status: Married, unmarried.
 - Gender: Male, female.
- Out of total 1850 women working in a factory, 549 were residing in labour area. Out of total married women of labour area, 250 had experience and 93 were inexperienced, the number of experienced and inexperienced women from other area were 87 and 400 respectively. The total number of inexperienced women was 1336 and out of them, 136 were from labour area. Out of total women, 1020 were unmarried. Among them, the number of experienced women from labour area and from other area were 163 and 14 respectively. Present these data in tabular form.
- There were 1250 skilled and 400 unskilled workers in a private company in the year 2011. There were 220 female workers and of them, 140 were unskilled. In the year 2012, the number of skilled workers was 1475 and of them, 1300 were males. Out of 250 unskilled workers, 200 were males. In 2013, there were 1700 skilled and 50 unskilled workers. Out of total workers, 250 were females of them 240 were skilled. In the year 2014, there were 2000 workers and of them, 2% were unskilled. Out of total workers, 300 were females and of them, 10 were unskilled. Present the above data in the form of table.

2.4 Diagrams

Importance and Limitations of diagrams in statistics :

In order to understand and represent huge and complex data into simple and attractive manner, classified data are presented in the form of graphs or diagrams. As per our syllabus, we will discuss only diagrams. The main characteristics of the data presented by graphs are self-explanatory, e.g. for the forecasting of weather conditions on television, various diagrams are used. Similarly, various departments of central and state government, industrial firms, etc. publish their annual activities in various publications by using diagrams.

Importance : The importance of diagram is clear from their uses given below :

- (1) Diagram represents the data in attractive, simple and concise form.
- (2) The characteristics of the data expressed by diagrams are remembered for longer time.
- (3) When two or more sets of data are presented by diagrams on the same scale, the comparative study of those data becomes very simple.
- (4) Diagrams have visual presentation of the data, so one can save time for studying the data.
- (5) The core characteristics of the data conveyed by diagrams are easily understood by the illiterates, less educated or even by children.
- (6) By using diagrams, industrialists and businessmen can make effective advertisement for their products.
- (7) Diagrams are very useful for conveying the messages related to social reforms effectively,
- (8) Pictorial diagrams are easy to understand irrespective of language barriers.

Limitations :

- (1) If the diagram is not drawn accurately then it leads to wrong interpretations.
- (2) Sometimes illusionary effect of diagrams misleads the public opinion.
- (3) There is a loss of accuracy of the data by diagrammatic representation.

2.4.1 Types of Diagrams

There are three main types of diagrams :

- (1) One dimensional diagram
- (2) Two dimensional diagrams
- (3) Pictorial diagrams

2.4.2 One Dimensional Diagram

A diagram drawn by considering only one characteristic of the data is called one dimensional diagram. We shall study the following four diagrams :

- (1) Bar diagram
- (2) Multiple or adjacent bar diagram
- (3) Simple divided bar diagram
- (4) Percentage divided bar diagram

2.4.2.1 Bar Diagram

Bar diagram is constructed by considering only one characteristic of the data. It is used to represent the information on different places, things or time on the diagram. For this purpose, the different places, things or time are taken on x -axis on a graph paper and bars with equal widths at equal distance are drawn, with the heights proportional to the measure of respective places, things or time on y -axis with appropriate scale. The diagram formed by such bars is called bar diagram. In a bar diagram, the logical order of bars should be maintained. When the data based on different places or things are given then it should be arranged in proper order before presenting it on the graph, so that the comparative study can be done easily. When the data based on time are given then they are presented as it is on a graph paper.

Illustration 23 : Information regarding the number of students studying in a college in different faculties in a year is given below. Represent it by a suitable diagram.

Faculty	Arts	Commerce	Science	Engineering	Medical	Law
Number of students	500	1300	900	400	200	300

Only one characteristic 'Faculty' is to be expressed diagrammatically. So, we draw bar diagram. Arranging the data in descending order of number of students of faculty, the following table is obtained.

Faculty	Commerce	Science	Arts	Engineering	Law	Medical
Number of students	1300	900	500	400	300	200

By taking faculty on x -axis and number of students on y -axis, the bar diagram is as under :

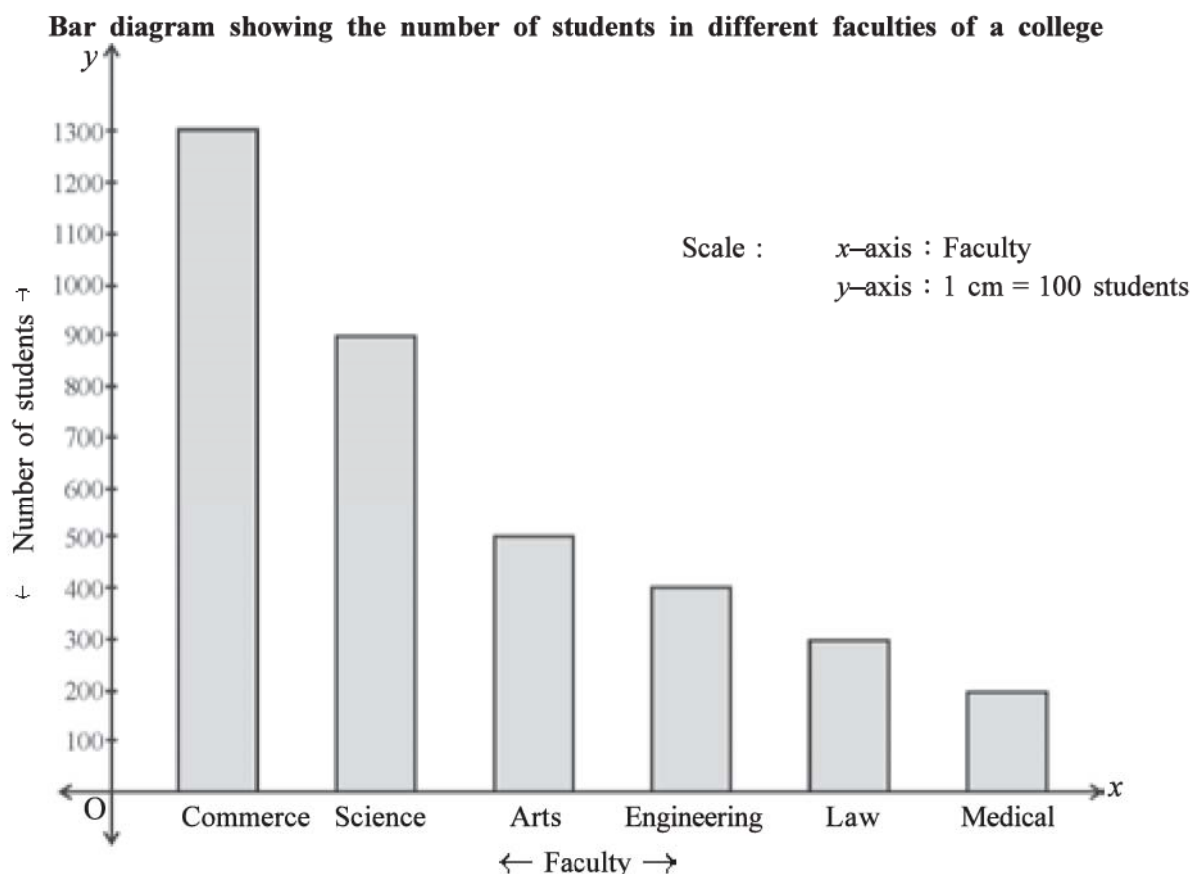
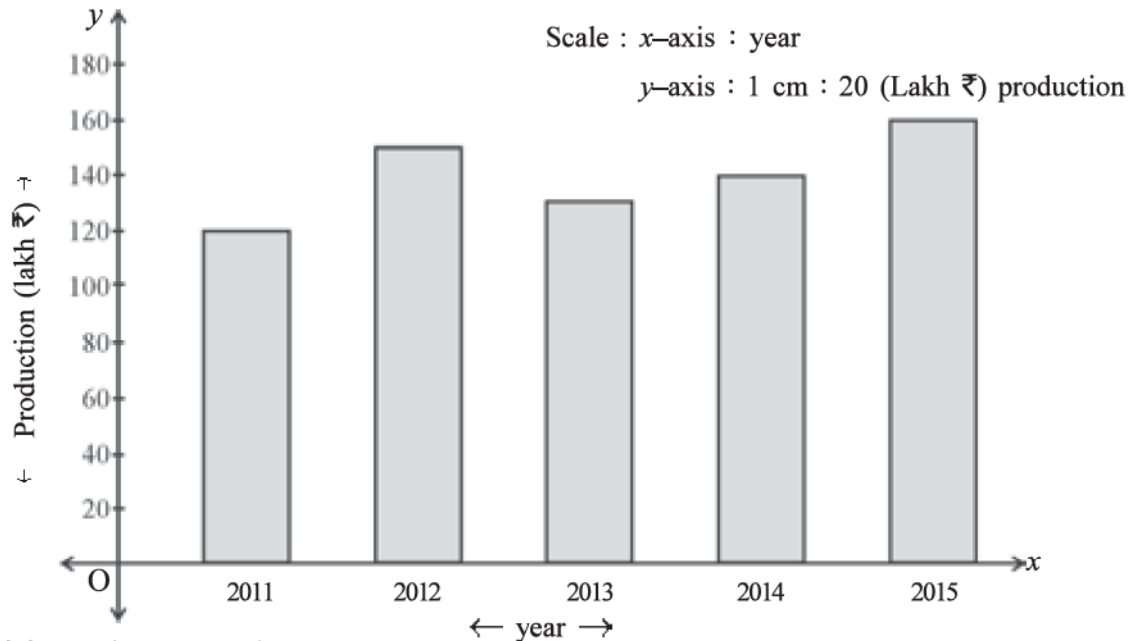


Illustration 24 : The information regarding the production (in lakh ₹) in a factory during five years is given below. Present it in a suitable diagram.

Year	2011	2012	2013	2014	2015
Production (Lakh ₹)	120	150	130	140	160

As the attribute (production) dependent on time is to be shown on the graph, we will draw bar diagram. By taking year on x -axis and production (in lakh ₹) on y -axis, the bar diagram is as under :

Bar diagram showing the production (in lakh ₹) in a factory during five years



2.4.2.2 Multiple Bar Diagram

If the data about different places, things or time are collected on more than one mutually related characteristics then such data can be presented by using multiple bar diagrams or adjacent bar diagrams by placing the related bars close to each other. If the data are related to time then the bars are drawn in the order of time but when the data are not related to time then they are arranged in ascending or descending order by considering any one of the characteristics and then they are presented on graph.

Illustration 25 : The information regarding the number of boys and girls studying in a college in different faculties in a year is given below. Represent it by a suitable diagram.

Faculty	Science	Commerce	Arts	Engineering	Medical	Law
Number of Boys	500	700	200	300	100	100
Number of Girls	400	600	300	100	100	200

The information for two characteristics is given for each faculty. So, we draw adjacent bar diagram. First of all, we arrange the given data in descending order of number of boys in each faculty. The given data can be represented as under :

Faculty	Commerce	Science	Engineering	Arts	Medical	Law
Number of Boys	700	500	300	200	100	100
Number of Girls	600	400	100	300	100	200

The following diagram is drawn taking faculty on x-axis and the number of boys/girls on y-axis.

Multiple bar diagram showing the number of boys and girls studying in different faculties of a college

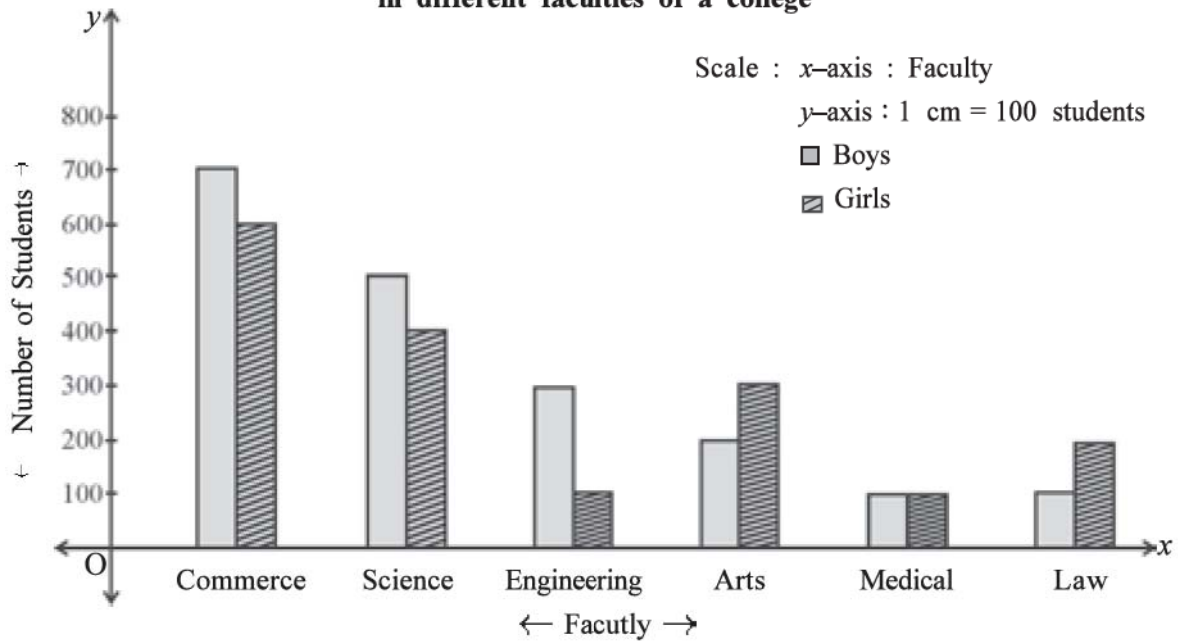
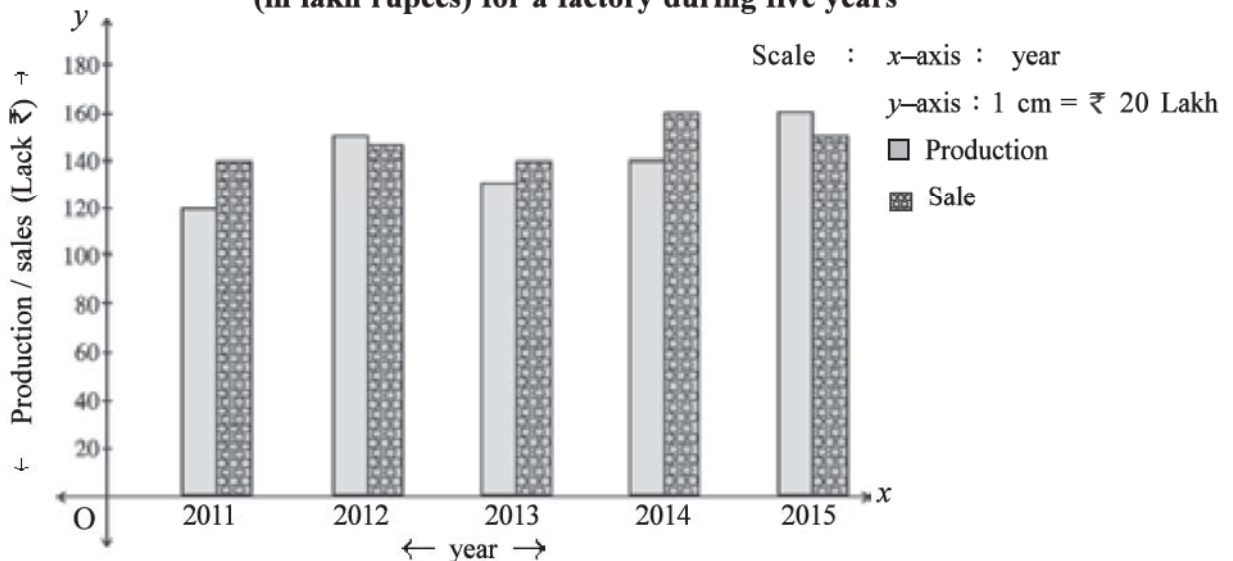


Illustration 26: The information regarding the production (in lakh ₹) and sales (in lakh ₹) for a factory during five years is given below. Present it in a suitable diagram.

Year	2011	2012	2013	2014	2015
Production (Lakh ₹)	120	150	130	140	160
Sales (Lakh ₹)	140	145	140	160	150

Year wise data regarding the production and sales for a factory are given here. So, we draw multiple bar diagram by taking years on x – axis and production/sales (in lakh ₹) on y – axis. The diagram is as follows :

Multiple bar diagram showing the production and sales (in lakh rupees) for a factory during five years



2.4.2.3 Simple divided Bar Diagram

If the data related to different places, things or times consist of several mutually related sub-data on different components are presented on bar diagram by dividing it in accordance with the sub-data, then the diagram obtained is called simple divided bar diagram. For example, if the information regarding the expenses on food, clothing, rent, fuel and miscellaneous for the livelihood of a family are given, then a bar of total expense which is divided into different segments indicated by various signs is called simple divided bar diagram. In short, when the total value is given, simple divided bar diagram is used for presenting its interrelated sub-data.

Illustration 27 : The data on monthly expenses of two different families living in a city are given below. Present it through appropriate diagram.

Monthly expense (₹)	Food	Clothing	Education	Fuel	Rent	Other	Total
Family A	8100	2700	2880	1800	1620	900	18,000
Family B	7000	2000	2000	3000	4000	2000	20,000

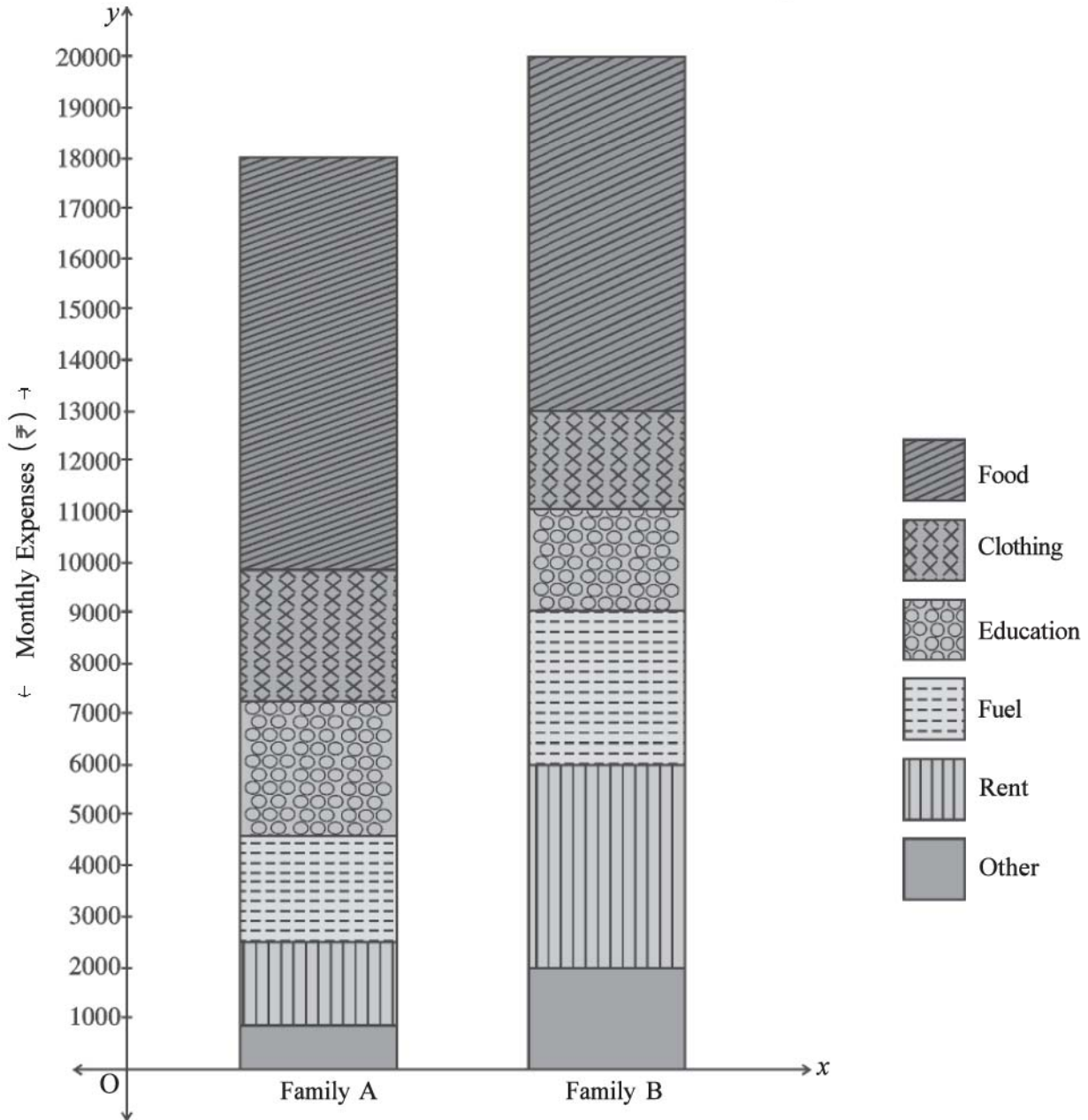
Information regarding the monthly expenses on different particulars of livelihood is given for two families. So, simple divided bar diagrams is to be drawn. By taking families on x – axis and expenses on y – axis, with appropriate scale, simple divided bar diagram is as under :

A bar of total expense is drawn on graph paper and then the different sub-data are separated by the lines according to the calculations shown in the following table.

Particulars	Family A		Family B	
	Expense	Split-up line	Expense	Split-up line
Food	8100	$18000 - 8100 = 9900$	7000	$20000 - 7000 = 13000$
Clothing	2700	$9900 - 2700 = 7200$	2000	$13000 - 2000 = 11000$
Education	2880	$7200 - 2880 = 4320$	2000	$11000 - 2000 = 9000$
Fuel	1800	$4320 - 1800 = 2520$	3000	$9000 - 3000 = 6000$
Rent	1620	$2520 - 1620 = 900$	4000	$6000 - 4000 = 2000$
Other	900	–	2000	–
Total	18,000	–	20,000	–

Simple divided bar diagram showing monthly expenses of two families

Scale : x-axis : Family
y-axis : 1 cm = ₹ 1000



2.4.2.4 Percentage divided Bar Diagram

In simple divided bar diagram, sub-data are presented in an attractive manner. But the mutually related sub-data cannot be effectively compared. To overcome this limitation, percentage divided bar diagram is used. The total value is considered as 100% and percentages of sub-data calculated on the basis of it are presented in divided bar diagram.

Illustration 28 : To compare the data on expenses of the families given in illustration 27, prepare an appropriate diagram.

Here we have to compare the sub-data on the monthly expenses of two families so we draw percentage divided bar diagram.

Particulars	Family A			Family B		
	Expense	Percentage	Split-up	Expense	Percentage	Split-up
Food	8100	45	$100 - 45 = 55$	7000	35	$100 - 35 = 65$
Clothing	2700	15	$55 - 15 = 40$	2000	10	$65 - 10 = 55$
Education	2880	16	$40 - 16 = 24$	2000	10	$55 - 10 = 45$
Fuel	1800	10	$24 - 10 = 14$	3000	15	$45 - 15 = 30$
Rent	1620	9	$14 - 9 = 5$	4000	20	$30 - 20 = 10$
Other	900	5	–	2000	10	–
Total	18,000	100	–	20,000	100	–

Percentage divided bar diagram showing monthly expenses of two families

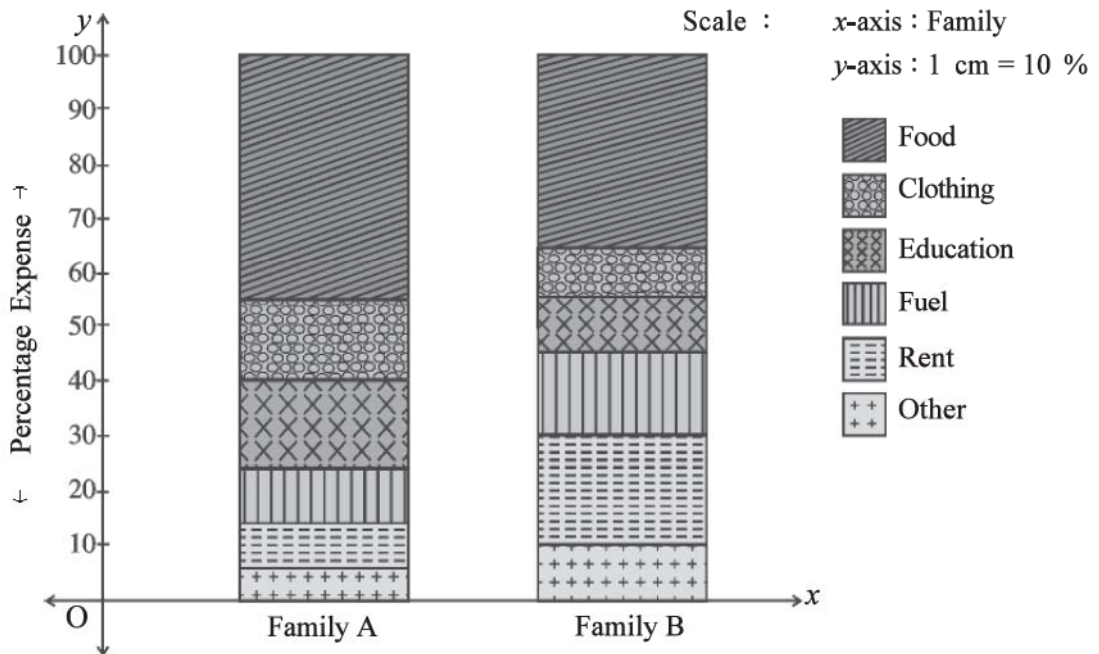


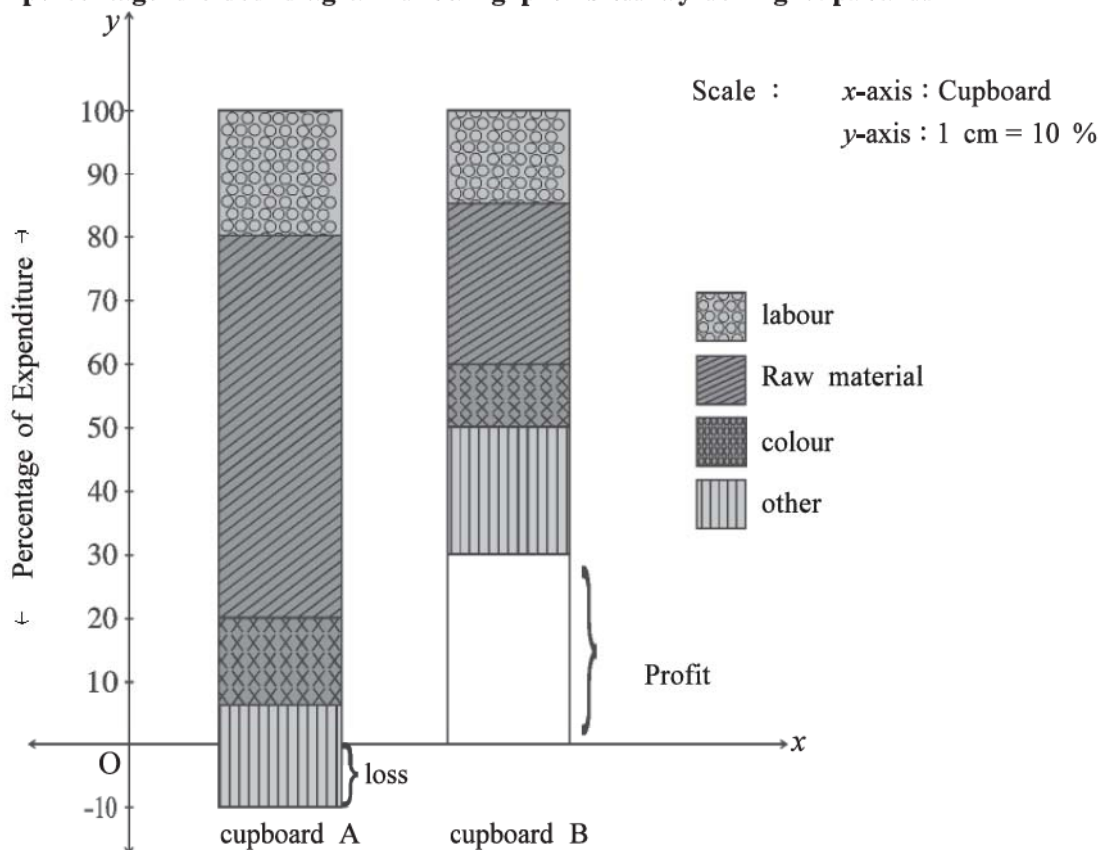
Illustration 29 : A factory produces and sells two types of cupboards, A and B. Cupboard A is not much attractive but is sturdy and hence it has more usage in industrial units. Whereas, cupboard B is good-looking, hence it is used for household purpose. The selling price of cupboard A is ₹ 5000 and that of cupboard B is ₹ 8000 per piece. The particulars about the making expenses are as follows. Draw a percentage divided bar diagram showing particulars of expenses of cupboards and the profit/loss by their sales.

Particulars of expenses	Cupboard A	Cupboard B
Labour	1000	1200
Raw material	3000	2000
Colours	700	800
Other	800	1600
Total	5500	5600

As we have to draw a diagram showing profit/loss by selling these cupboards, considering selling price of cupboard as 100 %, the percentage of the expenses are determined and are shown in percentage divided bar diagram.

Particulars	Cupboard A			Cupboard B		
	Expense	Percentage	Split-up line	Expense	Percentage	Split-up line
Labour	1000	20	$100 - 20 = 80$	1200	15	$100 - 15 = 85$
Raw material	3000	60	$80 - 60 = 20$	2000	25	$85 - 25 = 60$
Colours	700	14	$20 - 14 = 6$	800	10	$60 - 10 = 50$
Other	800	16	$6 - 16 = -10$	1600	20	$50 - 20 = 30$
Production cost	5500	110	–	5600	70	–
Selling price	5000	100	–	8000	100	–
Profit/loss	-500	-10	–	2400	30	–

A percentage divided diagram showing profit/loss by selling cupboards



2.4.3 Two Dimensional Diagram

One dimensional diagram is used to represent a single characteristic of the data. In such diagrams, either height or breadth is taken into consideration. But when the volume of the data is large, then in order to present it on a diagram, both length and breadth are taken into consideration. Thus total value is shown as an area in the diagram for which the diagrams like square, rectangle, circle or pie (sectorial) diagram are included. We shall study only circle diagrams and sectorial or pie diagrams.

2.4.3.1 Circle Diagram

When the volume of data regarding two or more places, things or time is large, then circular diagrams are used to present the data. The volume of the data is represented by area of a circle.

$$\text{Total volume of data} = \text{Area of circle} = \pi r^2$$

Where $\pi = 3.14$ or $\frac{22}{7}$, $r = \text{radius}$

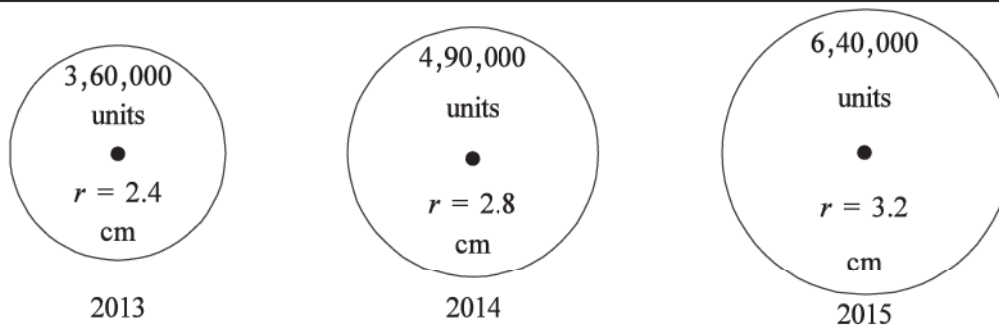
Area of a circle is proportional to the square of its radius. Therefore, in circle diagram, square roots of the volumes of different data are taken as the radii of the circle. The radii thus obtained are arranged in ascending or descending order and are drawn with centers on same line at equal distance from each other. If the data are given with respect to time then circles are drawn in order of time only. If the radius of circle is too large then it is divided by a constant and if it is too small then it is multiplied by a constant and then circle diagram is drawn.

Illustration 30 : The data on the production of an industrial unit during three years is as under. Present it by using circle diagram.

Year	Production (in units)
2013	3,60,000
2014	4,90,000
2015	6,40,000

As the given data are numerically large, radius is taken as the square root of each value which is further divided by 250.

Year	Production (Units)	Square root	Radius = square root/250
2013	3,60,000	600	2.4
2014	4,90,000	700	2.8
2015	6,40,000	800	3.2



2.4.3.2 Pie-diagram

If the data related to different places, things or times consist of several mutually related sub-data on different components are numerically large, then instead of divided bar diagram, pie diagram is used. The total volume of data is represented by a circle of suitable radius and this circle is then divided into sectors to present the sub-data. In the diagram, total volume of the data is taken as 360° and the volumes of sub-data are expressed in terms of measures of angle and are presented on circle with respective circular sectors.

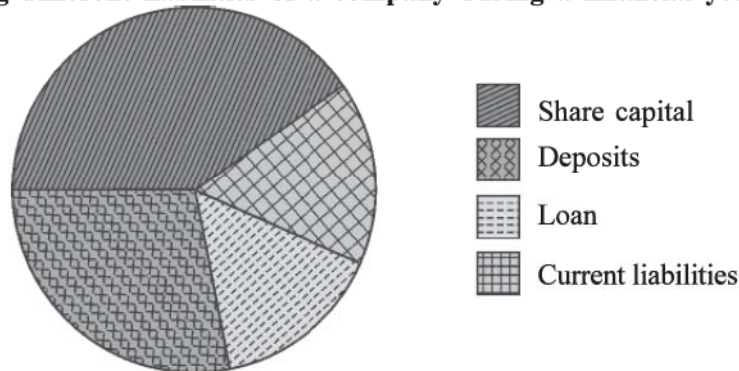
Illustration 31 : In a ledger of a company, the liabilities in the final balance sheet are as under.
Present the data by a pie diagram.

Particulars	Share capital	Deposits	Loan	Current liabilities	Total
Liabilities (₹)	12,00,000	8,00,000	4,00,000	4,80,000	28,80,000

Measure of angle in degrees for the liabilities of different particulars can be calculated as under :

Particulars	Liabilities (₹)	Angle (degree)
Share capital	12,00,000	$\frac{1200000}{2880000} \times 360^\circ = 150^\circ$
Deposits	8,00,000	$\frac{800000}{2880000} \times 360^\circ = 100^\circ$
Loan	4,00,000	$\frac{400000}{2880000} \times 360^\circ = 50^\circ$
Current liabilities	4,80,000	$\frac{480000}{2880000} \times 360^\circ = 60^\circ$
Total	28,80,000	360°

Pie diagram showing different liabilities of a company during a financial year



Activity

For your family, collect information regarding monthly expenses on food, education, fuel and miscellaneous and represent it through a pie diagram.

Illustration 32 : Information regarding the annual expense on different particulars of two middle class families is given below. Draw a pie diagram for the data.

Particulars	Expenditure (₹)	
	Family A	Family B
Food	40000	50,000
Clothing	10000	20,000
Rent	25000	30,000
Education	10000	32,000
Miscellaneous	5000	28,000
Total	90,000	1,60,000

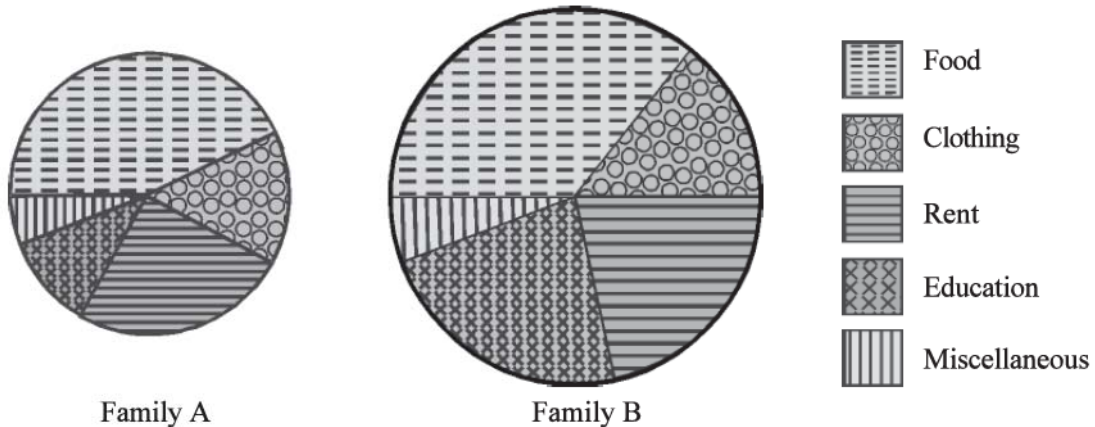
We shall draw two circles for two families with radii proportional to the square root of the total expenditure of two families. By taking total expenditure as 360° the measures of angles of sub-data are obtained in terms of degrees.

$$\text{Radius for family A} = \frac{\sqrt{90000}}{100} = \frac{300}{100} = 3 \text{ cm}$$

$$\text{Radius for family B} = \frac{\sqrt{160000}}{100} = \frac{400}{100} = 4 \text{ cm}$$

Particulars	Expenditure (₹)			
	Family A		Family B	
	Expense	Degree	Expense	Degree
Food	40,000	$\frac{40000}{90000} \times 360^\circ = 160^\circ$	50,000	$\frac{50000}{160000} \times 360^\circ = 112.5^\circ$
Clothing	10,000	= 40°	20,000	= 45°
Rent	25,000	= 100°	30,000	= 67.5°
Education	10,000	= 40°	32,000	= 72°
Miscellaneous	5000	= 20°	28,000	= 63°
Total	90,000	= 360°	1,60,000	= 360°

Pie diagram showing annual expenditure of two families



Activity

Represent the above information in percentage divided bar diagram.

2.4.4 Pictogram

A diagram in which the data are represented by appropriate pictures is called pictogram. The pictures are selected with reference to the data. For example, if the data related to the population are given then it can be shown by symbols of human. The pictures are drawn in proportion to the amount of the given data. Pictogram draws quick attention of the viewer. If the data are presented by pictograms then it can be easily understood by the less educated people and by the children. Also, pictogram has no barrier of language. The main disadvantage of this method is that it has limited usage in statistical analysis of data.

Illustration 33 : Information regarding the population of five cities is given below. Present it by a pictogram.

City	A	B	C	D	E
Population	20,000	40,000	60,000	80,000	90,000


Here, one figure of human  = 20,000

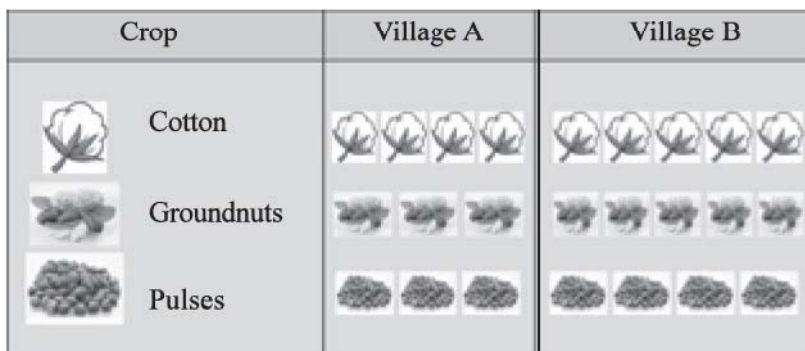


Illustration 34 : The data regarding area under cultivation of various crops at farms of two villages is given below. Present it through pictogram.

Crop	Area under Cultivation (hector)	
	Village A	Village B
Cotton	400	500
Groundnut	300	500
Pulses	300	400

By considering the symbol of the respective crop, the pictogram drawn is as under :

Pictogram showing area under cultivation in two villages



Here, one figure of crop = 100 acres of area under cultivation.

EXERCISE 2.3

1. Following is the data on the number of employees working in various Government departments. Present it with a suitable diagram.

Department	Road transport	Railway	Income tax	Finance	Planning
Number of employees	4000	6000	3000	2500	1500

2. The data on the profit of a company is as under. Present it by an appropriate diagram :

Year	2010	2011	2012	2013	2014	2015
Profit (crore ₹)	10	5	-2	4	8	6

3. Changes in the price of share of 5 companies at Bombay Stock Exchange during an interval of 15 days are as follows. Present them by a suitable diagram :

Company	A	B	C	D	E
Price of share (₹)	40	20	100	80	30
Price of same share after 15 days (₹)	60	30	150	60	10

4. Information regarding the birth rate and death rate of 5 countries is as under. Present it by a suitable diagram.

Country	U.S.A.	Japan	Bharat	Germany	U.K.
Birth rate	16.5	20.8	34.2	16.4	15.2
Death rate	10.2	12.2	20.4	10.3	12.0

5. Information regarding the age of persons living in two different regions is as under. Present it by using an appropriate diagram :

Age	Less than 15 (Child)	15 to 35 (Young)	35 to 60 (Adult)	More than 60 (Old)	Total
Region A	480	360	240	120	1200
Region B	350	250	200	200	1000

6. Draw the percentage divided bar diagram for the data given in example 5.
7. A car production company has produced the following number of cars during three years. Represent it through a circle diagram.

Year	2012	2013	2014
Production of Cars	25,600	1,02,400	1,60,000

8. The following data represent the percentage sales of copies of daily newspapers. Represent it by pie diagram.

News paper	P	Q	R	S	Total
Percentage of selling	25	23	24	28	100

9. Represent the following information by a pictogram :

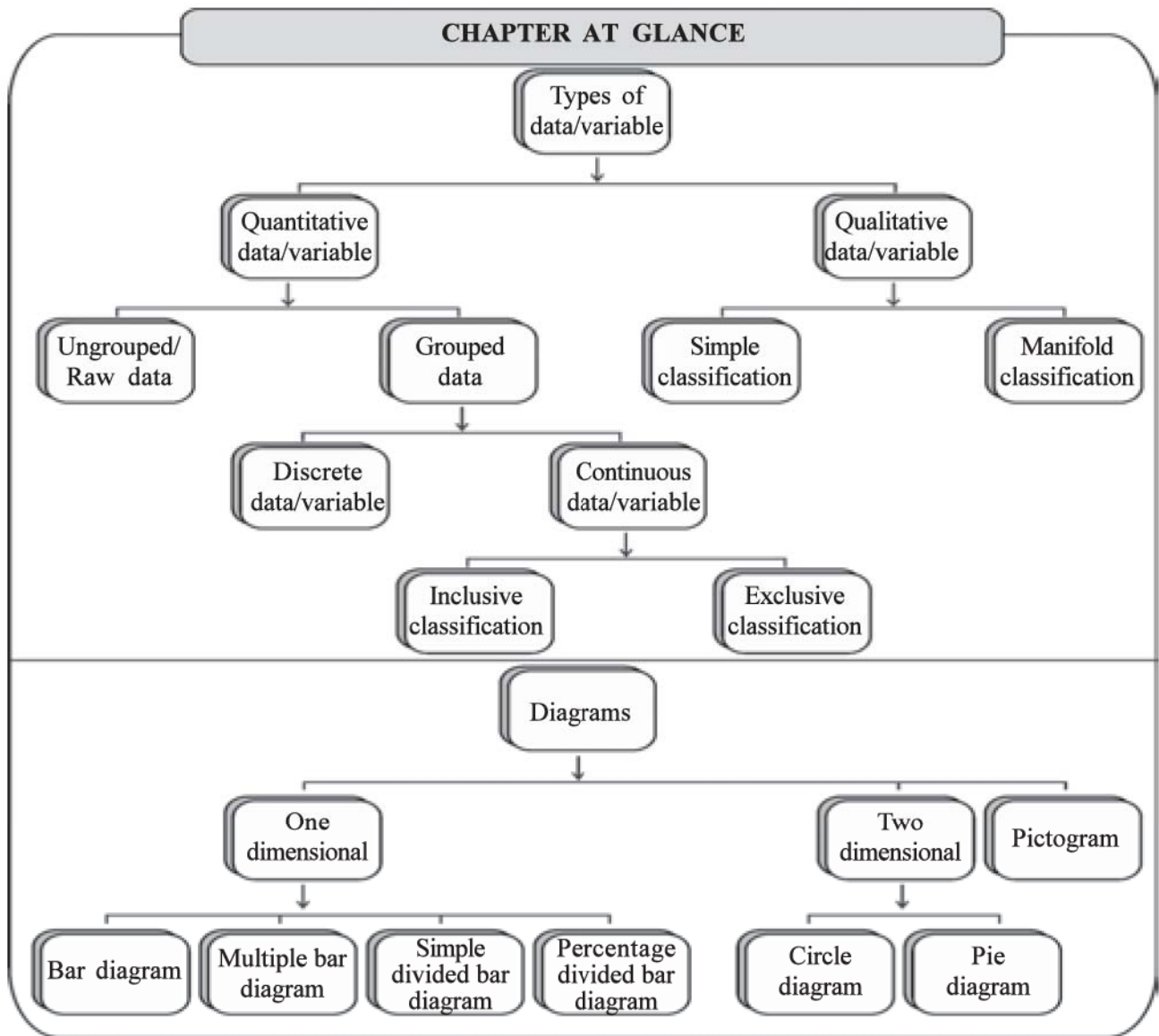
Year	2010	2011	2012	2013	2014
Production of mangoes (kg.)	1,00,000	1,50,000	2,50,000	1,50,000	75,000

10. Details of production of electric bulb by two well-known companies are as under. Represent them by a pictogram.

Year	Production of electric bulbs (lakh units)	
	Company A	Company B
2012	50	100
2013	100	150
2014	175	200
2015	200	200

Summary

- A variable assuming definite value between two specified limits is called discrete variable.
- A variable assuming any value between two specified limits is called continuous variable.
- A process of arranging raw data into short and systemic manner is called classification.
- A numeric value indicating the repetition of the value of observation is called frequency for that observation.
- A table showing the different values of discrete variable with the respective frequencies is called discrete frequency distribution.
- When the values of observations of raw data are classified in different classes then the table is called continuous frequency distribution.
- When the range of discrete variable is large then the data are expressed by inclusive continuous frequency distribution.
- Sum of the frequencies upto the value of observation or class is called cumulative frequency and its distribution is called cumulative frequency distribution.
- While preparing cumulative frequency distribution, inclusive classes should be converted into exclusive classes.
- In exclusive classes, class limits and class boundary points are identical.
- When the range of raw data is too large then a frequency distribution with unequal class length is to be obtained.
- Diagrams are used to present classified data in simple and attractive form.
- In one dimensional diagram, the diagram is drawn by considering only one characteristic of the data.
- To express one characteristic of a data, bar diagram is used, whereas for representing more than one characteristics, multiple bar diagram is used.
- To represent various sub-data of the classified data, divided bar diagram is used and to compare them, percentage divided bar diagram is used.
- When numerically large data is given then pie diagram is used to represent it. It is also used to compare large data.
- Pictorial representation of data is called pictogram.



Formulae :

- (1) Range of data $R = \text{Maximum value} - \text{Minimum value}$
- (2) Class length $C = \frac{\text{Range}}{\text{No. of classes}}$
- (3) Lower boundary point = $\frac{\text{lower limit of that class} + \text{upper limit of previous class}}{2}$
 Upper boundary point = $\frac{\text{upper limit of that class} + \text{lower limit of succeeding class}}{2}$
- (4) Mid point or mid value of a class = $\frac{\text{value of upper limit} + \text{value of lower limit}}{2}$
- (5) Lower boundary point = $\text{mid value} - \frac{1}{2} (\text{Class length})$
 Upper boundary point = $\text{mid value} + \frac{1}{2} (\text{Class length})$

EXERCISE 2

Section A

Choose the correct option for the following multiple choice questions :

- Which of the following variables is discrete ?
 - Height of a person
 - Weight of a commodity
 - Area of a ground
 - Number of children per family.
- Which of the following variables is continuous ?
 - Number of errors per page of a book
 - Number of cars produced
 - Number of accidents on road
 - Monthly income of a person
- Name the method of classification of raw data related to daily demand of a product.
 - Classification of attribute data
 - Classification of numeric data
 - Raw distribution
 - Manifold classification
- Name the type of classification of the data related to the occupation and education of a person living in a certain region.
 - Tabulation
 - Classification of numeric data
 - Raw distribution
 - Discrete frequency distribution
- In continuous frequency distribution, what is the class length of a class ?
 - Average of two successive lower boundary points.
 - Average of class limits
 - Difference between upper boundary point and lower boundary point of that class.
 - Average of upper boundary point and lower boundary point of the class.
- Range of an ungrouped data is 55 and it is divided into 6 classes. Then what is the class length ?
 - 10
 - 9
 - 9.17
 - 10.17
- Inclusive classes for a distribution are 10-19.5, 20-29.5, 30-39.5. What are the exclusive class limits for the second class ?
 - 19.5 - 29.5
 - 19.75 - 29.75
 - 20 - 30
 - 19 - 29
- A discrete variable has values 0, 1, 2, 3, 4 with the respective frequency 2, 4, 6, 8, 14. What is the value of 'more than' type cumulative frequency when the value of variable is 2 ?
 - 28
 - 12
 - 34
 - 6
- A continuous distribution has classes 0 – 9, 10 – 19, 20 – 29, 30 – 39 with the respective frequencies 10, 20, 40, 10. What is the 'less than' type cumulative frequency for the boundary point 29.5 ?
 - 30
 - 50
 - 70
 - 80

10. For a continuous variable, classes are 1 - 1.95, 2 - 2.95, 3 - 3.95, 4 - 4.95, 5 - 5.95 then what is the lower boundary point of the second class ?
 (a) 1.995 (b) 2 (c) 2.975 (d) 1.975
11. Which of the following statement is/are true ?
 Statement 1: A method of representing the large and complex data in simple and attractive manner is called diagram.
 Statement 2: Self-explanatory representation of main characteristics of the data is called diagram.
 Statement 3: Representation of comparative study of data is called diagram.
 (a) Only statement 1 is true. (b) Only statements 1 and 2 are true.
 (c) Statements 1,2 and 3 are true. (d) All three statements are false.
12. The class intervals for a continuous variable are 0 - 99, 100 - 199, 200 - 299, 300 - 399, 400 - 499. What is the mid value of the second class ?
 (a) 149.5 (b) 150 (c) 199.5 (d) 99.5
13. What do we call a table that shows designation, gender and marital status of employees of a company ?
 (a) Simple classification (b) Classification of numeric data
 (c) Manifold classification (d) Simple table
14. Which of the following diagrams is used to represent sub-data of classified information ?
 (a) Bar diagram (b) Divided bar diagram
 (c) Multiple bar diagram (d) Pictogram
15. Which of the following diagrams is used for comparing the sub-data of the classified data ?
 (a) Pictogram (b) Pie chart
 (c) Bar diagram (d) Divided bar diagram

Section B

Give answer in one sentence for the following question :

1. Define discrete variable.
2. Define continuous variable.
3. What is classification ?
4. State the types of classification.
5. Define the frequency of an observation.
6. State the method to determine number of classes on the basis of range of data and class length.
7. When should one form a frequency distribution with unequal class lengths ?
8. Define cumulative frequency.
9. Define 'less than' type cumulative frequency distribution for discrete data.
10. Define 'more than' type cumulative frequency distribution for continuous data.
11. Write a formula for finding mid value of a class.
12. Define tabulation.
13. Define manifold classification.
14. What is the characteristic of the best table to represent qualitative data ?

15. What is the main disadvantage of classification of data ?
16. In statistical study, what is the main objective of a diagram ?
17. State the types of diagrams.
18. For which type of data, multiple bar diagram is drawn ?
19. When do we draw divided bar diagram ?
20. State the main objective of percentage divided bar diagram.

Section C

Answer the following questions :

1. Define quantitative and qualitative data.
2. Define discrete frequency distribution with illustration.
3. Define continuous frequency distribution with illustration.
4. Explain the definition of inclusive continuous frequency distribution.
5. Explain the definition of exclusive continuous frequency distribution.
6. Write formulae for obtaining class boundary points from inclusive class limits.
7. Find mid values of each class of the following frequency distribution.

Class	0 - 9	10 - 24	25 - 49	50 - 74	75 - 100
Frequency	10	20	30	20	10

8. For the frequency distribution given in the above problem, find the class length of each class.
9. Prepare 'less than' type cumulative frequency distribution from the following.

Observation	10	20	30	40	50
Frequency	10	30	30	20	10

10. Demand of a certain item is classified as good, moderate and weak. On the basis of a study for entire year, it is known that the demand was moderate during 22 weeks, whereas the demand was weak during 18 weeks. Present this information in a table.

11. Complete the following table :

	Attribute A			Attribute B			Total		
Year	Sub-data-1	Sub-data-2	Total	Sub-data-1	Sub-data-2	Total	Sub-data-1	Sub-data-2	Total
2014	200		300	100		200			
2015		400		150	300		300		

12. Differentiate between inclusive and exclusive continuous frequency distribution.
13. State the limitations of diagram.
14. What are one dimensional diagrams? State their names.
15. Explain two dimensional diagrams in brief.

16. Represent the following data through a bar diagram :

(1)

Year	2011	2012	2013	2014	2015
Production (in crore ₹)	3.5	4.2	5.8	7.4	10.2

(2)

Stream	Arts	Commerce	Science	Engineering	Other
Number of students	5900	10,200	6000	4500	8000

Section D

Solve the following :

1. What is the necessity of classification in statistical study ?
2. Explain the classification of numeric data with an appropriate illustration.
3. Explain the classification of qualitative data with a suitable illustration.
4. Write a short note on 'cumulative frequency distribution'.
5. Discuss the points for constructing continuous frequency distribution.
6. State the guiding rules for the construction of a table.
7. State the uses of tabulation.
8. Obtain the original frequency distribution from the following data.

Mid value	250	350	450	550	650	750	Total
Frequency	20	80	80	40	60	20	300

9. Out of 40 persons working in an office, 60% are females and remaining 40 % are males. 50 % of males are married, where as the ratio of married and unmarried females is 5:3. Present this information in a table.
10. Information regarding the monthly income of 100 workers is given below. Obtain original frequency distribution from it.

Less than Monthly income	2400	2900	3400	3900	4400	4900	5400	5900	6400
Number of worker	0	3	12	30	55	78	88	95	100

11. Marks of 200 students in an examination are as under. Obtain the original frequency distribution.

Marks	10 - 100	20 - 100	30 - 100	40 - 100	50 - 100	60 - 100	70 - 100	80 - 100	90 - 100
No. of Students	200	180	140	90	55	30	8	2	1

12. From the data given below, obtain original frequency distribution.

Mid value	12.5	17.5	22.5	27.5	32.5	37.5	42.5	47.5
Frequency	12	18	16	22	14	10	6	2

13. There are 1000 buses used for the public transport in Ahmedabad city. Of them, 350 are used as BRTS and remaining as AMTS. Out of total 400 air conditioned buses, 250 were used as BRTS. Present this information in a suitable table.

14. Out of 1500 students of a college, 900 were boys and of them, 250 were in science stream. 250 girls were in commerce stream. Present these data in an appropriate table.
15. Explain the importance of diagrams in statistical study.
16. Write a short note on one dimensional diagrams.
17. Write a short note on two dimensional diagrams.
18. Explain pictogram with an illustration.
19. The agricultural production index numbers for two different states are as under. Present them by using suitable diagram.

Year	2011	2012	2013	2014	2015
State A	139	147	152	162	170
State B	110	115	125	140	150

20. Area (in sq.mt.) of 5 different regions is as under. Draw a pie diagram.

Region	A	B	C	D	E
Area	5	8	29	44	71

21. Production of a commodity in three different factories is as under. Present it through suitable diagram.

Factory	P	Q	R
Production (lakh ₹)	256	576	1024

Section E

Solve the following :

1. Number of mangoes received from different trees of mangoes in a farm during a season of 30 days is as under. Prepare a frequency distribution by taking class length 5.
 94 96 100 104 122 107 108 106 119 120
 98 123 102 125 95 125 115 104 114 109
 128 112 103 92 114 101 113 118 124 118
2. The data regarding the earnings (₹) of 40 rickshaw drivers during a certain day are as follows. Prepare a frequency distribution having one class as 220 - 239 and class length 20.
 285 215 200 225 255 250 235 242 298 312
 328 294 266 335 330 270 315 275 245 265
 210 235 275 305 332 355 307 230 348 350
 310 290 264 228 236 336 356 322 215 345
3. Information on monthly water consumption (in units) of 50 residents of a region is as under. By taking one of the classes as 25 - 30, prepare exclusive continuous frequency distribution.
 24 34 41 55 45 25 40 38 40 44
 28 35 40 48 35 44 27 57 42 30
 28 26 42 49 47 33 52 52 28 34
 36 30 44 33 31 30 39 25 24 47
 28 36 32 57 25 29 35 44 50 56

4. The data obtained by inquiring price of an item at 50 different shops are as under. Prepare a frequency distribution having the last class 85-90.

82 75 73 70 84 79 79 77 80 66
 70 70 72 62 64 80 85 64 75 65
 66 75 71 82 69 70 72 80 66 70
 79 69 80 63 66 75 68 78 86 66
 85 66 69 85 70 60 70 75 79 86

5. Obtain 'less than' type and 'more than' type cumulative frequency distribution from the following frequency distribution, .

Class	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	Total
Frequency	3	8	10	5	15	8	1	50

6. The following data refer to the daily absence of workers in a factory during 30 days. Prepare an appropriate frequency distribution and hence obtain 'less than' type cumulative frequency distribution.

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

7. There were 850 students studying in higher standards of a school. The number of students in standard 10, 11 and 12 were in the proportion 8:5:4. In standard 10, the number of boys is 30% of the number of students in the school. In standard 11, the numbers of boys and girls are equal. In standard 12, the number of boys is three times the number of girls. Present the above data in a tabular form.
8. In the year 2013, there were 1200 students studying in a school and of them, 400 were girls. 50 girls were not residing in hostel. In all 600 boys were residing in hostel. In the year 2014, there is an increase of 20% in the number of boys and the number of girls increased by 30%. During this year, 260 boys and 100 girls were not residing in hostel. In the year 2015, 140 boys and 100 girls were newly admitted in the school and all of them resided with the hostel students. Present above data in a tabular form.
9. Present the following data in an appropriate tabular form. A bank receives 2000 applications as a response to the job advertisement. Of the total applicants, 50% were graduates, 40% were post graduates and remaining 10 % have professional degree. Among the graduates, 60% were males and of them, 25% were married. 40% female graduates were married. Among the post graduates, 60% were males and 40% of them were married. Among post graduate females, 50% were married. 30% of the females had professional degree and of them, 60% were married. The number of married and unmarried males having professional degree was equal.
10. The following table represents the number of workers of a factory according to their gender, residence and year :

Year	Local			Non-local			Total		
	Men	Women	Total	Men	Women	total	Men	Women	Total
2010	1200	300	1500	300	200	500	1500	500	2000
2015	2000	600	2600	300	100	400	2300	700	3000

Answer the following questions using the above table :

- (1) What is the percentage increase in the total number of workers during the period of five years?
- (2) Find the percentage decline in the number of non-local workers in the year 2015.
- (3) Find the percentage increase in the number of men and women during the period of 5 years.

11. A mobile phone manufacturing company produces and sells two types of mobile phones. The particulars about it are given in the following table. Present it by a suitable diagram :

Particulars	Mobile A	Mobile B
Raw material	5000	6000
Assembly expense	3000	3000
Other expense	4000	4500
Total expense	12,000	13,500
Selling price	13,000	15,000

12. Information regarding the average monthly expenses (in ₹) of two families is as under. Present it through a pie diagram.

Particulars	Family A	Family B
Food	20,000	16,000
Fuel	5000	4000
Transportation	10,000	8800
House rent	15,000	18,000
Other	22,000	18,000

Section F

Solve the following :

1. A sample of 25 lenses is selected from a day's production of a company manufacturing eye lenses. The thicknesses (in millimeter) of these selected lenses are as under. Distribute these data into five classes of equal length.

1.518 1.509 1.527 1.505 1.520 1.511 1.518 1.522 1.528 1.528 1.520
 1.520 1.514 1.508 1.525 1.506 1.519 1.523 1.521 1.517 1.514 1.515
 1.516 1.521 1.507

If the company decides that the lenses having thicknesses less than 1.510 and more than 1.525 are considered as defective then what percent of lenses in the sample are defective ?

2. The data related to variations in the price of a share for 30 days in a share market are as under. Prepare an exclusive continuous classification having class limits of one of the classes as 18.5 - 20.5.

10.50	14.70	17.20	15.20	14.50	19.20	15.80	19.30
18.40	20.50	18.70	14.90	18.50	16.90	10.50	12.50
13.60	12.50	18.50	18.60	14.00	16.20	13.30	13.30
18.60	17.60	20.20	14.50	20.80	14.90		

On the basis of this frequency distribution, answer the following questions.

- (1) What is mid value of the 4th class ?
 - (2) Find the number of days during which the price of share is less than ₹ 16.50.
 - (3) Find the number of days during which the price of share is at least ₹ 19.50.
3. Owner of a factory has decided to produce 50 mixers used as household equipment, but the daily production of mixers changes due to variation in the number of workers. A variation in production of mixers with respect to a pre-decided number of production (100 units) during 40 days is recorded as under. Prepare an exclusive continuous frequency distribution having class length 6 and mid value of one of the classes as 3. Also prepare 'less than' and 'more than' cumulative frequency distributions.

7	6	12	16	12	18	11	-5	10	3	10	7	8	
14	-10	16	-7	20	9	12	-2	0	5	-4	23	6	
-3	4	4	3	4	2	0	22	1	5	-1	5	19	6

4. The data regarding the height (in cm.) of 30 students of a school are as under. Prepare an inclusive continuous frequency distribution of 6 classes and hence prepare 'less than' and 'more than' cumulative frequency distributions.

141	145	152	150	150	159	148	163	162	151	155	148
145	162	161	152	168	153	149	148	162	158	157	160
153	149	154	165	141	149						

On the basis of it, answer the following questions :

- (1) If participation in the N.C.C. activities requires a minimum height of 160 cm. then how many students are eligible to participate ?
 - (2) Find the number of students having height from 153 cm. to 163 cm.
 - (3) Find maximum height of one third of the students having minimum height.
5. The students of a university were classified according to faculty and gender. 60% of total 40,000 students were boys. The number of girls in engineering faculty was three times the number of girls in commerce faculty. 15% and 10% of the total number of university students were boys and girls respectively who belonged to medical faculty. 20% of the total number of students in the university belonged to faculty of science and among these students, the number of girls were one-seventh of the number of boys. 7% and 17% of the total number of students of arts faculty were boys and girls respectively. 3.75% of the total number of students of the university belonged to the commerce faculty and the proportion of boys and girls among them was 3:7.

Present the above data in an appropriate table.



Prof. C. R. Rao
(1920)

Prof. C.R. Rao is an Indian born, naturalized American, mathematician and statistician. He is currently professor emeritus at Penn State University and Research Professor at the University of Buffalo. Rao has been honoured by numerous colloquia, honorary degrees and festschrifts and was awarded the US National Medal of Science in 2002. The American Statistical Association has described him as “a living legend whose work has influenced not just statistics, but has had far reaching implications for fields as varied as economics, genetics, anthropology, geology, national planning, demography, biometry and medicine.” The Times of India listed Rao as one of the top 10 Indian scientists of all time.



Statistics may rightly be called as a science of averages.

– Sir A. L. Bowley

3

Measures of Central Tendency

Contents :

3.1 Meaning

3.2 Characteristics of good measure of central tendency

3.3 Mean

3.3.1 Meaning, Advantages and Disadvantage

3.3.2 Combined mean and Weighted mean

3.3.3 Geometric mean

Meaning, Advantages and Disadvantages

3.4 Measures of positional averages: Median, Quartiles, Deciles, Percentiles

3.4.1 Meaning, Advantages and Disadvantages

3.5 Mode

3.5.1 Meaning, Advantages and Disadvantages

3.5.2 Graphical Method

3.6 Comparative study of mean, median and mode

3.1 Meaning

The large volume of statistical data can be organized by classification and tabulation. This shows certain characteristics of the given data. The diagrams and graphs drawn for the data show its trends and patterns. It helps in visual interpretation and comparison of the data. We need more concise and numerical representation of the data for further statistical analysis. Let us understand this with an illustration.

Suppose a person is planning his monthly budget. The expense on every commodity is a variable which changes according to the quantity consumed and the market price. Suppose he wants to decide the amount to be spared for milk. He has the values of milk expenses of past 10 months. He wants a representative value from these data to provide for the expenses on milk in his budget.

A representative value of more than one set of data can be used for their comparison and further for taking future decisions. We will illustrate this by the following situation.

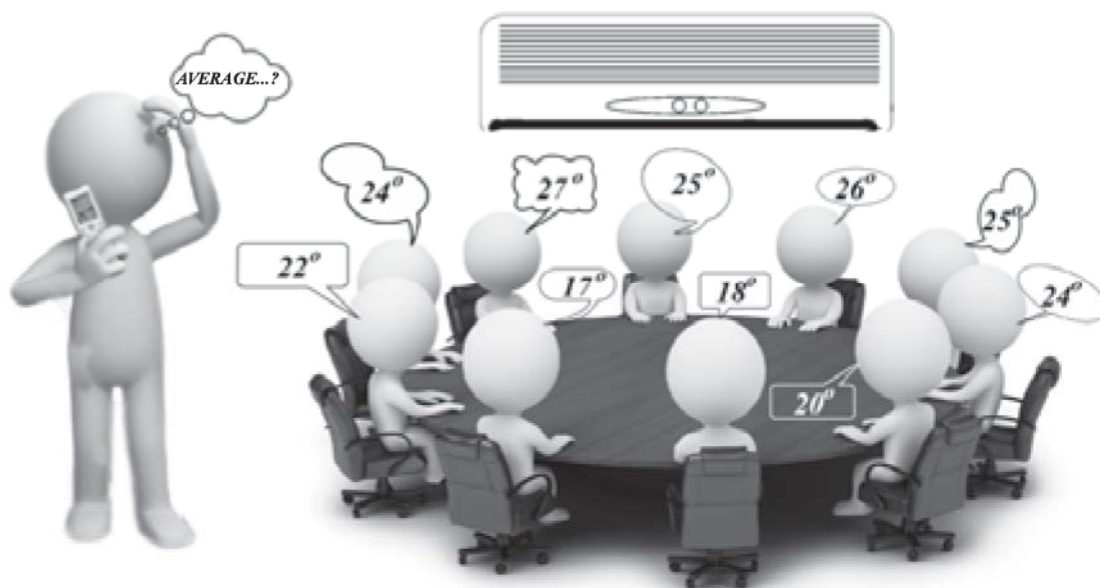
Suppose a company wants to compare the sales of two products produced by it. The sales vary from day to day. The company has data about sales of last 50 days. The pattern displayed by the sales of two products can be compared from the frequency distributions obtained from these data. But the company needs some exact measures to describe the data of sales of their two products for further conclusion and comparison.

In most of the graphs drawn for various frequency distributions, one can observe a common pattern that the values of the variable are concentrated around a certain central value. This characteristic of data is called as **central tendency** and the central value around which the values of the variable are concentrated is called as **measure of central tendency** or **average** value. The average value thus can be taken as a representative for the whole set of data. It is used for further analysis, interpretation and comparison.

Thus an average value

- Presents the data in a concise form
- Shows special characteristics of the data
- Helps in comparison between two or more sets of data

Different measures of central tendency can be obtained for the collected data. The choice of average depends upon the type of data, the purpose of average value and its further application.



3.2 Characteristics of Good Measure of Central Tendency

An average with following characteristics can be called as an ideal average :

- (1) It should be well defined and rigid.
- (2) It should be easy to understand and calculate.
- (3) It should be based on all the observations of the data.
- (4) It should be suitable for further algebraic operations.
- (5) It should be a stable measure. It means that values of averages found for different samples of same size from the same population should be almost same.
- (6) It should not be unduly affected by a few very large or very small observations.

We will discuss the following measures of central tendency which are widely used in data analysis.

- (1) Mean (2) Median and other positional averages (3) Mode.

3.3 Arithmetic Mean or Mean

This is one of the most commonly used averages.

3.3.1 Meaning

Arithmetic mean is defined as the value obtained by sum of all observations divided by the total number of observations.

The arithmetic mean of variable x is denoted by \bar{x} .

Calculation of mean :

For raw or ungrouped data :

Suppose x_1, x_2, \dots, x_n are the n observations in the data, then Arithmetic mean is $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

$$= \frac{\sum x_i}{n}$$

Where $\sum x_i = x_1 + x_2 + \dots + x_n =$ Sum of observations x_1, x_2, \dots, x_n

and $n =$ number of observations

Note : For the sake of simplicity while solving examples, we will not write the suffix i . Thus we will take x instead of x_i , d instead of d_i and f instead of f_i .

Illustration 1 : The following data show the number of scooters repaired daily at a garage. Find the mean number of scooters repaired per day.

7, 13, 4, 8, 6, 9, 10, 4

Here $n = 8$

$$\begin{aligned} \text{Mean } \bar{x} &= \frac{\sum x}{n} = \frac{x_1 + x_2 + \dots + x_8}{8} \\ &= \frac{7 + 13 + 4 + 8 + 6 + 9 + 10 + 4}{8} \\ &= \frac{61}{8} \\ &= 7.625 \\ &\approx 7.63 \end{aligned}$$

Thus, the mean number of scooters repaired per day at the garage is 7.63 scooters.

Short cut method :

If the values of observations are very large, the calculation can be simplified by using assumed mean A . A is some constant value, preferably around the centre of all the observations. Assumed mean A is subtracted from observations x_1, x_2, \dots, x_n and the deviations are denoted by d_1, d_2, \dots, d_n .

$$d_1 = x_1 - A, d_2 = x_2 - A, \dots, d_n = x_n - A$$

The mean \bar{x} is obtained as follows :

$$\bar{x} = A + \frac{\sum d_i}{n}$$

Where A = Assumed mean

$$\sum d_i = d_1 + d_2 + \dots + d_n$$

and n = number of observations.

Note : The choice of value of A does not change the value of mean.

Illustration 2 : The rainfall (in mm.) at 10 different places of a district was recorded as :

126, 110, 91, 115, 112, 80, 101, 93, 97, 113

Find mean rainfall.

As the observations have large values, we will calculate mean by short cut method. We will select assumed mean $A = 100$. The observations x and deviations $d = x - A$ are shown in the following table :

Rainfall (mm) x	126	110	91	115	112	80	101	93	97	113	Total
$d = x - A, A=100$	26	10	-9	15	12	-20	1	-7	-3	13	38

Here, $n = 10$

$$\begin{aligned} \text{Mean } \bar{x} &= A + \frac{\sum d}{n} \\ &= 100 + \frac{38}{10} \\ &= 100 + 3.8 \\ &= 103.8 \end{aligned}$$

Thus, the mean rainfall is 103.8 mm.

Illustration 3: The mean weight of a group of 20 persons was found to be 55 kg. Later, it was discovered that one of them reported her weight as 45 kg which was actually 54 kg. Find the correct mean of their weights.

Here, $\bar{x} = 55$ and $n = 20$

$$\text{Mean } \bar{x} = \frac{\sum x}{n} = 55$$

$$\therefore \frac{\sum x}{20} = 55$$

$$\therefore \sum x = 55 \times 20 = 1100$$

Thus the sum of observations is 1100 which includes a wrong value 45 instead of the correct value 54.

To find the correct sum of observations, we subtract the wrong observation and add the correct observation.

$$\begin{aligned} \therefore \text{correct } \Sigma x &= 1100 - 45 + 54 \\ &= 1109 \end{aligned}$$

$$\begin{aligned} \therefore \text{correct mean} &= \frac{\text{correct } \Sigma x}{n} \\ &= \frac{1109}{20} \\ &= 55.45 \end{aligned}$$

Thus, the correct mean weight is 55.45 kg.

For grouped data :

For discrete frequency distribution :

Suppose x_1, x_2, \dots, x_k are the observations in the given data with frequencies f_1, f_2, \dots, f_k respectively.

Here, n = total number of observations

$$= f_1 + f_2 + \dots + f_k = \Sigma f_i$$

The frequency of x_1 is f_1 means observation x_1 is repeated f_1 times. The sum of all x_1 observations will be $f_1 \times x_1$ or $f_1 x_1$. Similarly, sum of all x_2 observations will be $f_2 x_2$ and so on.

$$\begin{aligned} \text{Mean } \bar{x} &= \frac{\text{sum of all observations}}{\text{total number of observations}} \\ &= \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{n} \\ &= \frac{\Sigma f_i x_i}{n} \end{aligned}$$

Where $\Sigma f_i x_i = f_1 x_1 + f_2 x_2 + \dots + f_k x_k$

Illustration 4 : The following table shows the number of children per family in a certain area. Find the mean number of children per family.

No. of children	0	1	2	3	4	5
No. of families	4	8	23	8	6	3

Here we have $k = 6$ values for the variable x .

The calculations to find mean are shown in the following table :

No. of children x	No. of families f	fx
0	4	0
1	8	8
2	23	46
3	8	24
4	6	24
5	3	15
Total	$n = 52$	117

$$\begin{aligned}\text{Mean } \bar{x} &= \frac{\sum fx}{n} \\ &= \frac{117}{52} \\ &= 2.25\end{aligned}$$

Thus, mean for number of children per family is 2.25 children.

Short cut method :

As done earlier for the raw data, an assumed mean A can be suitably chosen and the deviations of values x_1, x_2, \dots, x_k can be taken from A . Further, if all these deviations have a common factor c , we can further simplify the calculations by dividing all the deviations by c .

Thus we will have the values $d_1 = \frac{x_1 - A}{c}, d_2 = \frac{x_2 - A}{c}, \dots, d_k = \frac{x_k - A}{c}$

Now, the formula for mean is written as follows :

$$\text{Mean } \bar{x} = A + \frac{\sum f_i d_i}{n} \times c$$

Where $\sum f_i d_i = f_1 d_1 + f_2 d_2 + \dots + f_k d_k$

and $n =$ total number of observations

$$= f_1 + f_2 + \dots + f_k = \sum f_i$$

Note : The choice of values of A and c does not change the value of mean.

Illustration 5 : The time (in minutes) taken by a bus to travel between two towns on different days is shown in the following table.

Time (min.)	110	113	120	122	126
No. of days	7	17	11	10	5

Find the mean travel time.

Here the values of observations are large. We will take assumed mean $A = 120$.

The deviations are $110 - 120 = -10$, $113 - 120 = -7$, $120 - 120 = 0$, $122 - 120 = 2$, $126 - 120 = 6$.

There is no common factor in them other than 1. Hence we will take $c = 1$

Thus we will have $d = \frac{x - A}{c} = \frac{x - 120}{1} = x - 120$

The calculations for mean will be as follows :

Time (minute) x	No. of days f	$d = x - A$ $A = 120$	fd
110	7	-10	-70
113	17	-7	-119
120	11	0	0
122	10	2	20
126	5	6	30
Total	$n = 50$		- 139

$$\begin{aligned}\text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\ &= 120 + \frac{(-139)}{50} \times 1 \\ &= 120 - 2.78 \\ &= 117.22\end{aligned}$$

Thus, the mean travel time for the bus is 117.22 minutes

Illustration 6: The price of an item changes from shop to shop. The following data are available.

Find the mean price.

Price (₹)	206	212	218	220	224	230
No. of shops	5	8	9	14	3	1

As the observations are large, we will calculate the mean using short cut method in which, we will select $A = 220$. The deviations of all observations from A will be $-14, -8, -2, 0, 4, 10$. These deviations have the highest common factor $c = 2$.

$$\text{Hence we will take } d = \frac{x-A}{c} = \frac{x-220}{2}$$

Calculations for mean :

Price (₹) x	No. of shops f	$d = \frac{x-A}{c}$ $A = 220, c = 2$	fd
206	5	-7	-35
212	8	-4	-32
218	9	-1	-9
220	14	0	0
224	3	2	6
230	1	5	5
Total	$n = 40$		-65

$$\begin{aligned}\text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\ &= 220 + \frac{(-65)}{40} \times 2 \\ &= 220 + \frac{(-130)}{40} \\ &= 220 - 3.25 \\ &= 216.75\end{aligned}$$

Thus the mean price of this item is ₹ 216.75.

For continuous frequency distribution :

When we transform the data into a continuous frequency distribution, each frequency shows the number of observations in that class. Thus we do not know each observation of that class. The mid-value of that class is taken as a representative for all the values in that class.

For example, let us consider a class 0-5 with frequency 7. Exact values of these 7 observations are not known. Hence the mid-value 2.5 is assumed for all 7 observations of this class which actually can be any number from 0 to 5.

Considering the mid-value of each class as x , the mean can be calculated by the same method as the one used for the discrete frequency distribution described earlier.

Thus the mean will be calculated as follows :

$$\text{Mean } \bar{x} = A + \frac{\sum f_i d_i}{n} \times c$$

Where A = Assumed mean

c = common factor among $x_i - A$

$$d_i = \frac{x_i - A}{c}$$

f_i = frequency of the class with mid-value x_i

$$\sum f_i d_i = f_1 d_1 + f_2 d_2 + \dots + f_k d_k$$

n = total number of observations

$$= f_1 + f_2 + \dots + f_k = \sum f_i$$

Illustration 7: The following data represent monthly income (in ₹) of workers in a factory. Find their mean income.

Income (₹)	2000	3000	4000	5000	6000	7000	8000
	- 3000	- 4000	- 5000	- 6000	- 7000	- 8000	- 9000
No. of workers	2	3	7	15	25	16	12

We first find the mid-value of each class.

$$\text{mid-value} = \frac{\text{upper limit of the class} + \text{lower limit of the class}}{2}$$

These mid-values will be 2500, 3500, 4500, 5500, 6500, 7500, 8500. We will take $A = 5500$.

The deviations $x - A$ will be -3000, -2000, -1000, 0, 1000, 2000, 3000 respectively.

As these deviations have the highest common factor $c = 1000$, we will take $d = \frac{x - A}{c} = \frac{x - 5500}{1000}$.

Calculations for mean are shown in the following table :

Income (₹)	No. of workers f	Mid-value x	$d = \frac{x-A}{c}$ $A = 5500, c = 1000$	fd
2000 - 3000	2	2500	-3	-6
3000 - 4000	3	3500	-2	-6
4000 - 5000	7	4500	-1	-7
5000 - 6000	15	5500	0	0
6000 - 7000	25	6500	1	25
7000 - 8000	16	7500	2	32
8000 - 9000	12	8500	3	36
Total	$n = 80$			74

$$\begin{aligned}
 \text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\
 &= 5500 + \frac{74}{80} \times 1000 \\
 &= 5500 + \frac{74000}{80} \\
 &= 5500 + 925 \\
 &= 6425
 \end{aligned}$$

Thus, the mean income of these workers is ₹ 6425.

Illustration 8 : The data about weight (in grams) of mangoes from a tree are as follows. Moreover, the minimum weight for these mangoes is 410 grams. Find the mean weight of mangoes.

Weight of mango (gram)	Below 420	Below 430	Below 440	Below 450	Below 460	Below 470
No. of mangoes	14	34	76	130	165	180

This table shows cumulative frequencies of 'less than' type. We can find the frequencies from these cumulative frequencies by subtracting the frequencies of successive classes. The lower limit of first class is given as 410 grams.

Thus the frequency distribution will be obtained as follows :

Weight of mango (gram)	410 - 420	420 - 430	430 - 440	440 - 450	450 - 460	460 - 470
No. of mangoes	14	20	42	54	35	15

The mid-values of the classes are 415, 425, ..., 465.

If we choose $A = 435$, the values obtained as deviations $-20, -10, \dots, 30$ will have the highest common factor $c = 10$. We will take $d = \frac{x-A}{c} = \frac{x-435}{10}$.

Calculation for mean is as follows :

Weight of mango (gram)	No. of mangoes f	Mid-value x	$d = \frac{x-A}{c}$ $A = 435, c = 10$	fd
410 - 420	14	415	-2	-28
420 - 430	20	425	-1	-20
430 - 440	42	435	0	0
440 - 450	54	445	1	54
450 - 460	35	455	2	70
460 - 470	15	465	3	45
Total	$n = 180$			121

$$\begin{aligned}
 \text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\
 &= 435 + \frac{121}{180} \times 10 \\
 &= 435 + \frac{1210}{180} \\
 &= 435 + 6.7222 \\
 &= 441.7222 \\
 &\approx 441.72
 \end{aligned}$$

Thus, the mean weight of mangoes is 441.72 gm.

Activity

Take $A = 415$ and suitable value of c in the above example and calculate mean.

Now take $A = 440$ and find the deviations. What is the common factor c ? Again calculate mean with these values of A and c .

Observe that all the answers for mean are same.

Illustration 9 : The distribution of annual sales tax of different companies in a zone is given below.

Find the mean sales tax of these companies :

Sales tax (thousand ₹)	0 - 10	10 - 20	20 - 30	30 - 50	50 - 70
No. of companies	3	14	32	40	21

The class lengths of this distribution are not same. The mid values of classes are 5, 15, 25, 40, 60. Taking $A = 25$ the deviations are -20, -10, 0, 15, 35 which will have the highest common factor $c = 5$

. Thus we will take $d = \frac{x-A}{c} = \frac{x-25}{5}$.

The calculation for mean in the following table :

Sale Tax (thousand ₹)	No. of companies f	Mid-value x	$d = \frac{x-A}{c}$ $A = 25, c = 5$	fd
0 - 10	3	5	-4	-12
10 - 20	14	15	-2	-28
20 - 30	32	25	0	0
30 - 50	40	40	3	120
50 - 70	21	60	7	147
	$n = 110$			227

$$\begin{aligned}
 \text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\
 &= 25 + \frac{227}{110} \times 5 \\
 &= 25 + \frac{1135}{110} \\
 &= 25 + 10.3182 \\
 &= 35.3182 \\
 &\approx 35.32
 \end{aligned}$$

Thus, mean sales tax paid by these companies is ₹ 35.32 thousand.

Advantages and disadvantages of mean :

Advantages :

Mean is the most popular measure of central tendency because of its following advantages.

- (1) It is rigidly defined. It has a fixed mathematical formula.
- (2) It is easy to understand and calculate.
- (3) It is based on all observations.
- (4) It is suitable for further algebraic operations.
- (5) It is comparatively more stable measure. It means that means of all samples of same size from the same population have comparatively less variation.
- (6) All the observations are given equal importance in the calculation of mean.

Disadvantages :

The following disadvantages should be also taken into consideration before using mean as a measure of central tendency.

- (1) It is unduly affected by too large or too small values.
- (2) It cannot be calculated for data having classes with open ends.
- (3) Its exact value cannot be found graphically or by inspection.
- (4) If few observations are missing, exact value of mean cannot be found.
- (5) Mean is not a good representative value for the data which are not evenly spread around their average.
- (6) Using mean as an average is not appropriate if observations have varying importance.

Some important results about mean :

(1) The sum of deviations of all observations from mean is always zero. Symbolically, the deviations can be shown as $x_i - \bar{x}$ and hence $\sum (x_i - \bar{x}) = 0$

For example, consider 4 values 1, 7, 5, 3.

Their mean $\bar{x} = \frac{\sum x}{n} = \frac{1+7+5+3}{4} = \frac{16}{4} = 4$

The deviations from mean are shown in the following table :

x	1	7	5	3	Total
$(x - \bar{x})$	-3	3	1	-1	$\sum (x - \bar{x}) = 0$

The sum of deviations from any other value cannot be zero.

Activity

Find the deviations of above observations from 5. What is their sum? Is it zero ?

Now, take any value of your choice other than mean and verify that the sum of deviations from this value is not zero.

(2) If each observation from x_1, x_2, \dots, x_n is multiplied by a non zero constant b and another constant a is added to it, we get a new set of observations. We will denote these values by y_1, y_2, \dots, y_n , where $y_1 = bx_1 + a, y_2 = bx_2 + a, \dots, y_n = bx_n + a$.

The mean of y_1, y_2, \dots, y_n will be $\bar{y} = \frac{\sum y}{n} = \frac{y_1 + y_2 + \dots + y_n}{n}$

If we know mean \bar{x} , we can use the formula $\bar{y} = b\bar{x} + a$ to find \bar{y} which is the mean of y .

Activity

Note down the ages of 10 neighbours around your house and find their mean \bar{x} . What will be their ages after two years ? Find the mean \bar{y} of new set of numbers y you calculated. Here, age after two years for each person is $y = x + 2$. Now observe that $\bar{y} = \bar{x} + 2$

EXERCISE 3.1

- The weekly growths (in cm.) in saplings grown in a nursery are:
1.0, 3.2, 1.4, 1.9, 2.4, 1.6, 1.4, 2.1, 1.3, 1.5.
Find the mean growth.
- The mean age of 4 participants in the team of a relay race was calculated to be 24 years. Later it was found that one of the participant's age was actually 27 years, which was wrongly recorded as 25 years. If there is rule wherein it is not possible to participate in this race if the mean age is more than 25 years, can they participate in this race even after the correction of age?
- The following table gives the diameters (in mm.) of different screws selected from a large consignment. Find the mean diameter.

Diameter of screw (mm.)	30	35	40	45	50	55
No. of screws	4	10	15	8	5	3

4. The marks in a test for a group of students are as follows. Find the mean marks of these students.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
No. of students	3	5	12	16	11	5	4

5. The following information is available on the talk time (in min.) noted for 70 calls of a certain mobile phone user. Find the mean talk time.

Talk time (min.)	Less than 4	Less than 8	Less than 12	Less than 16	Less than 20
No. of calls	20	42	57	65	70

6. The information of profits (in lakh ₹) of 50 firms is given below. Find mean profit.

Profit (Lakh ₹)	0–7	7–14	14–21	21–28	28–35
No. of firms	4	9	18	12	7

7. The distribution of demand of an item on different days is as follows. Find the mean demand.

Demand (units)	5–14	15–24	25–34	35–49	50–64	65–79
No. of days	4	17	19	22	18	10

*

3.3.2 Combined Mean and Weighted Mean

Combined Mean :

If we know the means of two or more groups of observations, we can find mean of the combined group. Such a value is called as combined mean. It is denoted by \bar{x}_c .

Suppose $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ are means of k groups having n_1, n_2, \dots, n_k observations respectively.

The formula for combined mean is as follows :

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \dots + n_k\bar{x}_k}{n_1 + n_2 + \dots + n_k}$$

Illustration 10 : A factory owner knows that the mean monthly production from January to March is 350 items, from April to August it is 254 items and from September to December it is 315 items. Find the mean monthly production for that year.

Here $n_1 = 3$ months, $n_2 = 5$ months, $n_3 = 4$ months,

$$\bar{x}_1 = 350, \quad \bar{x}_2 = 254, \quad \bar{x}_3 = 315$$

$$\begin{aligned} \text{Combined mean } \bar{x}_c &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3} \\ &= \frac{3(350) + 5(254) + 4(315)}{3 + 5 + 4} \\ &= \frac{1050 + 1270 + 1260}{12} \\ &= \frac{3580}{12} \\ &= 298.3333 \\ &\approx 298.33 \end{aligned}$$

Thus, mean monthly production of the factory over the year is 298.33 items.

Illustration 11: The ratio of women and men employees in an office is 1:2. If the means of ages of women and men are 34 years and 37 years respectively, find the mean age of all the employees of the office.

Suppose the no. of women = a . Then the no. of men = $2a$ as the ratio is 1:2.

Taking $n_1 = a$ and $n_2 = 2a$, $\bar{x}_1 = 34$ and $\bar{x}_2 = 37$

$$\begin{aligned} \text{Combined mean } \bar{x}_c &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \\ &= \frac{a(34) + 2a(37)}{a + 2a} \\ &= \frac{34a + 74a}{3a} \\ &= \frac{108a}{3a} \\ &= 36 \end{aligned}$$

Thus, mean age of all the employees in the office is 36 years.

Illustration 12: A team has to score with a run rate of 8.25 in 20 overs to win the match. The run rate at the end of 12 overs is 7.25 What should be the least run rate to be maintained in the remaining overs to win the match ?

$$\begin{aligned} \text{We know that run rate} &= \frac{\text{Total number of runs}}{\text{Total number of overs}} \\ &= \text{mean of runs} \end{aligned}$$

Thus we will take run rate as mean of runs per over.

Here $n_1 = 12$, $n_2 = 8$

$$\begin{aligned} \bar{x}_c &= \text{mean number of runs from all 20 overs} \\ &= 8.25 \text{ runs} \end{aligned}$$

$\bar{x}_1 =$ mean runs in the first 12 overs = 7.25 runs.

$$\text{Combined mean } \bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$\therefore 8.25 = \frac{12(7.25) + 8\bar{x}_2}{12 + 8}$$

$$\therefore 8.25 = \frac{87 + 8\bar{x}_2}{20}$$

$$\therefore 8.25 \times 20 = 87 + 8\bar{x}_2$$

$$\therefore 165 = 87 + 8\bar{x}_2$$

$$\therefore 8\bar{x}_2 = 165 - 87 = 78$$

$$\begin{aligned} \therefore \bar{x}_2 &= \frac{78}{8} \\ &= 9.75 \end{aligned}$$

Thus, minimum run rate for this team in the last 8 overs should be 9.75 runs to win the match.

Weighted Mean :

We said that using arithmetic mean is not appropriate if the importance of all observations is not same. A special mean called as weighted mean can be found in such cases. Weighted mean is denoted by \bar{x}_w . Each observation is assigned a numerical value called weight according to its importance. The most important value is given maximum weight.

Suppose w_1, w_2, \dots, w_n are the weights assigned to observations x_1, x_2, \dots, x_n respectively.

The formula for weighted mean is given as follows :

$$\text{Weighted mean } \bar{x}_w = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n}$$
$$= \frac{\sum w_i x_i}{\sum w_i}$$

$$\text{Here } \sum w_i x_i = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$$

$$\text{And } \sum w_i = w_1 + w_2 + \dots + w_n$$

= sum of weights

Illustration 13: A student gets 35 marks in theory paper, 15 marks in practical examination and 5 marks in oral examination of a subject. The school gives weights 4, 2 and 1 respectively to these types of examinations. Find the weighted mean of marks for this student.

$$\text{Here } x_1 = 35, \quad x_2 = 15, \quad x_3 = 5 \text{ and}$$
$$w_1 = 4, \quad w_2 = 2, \quad w_3 = 1$$

$$\text{Weighted mean } \bar{x}_w = \frac{w_1x_1 + w_2x_2 + w_3x_3}{w_1 + w_2 + w_3}$$
$$= \frac{4(35) + 2(15) + 1(5)}{4 + 2 + 1}$$
$$= \frac{140 + 30 + 5}{7}$$
$$= \frac{175}{7}$$
$$= 25$$

Thus, the weighted mean of marks of this student is 25.

EXERCISE 3.2

1. The mean daily wage paid to 75 skilled workers of a factory was ₹ 280 whereas the mean daily wage paid to 125 unskilled workers was ₹ 150. Find the mean wage of all the workers.
2. Find the weighted mean of the percentage change in prices from the following data :

Food item	Rice	Wheat	Tea	Sugar	Pulses
Weight	7	10	5	8	2
Percentage change in price	134	125	115	97	120

3. 2 officers, 10 clerks and 3 peons contributed for a staff picnic. The contribution collected per person is shown in the following table.

Officer	Clerk	Peon
₹ 1000	₹ 500	₹ 200

Find the mean contribution per person using weighted mean.

4. The mean marks of a student in 7 theory papers is 62. What should be the mean marks in 3 practical examinations so that his mean marks for the entire examination is 68 ?
(The marks of each theory paper and practical examination are same.)

*

3.3.3 Geometric Mean :

Suppose we are studying a variable that changes over time. If we want to find average rate of change in the variable, the application of arithmetic mean is not appropriate.

Let us understand this with an example.

Consider an item with a price of ₹ 200. If the price increases after one month by 50% and by 25% in the next month, the prices in the successive months will be $200 \times \frac{150}{100} = 300$ and $300 \times \frac{125}{100} = 375$ ₹ respectively.

If we find the average of percentage prices for the two months using arithmetic mean, it will be $\frac{150+125}{2} = 137.5$

If We take this average for finding price of that item 2 months later, it will be $200 \times \frac{137.5}{100} \times \frac{137.5}{100} = 378.18$ ₹ which is not ₹ 375 as calculated above.

Another average called as Geometric Mean is more appropriate here.

A geometric mean is defined as the n th root of the product of n positive observations and it is denoted by G .

Thus, for n observations x_1, x_2, \dots, x_n , $G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$

Let us find geometric mean in the above example. The two numbers 150 and 125 which show the percentage prices will have $G = \sqrt{150 \times 125} = 136.93$

Now applying this average, the price after two months will be $200 \times \frac{136.93}{100} \times \frac{136.93}{100} = 375$ ₹ which is same as the value calculated earlier.

Note: If the given variable increases by $p\%$ then increased percentage value is written as $(100 + p)$ whereas, if the value has decreases by $p\%$ then decreased percentage value is written as $(100 - p)$. For example, if the price decreases 20% in a certain month, the percentage price for the next month is taken as $(100 - 20) = 80$.

Note : For the given observations x_1, x_2, \dots, x_n , arithmetic mean is always equal to or more than their geometric mean. That is $\bar{x} \geq G$

$\bar{x} = G$ only if all the observations have same value.

Illustration 14 : The population of an area increased by 15 %, 18 %, 13 %, 20 % respectively for four years. Find the average rate increase in the population.

Since the values of increase in population are given in percentages, we will use the geometric mean for the average value.

Considering the percentage increase in the population, we will get the observations

$$\begin{aligned}x_1 &= 100 + 15 = 115 & x_2 &= 100 + 18 = 118, \\x_3 &= 100 + 13 = 113 & x_4 &= 100 + 20 = 120\end{aligned}$$

$$\begin{aligned}G &= \sqrt[4]{x_1 \times x_2 \times x_3 \times x_4} \\&= \sqrt[4]{115 \times 118 \times 113 \times 120} \\&= \sqrt[4]{184009200} \\&= \sqrt{13564.9991} \\&= 116.4689 \\&\approx 116.47\end{aligned}$$

Thus, the average rate of growth in the population over these four years is 16.47 %.

Note : To find the 4th root in this example, the square root of 184009200 is taken and its square root is taken again. Similarly, to find the 8th root of a number the process of taking square root should be repeated three times.

Illustration 15 : The geometric mean of two numbers is 2. If one number is 4 times the other number, find the numbers.

Suppose the smaller number is $x_1 = a$.

Then the larger number which is 4 times this number will be $x_2 = 4a$

and $G = 2$.

$$\begin{aligned}G &= \sqrt{x_1 \times x_2} \\ \therefore 2 &= \sqrt{a \times 4a} \\ \therefore 2 &= \sqrt{4a^2} \\ \therefore 2 &= 2a \\ \therefore a &= 1\end{aligned}$$

Thus, we get the first number $x_1 = a = 1$ and the second number $x_2 = 4a = 4$.

Illustration 16: Find the arithmetic mean and geometric mean of two numbers 9 and 16 and verify that $\bar{x} > G$.

Here $x_1 = 9$, $x_2 = 16$ and $n = 2$

$$\text{Arithmetic mean } \bar{x} = \frac{\sum x}{n} = \frac{9+16}{2} = \frac{25}{2} = 12.5$$

$$\text{Geometric mean } G = \sqrt{x_1 \times x_2} = \sqrt{9 \times 16} = \sqrt{144} = 12$$

Since $\bar{x} = 12.5$ and $G = 12$ we can see that $\bar{x} > G$

Advantages of geometric mean :

- (1) It is rigidly defined.
- (2) It is based on all observations.
- (3) It is suitable for algebraic calculations.
- (4) It is comparatively less affected by too large or too small values.

Disadvantages of geometric mean :

- (1) It can be found only if all the observations have positive values.
- (2) It is difficult to calculate.
- (3) Its calculation becomes difficult if the number of observations is large.

Activity

Find the G.M. and A.M. of 4 values 1, 7, 5, 100.
Which average is more appropriate ? Why ?

EXERCISE 3.3

- (1) The following data show the number of books read by 8 students of a class during last month.
2, 1, 5, 9, 1, 3, 2, 4
Find the average number of books read using geometric mean.
- (2) The value of a machine depreciates at the rate of 10 % , 7 % , 5 % and 2 % in its first four years respectively. Find the average rate of depreciation using an appropriate method.
- (3) A taxi travelled 15 km on Monday and 254 km on Tuesday. Find the average distance travelled over these two days using geometric mean.

*

3.4 Measures of Positional Average

Median, Quartiles, Deciles, Percentiles :

We studied that mean is an appropriate average if we have data which are evenly distributed around the average and the data which do not have too large or too small values. It is said that mean does not become a good representative of data if these conditions are not satisfied. Another average is more suitable in such situations which is called as median. It is a positional average. In addition to median, quartiles, deciles and percentiles are also other positional averages.

3.4.1 Meaning

Median, quartiles, deciles and percentiles are called positional averages because their values are found using the value of an observation at a specific position among the values of variable in the ordered data.

Median :

Median is defined as the value of middlemost observation when the data are arranged in either ascending or descending order. It is denoted by M . In other words, 50% values of observations in the data are above the median and 50% observations have value less than the median.

Calculation of median :

For raw data :

As we have to find the value at the centre, the observations have to be arranged in ascending or descending order.

For n observations x_1, x_2, \dots, x_n , median is found as follows :

Median $M =$ value of the $\left[\frac{n+1}{2}\right]$ th observation

For example, if we have 15 observations, the value of the $\left[\frac{15+1}{2}\right]$ th that is the 8th observation will be exactly the central value, which is called as median.

Suppose the given data consists of 20 observations. Then, as $\frac{n+1}{2} = \frac{20+1}{2} = 10.5$, we say that the 10th and the 11th observations are both in the centre. In this case, median will be taken as the mean of these two central values.

Illustration 17 : The numbers of items produced by a factory in different weeks are 80, 85, 90, 92, 68, 80, 72, 63, 55. Find the median production.

Arrangement of observations in ascending order is as follows :

55, 63, 68, 72, 80, 80, 85, 90, 92

Here, $n = 9$

$$\begin{aligned} \text{Median } M &= \left\{ \frac{n+1}{2} \right\} \text{th observation} \\ &= \text{value of the } \left\{ \frac{9+1}{2} \right\} \text{th observation} \\ &= \text{value of the 5th observation} \\ &= 80 \end{aligned}$$

Thus, the median production of this factory is 80 items.

Illustration 18 : The daily profits of a hawker (in ₹) for last 10 days are given below. Find the median profit.

261.5, 257, 258.5, 260, 265, 249, 255.5, 262.5, 264, 267

Arranging in ascending order, the observations will be as follows :

249, 255.5, 257, 258.5, 260, 261.5, 262.5, 264, 265, 267

$$\begin{aligned} \text{Median } M &= \text{value of the } \left\{ \frac{n-1}{2} \right\} \text{th observation} \\ &= \text{value of the } \left\{ \frac{10-1}{2} \right\} \text{th observation} \\ &= \text{value of the 5.5th observation} \\ &= \frac{\text{value of the 5th observation} + \text{value of the 6th observation}}{2} \\ &= \frac{260 + 261.5}{2} \\ &= 260.75 \end{aligned}$$

Thus, median daily profits of the hawker is ₹ 260.75.

Illustration 19 : There are 11 employees in an office. The monthly salaries (in ₹) of the lowest paid 7 employees among them are 4500, 2100, 3400, 3600, 2500, 4200, 1500. What is the median monthly salary of all employees ?

Some of the observations are missing here. We do not know the salaries of highest paid 4 employees.

Suppose these values are a, b, c, d in their increasing order. These values are greater than the given observations as these are salaries of highest paid employees.

We will arrange these data in ascending order. 1500, 2100, 2500, 3400, 3600, 4200, 4500, a, b, c, d .

Here, $n = 11$

$$\begin{aligned} \text{Median } M &= \text{Value of the } \left\{ \frac{n+1}{2} \right\} \text{th observation} \\ &= \text{Value of the } \left\{ \frac{11+1}{2} \right\} \text{th observation} \\ &= \text{Value of the 6th observation} \\ &= 4200 \end{aligned}$$

Thus, the median salary of these employees is ₹ 4200.

For grouped data :

For discrete frequency distribution :

Suppose x_1, x_2, \dots, x_k are the values of a variable with frequencies f_1, f_2, \dots, f_k respectively.

A frequency distribution generally shows the observations arranged in ascending order. We shall use cumulative frequencies to find the median for a frequency distribution where observations are arranged in an ascending order.

Here, the median is found as follows :

Median $M =$ value of the $\left\lfloor \frac{n+1}{2} \right\rfloor$ th observation

Where $n = f_1 + f_2 + \dots + f_k = \sum f_i =$ total number of observations

Illustration 20 : The following table shows the record of absent students of class during a month.

Find the median of number of absent days per student.

No. of absent days of student	0	1	2	3	4	5
No. of students	8	12	18	9	5	1

We will find cumulative frequency distribution as shown below :

No of absent days x	0	1	2	3	4	5	Total
No of students f	8	12	18	9	5	1	53
cumulative frequency cf	8	20	38	47	52	53	–

Here $n = \sum f = 53$

Median $M =$ value of the $\left\lfloor \frac{n+1}{2} \right\rfloor$ th observation

$=$ value of the $\left\lfloor \frac{53+1}{2} \right\rfloor$ th observation

$=$ value of the 27th observation

It can be known from the cumulative frequencies that the 21st to the 38th observations have value 2.

Hence, the 27th observation has value 2. \therefore median $M = 2$ days.

Thus, the median number of absent days is 2 days.

Illustration 21 : The time required for typing a report by different typists is given in the following data. Find the median typing time using it.

Time for typing (min.)	10	11	12	13	14
No. of typists	5	7	8	15	5

The cumulative frequency distribution is as follows :

Time for typing x	10	11	12	13	14
No. of typists f	5	7	8	15	5
Cumulative frequency cf	5	12	20	35	40

Here, $n = \Sigma f = 40$

$$\begin{aligned} \text{Median } M &= \text{value of the } \left\lfloor \frac{n+1}{2} \right\rfloor \text{th observation} \\ &= \text{value of the } \left\lfloor \frac{40+1}{2} \right\rfloor \text{th observation} \\ &= \text{value of the 20.5th observation} \\ &= \frac{\text{value of the 20th observation} + \text{value of the 21st observation}}{2} \end{aligned}$$

The observations of this discrete frequency distribution are as follows in their increasing order : 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 12, 12, 12, 13, 13, 13, ... An inspection of above observations shows that the 20.5th observation is a mean of 12 and 13.

It can be known from the cumulative frequencies that the 13th to the 20th observations have value 12 and the 21th to the 35th observations have value 13.

Thus, the 20th and the 21st observations are 12 and 13 respectively.

$$\begin{aligned} \therefore M &= \frac{12+13}{2} \\ &= 12.5 \end{aligned}$$

Thus, the median time required for typing is 12.5 min.

For continuous frequency distribution :

A continuous frequency gives the values of the variable in the form of class intervals and they are generally arranged in ascending order. In such cases, we will use the cumulative frequencies to find the median. These cumulative frequencies will show us the class containing median. For this, we take Median

class = class containing the $\left\lfloor \frac{n}{2} \right\rfloor$ th observation

Where $n = f_1 + f_2 + \dots + f_k = \Sigma f_i =$ total number of observations

The following formula is used to find the median :

$$\text{Median } M = L + \frac{\frac{n}{2} - cf}{f} \times c$$

Where L = lower boundary point of the median class

cf = cumulative frequency of the class prior to median class

f = frequency of the median class

c = length of median class

Illustration 22 : The number of cheques deposited in a bank per day has the following frequency distribution :

No. of cheques	0 - 39	40 - 79	80 - 119	120 - 159	160 - 199
No. of days	2	14	23	7	4

Find the median number of cheques deposited.

We will find the cumulative frequency distribution as shown below:

No. of cheques	0 - 39	40 - 79	80 - 119	120 - 159	160 - 199
No. of days	2	14	23	7	4
Cumulative frequency	2	16	39	46	50

Here $n = \Sigma f = 50$

Median class = Class of the $\left\lfloor \frac{n}{2} \right\rfloor$ th observation

= Class of the $\left\lfloor \frac{50}{2} \right\rfloor$ th observation

= Class of the 25th observation

It can be known from the cumulative frequencies that all the observations from the 17th observation to the 39th observation lie in 80 - 119. Hence it will be the median class.

Since this is exclusive type of classification, we will obtain the class boundary points from the class limits. Hence the median class will be taken as 79.5 - 119.5.

Taking, $L = 79.5$, $cf = 16$, $f = 23$, $c = 40$,

$$\begin{aligned} \text{Median } M &= L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c \\ &= 79.5 + \frac{25 - 16}{23} \times 40 \\ &= 79.5 + \frac{9}{23} \times 40 \\ &= 79.5 + \frac{360}{23} \\ &= 79.5 + 15.6522 \\ &= 95.1522 \\ &= 95.15 \end{aligned}$$

Thus, the median for number of cheques deposited in the bank per day is 95.15 cheques.

Illustration 23: The data about monthly expenditure on petrol for 75 families is given in the following table. Find median expenditure on petrol for these families.

Expenditure (₹) on petrol	Upto 200	Upto 400	Upto 600	Upto 800	Upto 1000	Upto 1200
No. of families	2	8	17	32	57	75

We are given the cumulative frequencies. We will find the frequency distribution.

Expenditure (₹)	Upto 200	200 - 400	400 - 600	600 - 800	800 - 1000	1000 - 1200
No. of families	2	6	9	15	25	18
Cumulative Frequency cf	2	8	17	32	57	75

Here $n = \Sigma f = 75$

Median class = class of the $\left\lfloor \frac{n}{2} \right\rfloor$ th observation

= class of the $\left\lfloor \frac{75}{2} \right\rfloor$ th observation

= class of the 37.5th observation

It can be known from the cumulative frequencies that the 37th and the 38th observations both lie in the class 800 - 1000. Thus the median class will be 800-1000.

Taking, $L = 800$, $cf = 32$, $f = 25$, $c = 200$,

$$\begin{aligned} \text{Median } M &= L + \frac{\frac{n}{2} - cf}{f} \times c \\ &= 800 + \frac{37.5 - 32}{25} \times 200 \\ &= 800 + \frac{5.5}{25} \times 200 \\ &= 800 + \frac{1100}{25} \\ &= 800 + 44 \\ &= 844 \end{aligned}$$

Thus, the median monthly expenditure on petrol of these families is ₹ 844.

Illustration 24: The level of air pollution (in ppm) in a city on different days is as follows. Find the median level of pollution.

Level of pollution (ppm)	250 and above	270 and above	290 and above	310 and above	320 and above	330 and above	340 and above
No. of days	150	133	108	76	41	20	7

A cumulative frequency distribution of 'more than' type is given here. We shall obtain the frequency distribution as well as the cumulative distribution of 'less than' from it.

Level of pollution	250 - 270	270 - 290	290 - 310	310 - 320	320 - 330	330 - 340	340 and above
No. of days	17	25	32	35	21	13	7
Cumulative frequency cf	17	42	74	109	130	143	150

$n = \sum f = 150$. Here, the classes have unequal lengths.

$$\begin{aligned} \text{Median class} &= \text{class of the } \left\{ \frac{n}{2} \right\} \text{th observation} \\ &= \text{class of the } \left\{ \frac{150}{2} \right\} \text{th observation} \\ &= \text{class of the 75th observation} \end{aligned}$$

It can be known from the cumulative frequencies that the 75th observation lies in the class 310 - 320.

Hence median class will be 310 - 320.

Taking $L = 310$, $cf = 74$, $f = 35$, $c = 10$;

$$\begin{aligned}
 \text{Median } M &= L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c \\
 &= 310 + \frac{75 - 74}{35} \times 10 \\
 &= 310 + \frac{1}{35} \times 10 \\
 &= 310 + \frac{10}{35} \\
 &= 310 + 0.2857 \\
 &= 310.2857 \\
 &\approx 310.29
 \end{aligned}$$

Thus, the median for level of pollution 310.29 ppm

Advantages and disadvantages of median :

Advantages :

- (1) It is easy to calculate and to understand.
- (2) It can be found by inspection.
- (3) It can be located by graph.
- (4) It is the only available average when the frequency distribution has open ended classes.
- (5) It is less affected by too large or too small values.
- (6) It can be calculated even if certain data are missing.

Disadvantages :

- (1) It is not rigidly defined.
- (2) It is not based on all values.
- (3) It is not suitable for further algebraic operations.
- (4) It is less stable measure of central tendency as compared to mean.

Other positional averages :

We saw that the median divides the data in two equal parts. Sometimes we require values that divide data in more parts. We shall now study a few such positional averages.

Quartiles :

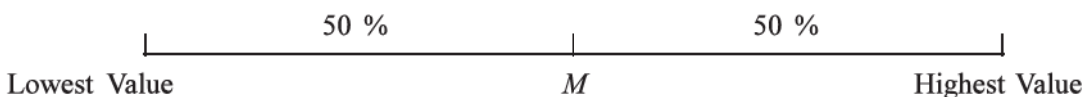
When the observations of the given data are arranged in ascending order, three values which divide the data in four equal parts are called quartiles. These three quartiles are denoted by Q_1 , Q_2 and Q_3 respectively.

First 25% values of the data will be less than or equal to Q_1 , next 25% values of the data will be between Q_1 and Q_2 ; and the further 25% values will be between Q_2 and Q_3 . Hence we have 50% data values lying between Q_1 and Q_3 whereas 25% observations have value above Q_3 .

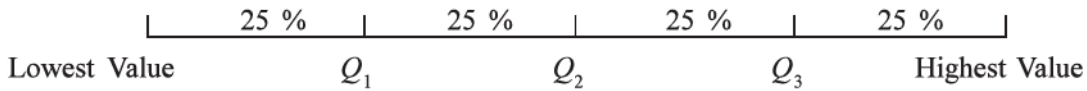
We can say that the j th quartile Q_j will divide the data such that $25j\%$ observations will be below Q_j . ($j = 1, 2, 3$)

Thus, quartile Q_2 will have $(25 \times 2)\%$ or 50% observations below Q_2 . Hence $Q_2 = \text{Median} = M$

For median :



For quartiles :



Deciles :

Suppose the observations are arranged in the ascending order. Nine values which will divide the data in ten equal parts are called as deciles which are denoted by D_1, D_2, \dots, D_9 respectively. 10% observations will have value less than D_1 , 20% observations will have value less than D_2 and so on. Thus, 10 j % observations will have value less than the j th decile D_j ($j = 1, 2, \dots, 9$).

We can see that $D_5 = M = Q_2$.

Percentiles :

Suppose the observations are arranged in the ascending order. The 99 values which divide the data in 100 equal parts are called percentiles which are denoted by P_1, P_2, \dots, P_{99} respectively. Here, 100 j % observations will have value less than the j th percentile P_j ($j = 1, 2, \dots, 99$).

We can see that $D_1 = P_{10}, D_2 = P_{20}, \dots, D_9 = P_{90}$ Similarly, $Q_1 = P_{25}$ and $Q_3 = P_{75}$

Moreover, $M = Q_2 = D_5 = P_{50}$

Since median, quartiles, deciles and percentiles are all positional averages, the calculation will have similar method. The following table shows the formula for finding the j th quartile Q_j , the j th decile D_j and the j th percentile P_j .

Type of Data	j th Quartile $j = 1, 2, 3$	j th Decile $j = 1, 2, \dots, 9$	j th Percentile $j = 1, 2, \dots, 99$
Raw data and Discrete frequency distribution	$Q_j =$ value of the $j\left\{\frac{n+1}{4}\right\}$ th observation	$D_j =$ value of the $j\left\{\frac{n+1}{10}\right\}$ th observation	$P_j =$ value of the $j\left\{\frac{n+1}{100}\right\}$ th observation
Continuous frequency distribution	Class of $Q_j =$ class of the $j\left\{\frac{n}{4}\right\}$ th observation $Q_j = L + \frac{j\left\{\frac{n}{4}\right\} - cf}{f} \times c$ Where $L =$ lower boundary point of class of Q_j $cf =$ cumulative frequency of the class prior to class of Q_j $f =$ frequency of the class of Q_j $c =$ length of class of Q_j	Class of $D_j =$ class of the $j\left\{\frac{n}{10}\right\}$ th observation $D_j = L + \frac{j\left\{\frac{n}{10}\right\} - cf}{f} \times c$ Where $L =$ lower boundary point of class of D_j $cf =$ cumulative frequency of the class prior to class of D_j $f =$ frequency of the class of D_j $c =$ length of class of D_j	Class of $P_j =$ class of the $j\left\{\frac{n}{100}\right\}$ th observation $P_j = L + \frac{j\left\{\frac{n}{100}\right\} - cf}{f} \times c$ Where $L =$ lower boundary point of class of P_j $cf =$ cumulative frequency of the class prior to class of P_j $f =$ frequency of the class of P_j $c =$ length of class of P_j

Illustration 25: Find Q_1 , D_7 , P_{40} for the following data showing runs scored by a batsman in his 20 innings.

32, 28, 47, 63, 71, 9, 60, 10, 96, 14, 31, 148, 53, 67, 29, 10, 62, 40, 80, 54

Arrangement of observations in ascending order is as following :

9, 10, 10, 14, 28, 29, 31, 32, 40, 47,
53, 54, 60, 62, 63, 67, 71, 80, 96, 148

Here, $n = 20$

Quartile $Q_1 =$ value of the $\left(\frac{n+1}{4}\right)$ th observation

$$= \text{value of the } \left(\frac{20+1}{4}\right)\text{th observation}$$

$$= \text{value of the 5.25th observation}$$

$$= \text{value of the 5th observation} + 0.25(\text{value of the 6th observation} -$$

value of the 7th observation)

$$= 28 + 0.25(29 - 28)$$

$$= 28 + 0.25$$

$$= 28.25$$

Decile $D_7 =$ value of the $7\left(\frac{n+1}{10}\right)$ th observation

$$= \text{value of the } 7\left(\frac{20+1}{10}\right)\text{th observation}$$

$$= \text{value of the 14.7th observation}$$

$$= \text{value of the 14th observation} + 0.7(\text{value of the 15th observation} -$$

value of the 14th observation)

$$= 62 + 0.7(63 - 62)$$

$$= 62 + 0.7$$

$$= 62.7$$

Percentile $P_{40} =$ value of the $40\left(\frac{n+1}{100}\right)$ th observation

$$= \text{value of the } 40\left(\frac{20+1}{100}\right)\text{th observation}$$

$$= \text{value of the 8.4th observation}$$

$$= \text{value of the 8th observation} + 0.4(\text{value of the 9th observation} -$$

value of the 8th observation)

$$= 32 + 0.4(40 - 32)$$

$$= 32 + 3.2$$

$$= 35.2$$

Thus the values of Q_1 , D_7 , P_{40} are 28.25 runs, 62.7 runs, and 35.2 runs respectively.

Illustration 26 : The following data refer to the milk quantity in 90 bags filled by a machine in a dairy.

Milk content (ml.)	485 - 490	490 - 495	495 - 500	500 - 505	505 - 510
No. of bags	5	21	33	23	8

Find Q_3 , D_2 , P_{55} and interpret them.

Let us find the table of cumulative frequencies for this continuous frequency distribution.

Milk content (ml.)	485 - 490	490 - 495	495 - 500	500 - 505	505 - 510
No. of bags	5	21	33	23	8
Cumulative frequency cf	5	26	59	82	90

Here, $n = 90$

For quartile Q_3 :

$$\begin{aligned} \text{Class of } Q_3 &= \text{Class of the } 3\left(\frac{n}{4}\right)\text{th observation} \\ &= \text{Class of the } 3\left(\frac{90}{4}\right)\text{th observation} \\ &= \text{Class of the } 67.5\text{th observation} \end{aligned}$$

It can be known from the cumulative frequencies that the 67th and the 68th observations are in the class 500 - 505.

Taking $L = 500$, $cf = 59$, $f = 23$, $c = 5$,

$$\begin{aligned} \text{Quartile } Q_3 &= L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c \\ &= 500 + \frac{67.5 - 59}{23} \times 5 \\ &= 500 + \frac{8.5}{23} \times 5 \\ &= 500 + \frac{42.5}{23} \\ &= 500 + 1.8478 \\ &= 501.8478 \\ &\approx 501.85 \end{aligned}$$

Thus, maximum quantity of milk is 501.85 ml in 75 % bags having least milk content.

For decile D_2 :

$$\begin{aligned} \text{Class of } D_2 &= \text{Class of the } 2\left(\frac{n}{10}\right)\text{th observation} \\ &= \text{Class of the } 2\left(\frac{90}{10}\right)\text{th observation} \\ &= \text{Class of the } 18\text{th observation} \end{aligned}$$

It can be found from the cumulative frequencies that the 18th observation is in the class 490-495.

Taking $L = 490$, $cf = 5$, $f = 21$, $c = 5$,

$$\begin{aligned}
 D_2 &= L + \frac{2\left(\frac{n}{10}\right) - cf}{f} \times c \\
 &= 490 + \frac{18-5}{21} \times 5 \\
 &= 490 + \frac{13}{21} \times 5 \\
 &= 490 + \frac{65}{21} \\
 &= 490 + 3.0952 \\
 &= 493.0952 \\
 &\approx 493.1
 \end{aligned}$$

Thus, maximum quantity of milk is 493.1 ml in 20% bags having least milk content.

For percentile P_{55} :

$$\begin{aligned}
 \text{Class of } P_{55} &= \text{class of the } 55\left(\frac{n}{100}\right)\text{th observation} \\
 &= \text{class of the } 55\left(\frac{90}{100}\right)\text{th observation} \\
 &= \text{class of the } 49.5\text{th observation}
 \end{aligned}$$

It can be found from the cumulative frequencies that the 49th and the 50th observations are in the class 495-500.

Taking $L = 495$, $cf = 26$, $f = 33$, $c = 5$,

$$\begin{aligned}
 P_{55} &= L + \frac{55\left(\frac{n}{100}\right) - cf}{f} \times c \\
 &= 495 + \frac{49.5-26}{33} \times 5 \\
 &= 495 + \frac{23.5}{33} \times 5 \\
 &= 495 + \frac{117.5}{33} \\
 &= 495 + 3.5606 \\
 &= 498.5606 \\
 &\approx 498.56
 \end{aligned}$$

Thus, maximum quantity of milk is 498.56 ml in 55% bags having least milk content.

Illustration 27 : We have the following information from a survey of 100 customers of a bank on their number of visits to the bank during a month.

No. of visists	0	1	2	3	4	5	6
No. of customers	12	22	40	15	6	4	1

Find the first quartile, 4th decile and 95th percentile.

No. of visits	0	1	2	3	4	5	6
No. of customers	12	22	40	15	6	4	1
Cumulative frequency <i>cf</i>	12	34	74	89	95	99	100

Here, $n = 100$

$$\begin{aligned} \text{First quartile } Q_1 &= \text{Value of the } \left(\frac{n+1}{4}\right)\text{th observation} \\ &= \text{Value of the } \left(\frac{100+1}{4}\right)\text{th observation} \\ &= \text{Value of the 25.25th observation} \end{aligned}$$

According to cumulative frequencies, the 25th and the 26th observations have value 1. $\therefore Q_1 = 1$
Thus, maximum number of visits is 1 among 25% least visiting customers.

$$\begin{aligned} \text{4th Decile } D_4 &= \text{Value of the } 4\left(\frac{n-1}{10}\right)\text{th observation} \\ &= \text{Value of the } 4\left(\frac{100-1}{10}\right)\text{th observation} \\ &= \text{Value of the 40.4th observation} \end{aligned}$$

According to cumulative frequencies, the 40th and the 41st observations have value 2. $\therefore D_4 = 2$

Thus, maximum number of visits is 2 among 40% least visiting customers.

$$\begin{aligned} \text{95th Percentile } P_{95} &= \text{Value of the } 95\left(\frac{n+1}{100}\right)\text{th observation} \\ &= \text{Value of the } 95\left(\frac{100+1}{100}\right)\text{th observation} \\ &= \text{Value of the 95.95th observation} \end{aligned}$$

According to cumulative frequencies, the 95th observation has value 4 and the 96th observation has value 5.

$$\begin{aligned} \therefore P_{95} &= \text{Value of the 95th observation} + 0.95 (\text{Value of the 96th observation} - \\ &\hspace{15em} \text{Value of the 95th observation}) \\ &= 4 + 0.95(5 - 4) \\ &= 4 + 0.95 \\ &= 4.95 \end{aligned}$$

Thus, maximum number of visits is 4.95 \approx 5 among 95% least visiting customers.

Illustration 28 : The following table shows data about monthly travelling expenses (in ₹) in a sample of 50 local bus commuters.

Monthly expenses	350 - 500	500 - 650	650 - 800	800 - 950	950 and above
No. of commuters	7	13	16	9	5

- (1) Find the limits for the expense of central 50% of the commuters.
- (2) Find the minimum expense of the highest spending 15% commuters.
- (3) Find the maximum expense in the lowest spending 10% commuters.

We will use the concepts of positional averages to answer the above questions. For this we find the cumulative frequencies for this continuous frequency distribution.

Monthly expenses(₹)	350 - 500	500 - 650	650 - 800	800 - 950	950 and above
No. of commuters	7	13	16	9	5
Cumulative frequency <i>cf</i>	7	20	36	45	50

Here, $n = 50$

(1) For any variable, central 50% observations have values between Q_1 and Q_3 . Hence we will find Q_1 and Q_3 .

$$\begin{aligned}
 \text{Class of } Q_1 &= \text{class of the } \left(\frac{n}{4}\right)\text{th observation} \\
 &= \text{class of the } \left(\frac{50}{4}\right)\text{th observation} \\
 &= \text{class of the 12.5th observation}
 \end{aligned}$$

It can be found from the cumulative frequencies that the 12th and the 13th observations are in the class 500-650.

Taking $L = 500$, $cf = 7$, $f = 13$, $c = 150$,

$$\begin{aligned}
 Q_1 &= L + \frac{\left(\frac{n}{4}\right) - cf}{f} \times c \\
 &= 500 + \frac{12.5 - 7}{13} \times 150 \\
 &= 500 + \frac{5.5}{13} \times 150 \\
 &= 500 + \frac{825}{13} \\
 &= 500 + 63.4615 \\
 &= 563.4615 \\
 &\approx 563.46
 \end{aligned}$$

$$\begin{aligned}
 \text{Class of } Q_3 &= \text{class of the } 3\left(\frac{n}{4}\right)\text{th observation} \\
 &= \text{class of the } 3\left(\frac{50}{4}\right)\text{th observation} \\
 &= \text{class of the 37.5th observation}
 \end{aligned}$$

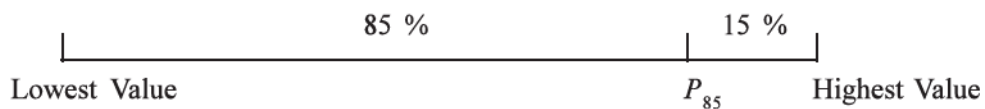
It can be known from the cumulative frequencies that the 37th and the 38th observations are in the class 800 - 950.

Taking $L = 800$, $cf = 36$, $f = 9$, $c = 150$,

$$\begin{aligned}
 Q_3 &= L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c \\
 &= 800 + \frac{37.5 - 36}{9} \times 150 \\
 &= 800 + \frac{1.5}{9} \times 150 \\
 &= 800 + \frac{225}{9} \\
 &= 800 + 25 \\
 &= 825
 \end{aligned}$$

Thus, central 50% commuters have monthly travelling expenses between ₹ 563.46 and ₹ 825.

(2) Here we want a value so that 15% observations should be above it, which means 85% observations should have value below it. Hence we find P_{85} .



$$\begin{aligned}
 \text{Class of } P_{85} &= \text{Class of the } 85\left(\frac{n}{100}\right)\text{th observation} \\
 &= \text{Class of the } 85\left(\frac{50}{100}\right)\text{th observation} \\
 &= \text{Class of the 42.5th observation}
 \end{aligned}$$

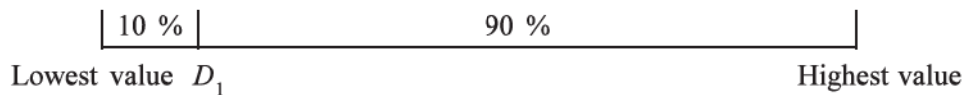
It can be known from the cumulative frequencies that the 42nd and the 43rd observations are in the class 800 - 950.

Taking $L = 800$, $cf = 36$, $f = 9$, $c = 150$,

$$\begin{aligned}
 \text{Class of } P_{85} &= \text{Class of } L + \frac{85\left(\frac{n}{100}\right) - cf}{f} \times c \\
 &= 800 + \frac{42.5 - 36}{9} \times 150 \\
 &= 800 + \frac{6.5}{9} \times 150 \\
 &= 800 + \frac{975}{9} \\
 &= 800 + 108.3333 \\
 &= 908.3333 \\
 &\approx 908.33
 \end{aligned}$$

Thus, minimum expense is ₹ 908.33 among highest spending 15% commuters.

(3) To find a value so that 10% observations are below it, we find D_1 .



$$\begin{aligned} \text{Class of } D_1 &= \text{class of the } \left(\frac{n}{10}\right)\text{th observation} \\ &= \text{class of the } \left(\frac{50}{10}\right)\text{th observation} \\ &= \text{class of the 5th observation} \end{aligned}$$

It can be found from the cumulative frequencies that the 5th observation is in the class 350 – 500.

Taking $L = 350$, $cf = 0$, $f = 7$, $c = 150$,

$$\begin{aligned} D_1 &= L + \frac{\left(\frac{n}{10}\right) - cf}{f} \times c \\ &= 350 + \frac{5-0}{7} \times 150 \\ &= 350 + \frac{5}{7} \times 150 \\ &= 350 + \frac{750}{7} \\ &= 350 + 107.1429 \\ &= 457.1429 \\ &\approx 457.14 \end{aligned}$$

Thus, the maximum expense among the lowest spending 10 % commuters is ₹ 457.14.

EXERCISE 3.4

1. Find all quartiles for the data given below about marks scored by 15 students in class test.

8, 6, 7, 0, 2, 4, 6, 5, 5, 4, 8, 9, 3, 6, 7

2. The following table shows data about the distance travelled (in km) by a salesman on different days.

Find median, Q_3 , D_8 , P_{62} and interpret them.

Distance travelled (km)	0–100	100–200	200–300	300–400	400–500	500–600
No. of days	5	18	24	7	5	1

3. The following table gives ages of 80 students selected from a college.

Age (years)	17	18	19	20	21	22	23
No. of students	11	14	22	15	8	6	4

Find median age. Also find Q_1 , D_4 , P_{32} for age and interpret them.

4. Use of following data to find the median salary of employees in a firm. Also find the lower limit for the richest 20% employees.

Salary (thousand ₹)	5 or more	10 or more	15 or more	20 or more	25 or more	30 or more
No. of employees	120	117	106	76	31	12

5. The following table shows the monthly expense for entertainment in a group of 100 students. Find the median of this expense.

Expense (₹)	Less than 200	200 - 400	400 - 600	600 - 700	700 - 800	800 and above
No. of students	8	23	40	17	7	5

6. The following data indicate records of hospital stays (in days) of 30 patients admitted to a hospital.

1, 10, 2, 6, 3, 4, 15, 1, 5, 9, 2, 4, 3, 1, 10,
7, 3, 5, 4, 2, 4, 8, 5, 3, 1, 9, 6, 2, 3, 7

Find the median stay. Further convert this information in a continuous frequency distribution (inclusive type) by taking classes of equal length starting from 1 –3. Find the median from the frequency distribution and compare it with your earlier answer.

*

3.5 Mode

We have earlier studied the mean and the median as the measures of central tendency. We shall now study 'mode' as another measure which is extensively used in business and commercial fields.

3.5.1 Meaning :

The value which gets repeated maximum number of times or the value occurring with maximum frequency in the given data is called as **mode**. It is denoted by M_o .

It is very often used in business to give a representative value for a set of data. For example, see the following statements :

- (1) On an average 3 languages are known to the students of this school.
- (2) The average height of the men in our country is 1.7 m.
- (3) The average daily production of our company is 50 items.
- (4) The average daily overtime put in by the workers of a factory is 3 hours.

The value which is repeated most number of times is considered in the calculation of the average in these situations. As per the first statement, it is implied that most of the students know three languages. Thus we can say that mode is used as an average here.

Calculation of mode :

For raw data and for discrete frequency distribution :

In these cases the mode can be found simply by inspection. We can find the mode as a value among the observations which is repeated maximum number of times or the one which has maximum frequency.

Illustration 29 : The numbers of books purchased by each of the 15 persons from a book store are as follows.

1, 0, 2, 2, 3, 4, 2, 7, 2, 2, 5, 4, 2, 1, 2.

Find the modal value for the number of books purchased.

We can see that the value of 2 is repeated 7 times which is more than the number of repetitions of any value of the other observations. Hence mode $M_o = 2$.

Thus, the mode of the number of books purchased is 2.

Illustration 30 : TV sets assembled by a TV manufacturing company in a month are tested. The following table shows the numbers of defects per TV set. Find the mode for the number of defects.

No. of defects	0	1	2	3	4	5
No. of TV sets	45	22	18	10	6	4

An inspection of frequencies shows that the observation 0 has the maximum frequency 45.

Hence $M_o = 0$

Thus the mode for the number of defects in TV sets is 0.

Note : According to the definition, the value of mode in this illustration is 0. But it cannot be taken as a measure of central tendency for the data as the value of mode is in the beginning of the data.

Mean or Median should be chosen as the measures of central tendency for such data or the value of mode should be found using empirical formula based on the values of mean and median which will be discussed in the later part of this chapter.

Illustration 31 : The number of trips made by 24 taxi drivers in a day are shown in the following data. Find the modal number of trips.

No. of trips	1	2	4	5	6	7
No. of drivers	3	7	4	7	2	1

The maximum frequency is 7 which is the frequency for the observations 2 and 5. Here, we say that this distribution has two modes $M_o = 2$ and $M_o = 5$.

Thus, the modes for the number of trips by the taxi driver are 2 and 5.

Note : Such a distribution is called as a bimodal distribution. Similarly there can be distributions with more than two modes.

Illustration 32: The number of patients arriving at a clinic each hour during working hours of a day are recorded as follows.

3, 5, 4, 2, 7, 8

Find the mode for number of patients.

As all the values are appearing only once, we can't find the most common observation.

Hence the mode for the number of patients cannot be found from the given data using the definition.

For continuous frequency distribution :

When the data are converted into a frequency distribution with classes, the exact values of the observations are not available.

Similar to median, for mode also, the class containing mode is found first and the value of mode is found using it.

The class having maximum frequency is called as modal class of the frequency distribution. The mode is further obtained using the following formula.

$$\text{Mode } M_o = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$$

Where L = lower boundary point of the modal class

f_m = frequency of the modal class

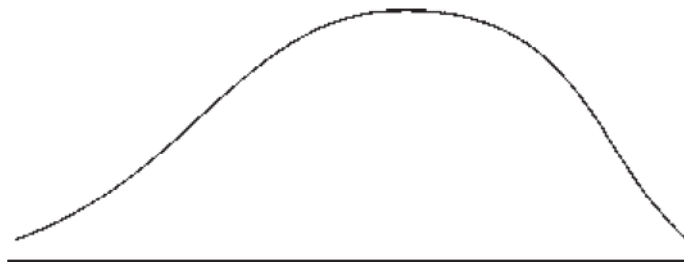
f_1 = frequency of the class prior to modal class

f_2 = frequency of the class succeeding to modal class

c = class length of the modal class

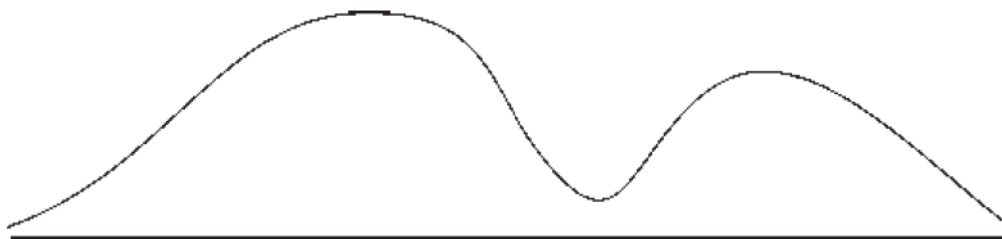
Note : The above formula can be used only if the distribution has classes of equal class length. Moreover, the formula can be used only in those cases where the maximum frequency is only for one class.

The frequency distribution in which the frequencies increase initially and then start decreasing after attaining the maximum frequency is called as a **regular frequency distribution**. Such distributions are also called as **unimodal distributions** as the distribution has only one mode. The frequency curve of such distributions is as follows :



Frequency curve of regular distribution

For bimodal distribution, the frequencies increase and then decrease but then again increase and decrease. Such a frequency distribution is called as an **irregular frequency distribution** whose frequency curve is as follows.



Frequency curve of irregular distribution

Illustration 33 : The following table gives output of workers in a factory. Find the modal output.

Output (no. of items)	150 - 160	160 - 170	170 - 180	180 - 190	190 - 200	200 - 210	210 - 220	220 - 230
No. of workers	4	5	19	33	48	22	12	6

The maximum frequency is 48 for the class 190-210. Hence the modal class is 190-200.

Taking, $L = 190, f_m = 48, f_1 = 33, f_2 = 22, c = 10,$

$$\begin{aligned}
 \text{Mode } M_o &= L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c \\
 &= 190 + \frac{48-33}{2(48)-33-22} \times 10 \\
 &= 190 + \frac{15}{96-33-22} \times 10 \\
 &= 190 + \frac{150}{41} \\
 &= 190 + 3.6585 \\
 &= 193.6585 \\
 &\approx 193.66
 \end{aligned}$$

Thus, the modal output is 193.66 items.

Illustration 34 : The number of cold drink bottles sold by a shopkeeper on different days is shown in the following table. Find the mode for the number of cold drink bottles sold.

No. of bottles	0 - 3	4 - 7	8 - 11	12 - 15	16 - 19	20 - 23
No. of days	3	11	16	20	18	12

The maximum frequency is 20 for the class 12 - 15. Hence the modal class is 12 - 15. This is an inclusive type of frequency distribution and hence we will take the boundary points of this class as 11.5 - 15.5.

Taking, $L = 11.5, f_m = 20, f_1 = 16, f_2 = 18, c = 4,$

$$\begin{aligned}
 \text{Mode } M_o &= L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c \\
 &= 11.5 + \frac{20-16}{2(20)-16-18} \times 4 \\
 &= 11.5 + \frac{4}{40-16-18} \times 4 \\
 &= 11.5 + \frac{16}{6} \\
 &= 11.5 + 2.6666 \\
 &= 14.1666 \\
 &\approx 14.17
 \end{aligned}$$

Thus, mode for sale of cold drink bottles is 14.17.

Empirical formula for mode :

We have observed that mode is not well defined in many cases. The noted statistician Karl Pearson established a relation between mean, median and mode by studying their values for different data sets. He observed that for data that are not evenly distributed around average, difference between mean and mode is approximately 3 times the difference between mean and median.

That is, (Mean – Mode) = 3 (Mean – Median)

The following formula is obtained to find the mode using this relation:

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

$$\text{This is written in notations as } M_o = 3M - 2\bar{x}$$

This formula to find mode is called as an empirical formula because the value obtained from observation and not from the theory. The value of mode found using this formula can be negative if the frequency distribution is not evenly distributed around the average.

This formula for mode is used in the following situations :

- Each observation appears just once in raw data.
- More than one observation in a frequency distribution appears with highest frequency.
- The continuous distribution has classes of unequal length.
- The frequency distribution is a mixed distribution that is a part of it is discrete and the rest is continuous.
- The right or left end of the curve of the frequency distribution is too extended.

Activity

Find the mean and median for the data given in Illustration 34 and verify the empirical formula.

Illustration 35 : The time between placing an order and its delivery was noted for a certain wholesaler as follows. Find the mode for this time.

Time (hours)	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 55
No. of orders	2	5	7	5	6	7	3

We can see that the maximum frequency is 7 which appears for two classes. Hence we shall use the empirical formula for finding the mode.

The following calculations are carried out to find mean and median :

Time (hrs.)	No. of orders f	Mid-value x	$d = \frac{x-A}{c}$ $A = 37.5$ $c = 5$	fd	Cumulative frequency cf
20 - 25	2	22.5	- 3	- 6	2
25 - 30	5	27.5	- 2	- 10	7
30 - 35	7	32.5	- 1	- 7	14
35 - 40	5	37.5	0	0	19
40 - 45	6	42.5	1	6	25
45 - 50	7	47.5	2	14	32
50 - 55	3	52.5	3	9	35
Total	$n = 35$			6	

$$\begin{aligned}
\text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\
&= 37.5 + \frac{6}{35} \times 5 \\
&= 37.5 + \frac{30}{35} \\
&= 37.5 + 0.8571 \\
&= 38.3571 \\
&\approx 38.36
\end{aligned}$$

$$\begin{aligned}
\text{Median class} &= \text{Class containing the } \left(\frac{n}{2}\right)\text{th observation} \\
&= \text{Class containing the } \left(\frac{35}{2}\right)\text{th observation} \\
&= \text{Class containing the 17.5th observation}
\end{aligned}$$

It can be known from the cumulative frequencies that, the 17th and the 18th observations lie in the class 35 - 40.

Taking $L = 35$, $cf = 14$, $f = 5$, $c = 5$,

$$\begin{aligned}
\text{Median } M &= L + \frac{\frac{n}{2} - cf}{f} \times c \\
&= 35 + \frac{17.5 - 14}{5} \times 5 \\
&= 35 + 3.5 \\
&= 38.5
\end{aligned}$$

$$\begin{aligned}
\text{Using the empirical formula, } M_o &= 3M - 2\bar{x} \\
&= 3(38.5) - 2(38.36) \\
&= 115.5 - 76.72 \\
&= 38.78
\end{aligned}$$

Thus, the mode for time between placing of an order and delivery is 38.78 hours.

Illustration 36 : The following table shows the number of visits to the dentist of the persons having toothache. Find the mode for number of visits.

No. of visits	1	2	3	4 - 7	7 - 10	10 - 15
No. of patients	7	11	18	9	4	1

The given distribution is mixed distribution and the class lengths are unequal. Thus, we will use the empirical formula for finding the mode. The calculations for mean and median are shown in the following table.

No. of visists	No. of patients f	Mid-value x	fx	Cumulative Frequency cf
1	7	1	7	7
2	11	2	22	18
3	18	3	54	36
4-7	9	5.5	49.5	45
7-10	4	8.5	34	49
10-15	1	12.5	12.5	50
Total	$n = 50$		179	

Note : The values of x are not too large. Hence we need not use the short cut method.

$$\text{Mean } \bar{x} = \frac{\sum fx}{n} = \frac{179}{50} = 3.58$$

$$\text{Median} = \text{Value of the } \left\{ \frac{n}{2} \right\} \text{th observation}$$

$$= \text{Value of the } \left\{ \frac{50}{2} \right\} \text{th observation}$$

$$= \text{Value of the 25th observation}$$

It can be known from the cumulative frequencies that the 25th observation has value 3.

$$\therefore M = 3 \text{ visits}$$

$$\begin{aligned} \text{Using the empirical formula, } M_o &= 3M - 2\bar{x} \\ &= 3(3) - 2(3.58) \\ &= 9 - 7.16 \\ &= 1.84 \end{aligned}$$

Thus, the mode for the number of visits to the dentist will be 1.84.

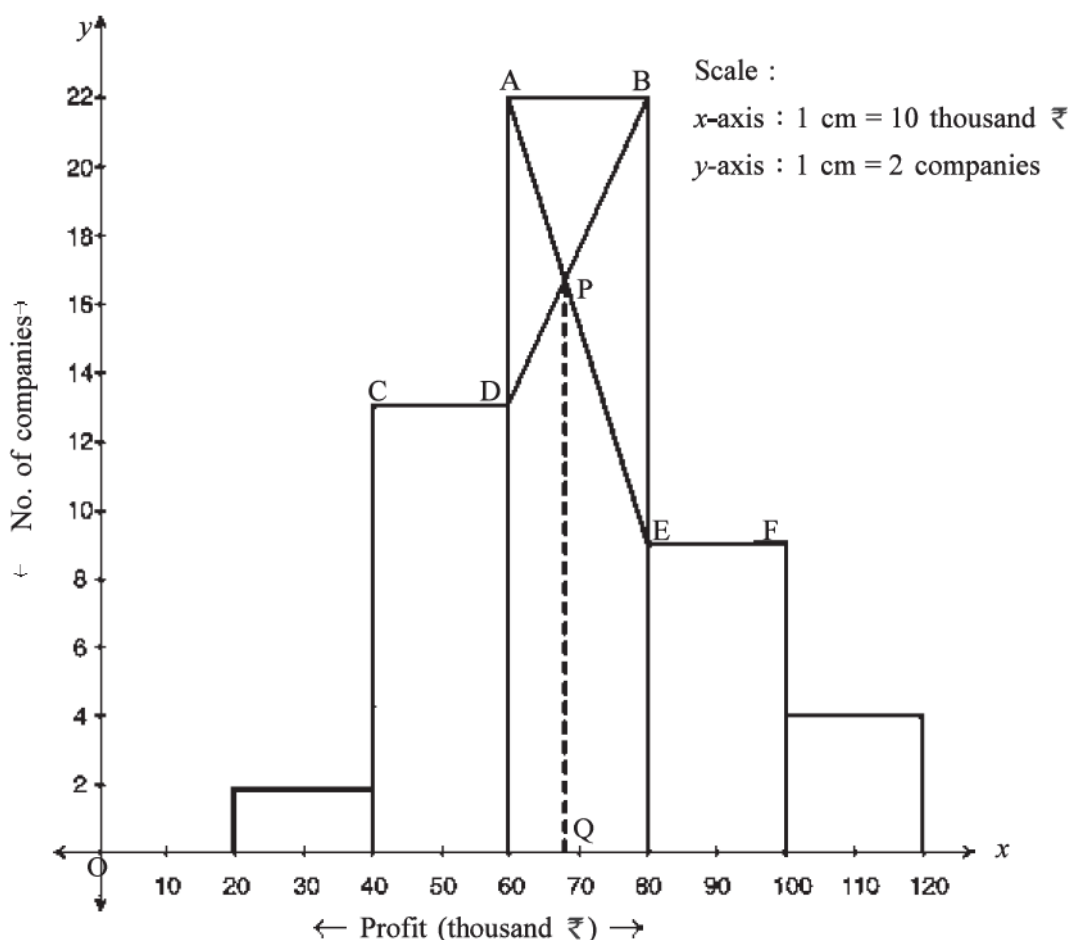
3.5.2 Graphical Method

The commonly used formula for mode (page no. 101) can not be used in a continuous frequency distribution having classes with unequal length. The mode can be found graphically for a continuous frequency distribution having classes with equal or unequal lengths. The histogram of the frequency distribution is used here. This method can be used only for unimodal distributions.

We shall understand this method with the help of the following data which are about the profits (in thousand ₹) of 50 small scale industrial units.

Profit (thousand ₹)	20 - 40	40 - 60	60 - 80	80 - 100	100 - 120
No. of compaines	2	13	22	9	4

The histogram for this frequency distribution is as follows :



To find the mode, we first consider the rectangle with maximum length which corresponds to the class interval with maximum frequency. In this case, it is the interval 60 - 80. Suppose the upper side of this rectangle is shown as AB. Now we consider the two rectangles adjacent to this rectangle. They will be the rectangles for the class intervals 40 - 60 and 80 - 100 respectively. We shall denote the upper sides of these rectangles as CD and EF respectively. Now we will draw a line segments AE joining points A and E as well as a segment BD. The point of intersection of AE and BD will be denoted by P. The point where a perpendicular drawn from the point P on x-axis meets x-axis will be denoted by Q. The distance of the point Q from the origin O will give us the value of mode. It can be seen from the above histogram that $OQ = 68$ (thousand ₹).

Hence, the mode of profits of these companies is ₹ 68 thousand.

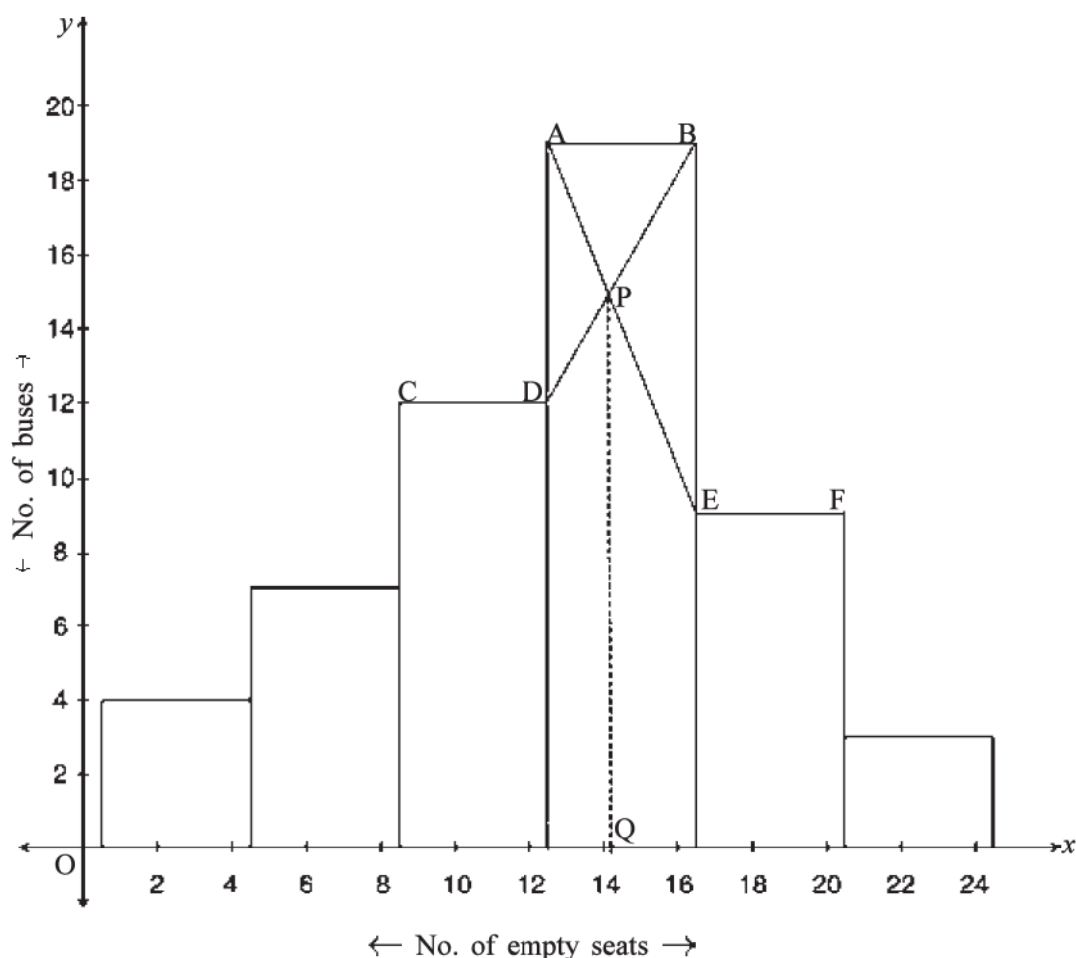
Illustration 37 : The following data show the number of empty seats per bus for the buses leaving from a depot. Find the modal number of empty seats using the graphical method.

No. of empty seats x	1 - 4	5 - 8	9 - 12	13 - 16	17 - 20	21 - 24
No. of buses	4	7	12	19	9	3

We can see that this is a continuous distribution of inclusive type. Therefore, first we find the class boundaries to draw the histogram.

No. of empty seats x	0.5-4.5	4.5-8.5	8.5-12.5	12.5-16.5	16.5-20.5	20.5-24.5
No. of buses	4	7	12	19	9	3

The histogram for this distribution is drawn below :



Following the method explained above, we find that $OQ = 14.2$. Hence modal number of empty seats for the buses leaving the depot is 14.2.

Illustration 38 : The distribution of the share prices of a company on different days is given in the following table. Find the mode for the share prices using the graphical method.

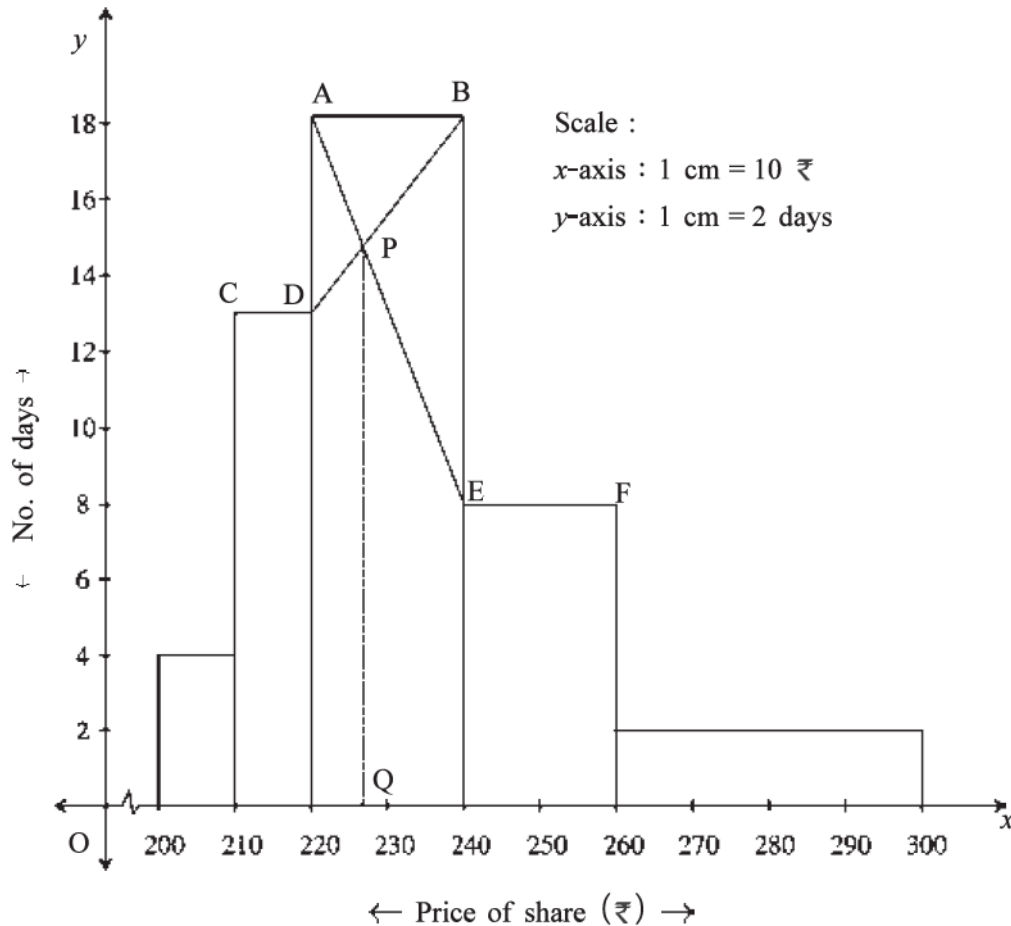
Share price (₹)	200 - 210	210 - 220	220 - 240	240 - 260	260 - 300
No. of days	4	13	36	16	8

We have to draw the histogram to find the mode by graphical method. This frequency distribution does not have equal class lengths and hence we first find the proportionate frequencies with respect to the smallest class length which are shown in the following table :

Share price	Class length	Frequency	Proportionate frequency
200 - 210	10	4	$\frac{4}{10} \times 10 = 4$
210 - 220	10	13	$\frac{13}{10} \times 10 = 13$
220 - 240	20	36	$\frac{36}{20} \times 10 = 18$
240 - 260	20	16	$\frac{16}{20} \times 10 = 8$
260 - 300	40	8	$\frac{8}{40} \times 10 = 2$

Proportionate frequency of each class = $\frac{\text{class frequency}}{\text{class length}} \times \text{minimum class length}$
 Here the minimum class length is 10.

The histogram drawn using the proportionate frequencies is as follows :



Using the graphical method, $OQ = 227 ₹$

Hence, the mode for share prices is ₹ 227.

Advantages and disadvantages of mode :

Advantages :

- (1) It is easy to understand and calculate.
- (2) It can be found merely by inspection.
- (3) It is not affected by too large or too small values.
- (4) Its value can be found using graph.

Disadvantages :

- (1) It is not rigidly defined.
- (2) There can be more than one mode for the given variable whereas sometimes the mode cannot be found.
- (3) It is not based on all observations.
- (4) It has less stability in sampling as compared to mean.
- (5) It is not suitable for further mathematical calculations.

EXERCISE 3.5

1. The IQ levels of students in a class are given below. Find the modal value of IQ level of students.
146, 134, 143, 144, 138, 145, 153, 138, 138, 146, 140, 135.

2. The following table gives the number of cakes sold each day at a bakery. Find the mode for sale of cakes.

No. of cakes	10	12	13	16	17	18
No. of days	5	9	25	16	10	7

3. The distribution of ages of 48 persons in an old age home is given below. Which formula will be appropriate to find the mode ? Why ? Find the modal age of the persons in the old age home using the formula you have chosen.

Age (years)	50 - 60	60 - 65	65 - 70	70 - 85	85 - 100
No. of persons	6	10	19	9	4

4. Comment on the mode for the following data showing the time taken (in seconds) for 8 competitors in a running race.

25.2, 26.5, 28.6, 32.1, 29.0, 29.3, 31.3, 27.8

5. The table below shows the data about weights of 86 apples from a garden. Find the mode for the Weight of apples.

Weight of apple (gram)	120 - 130	130 - 140	140 - 150	150 - 160	160 - 170	170 - 180	180 - 190
No. of apples	8	13	19	23	10	8	5

Also find the mode of this distribution graphically.

6. The data about the monthly house rent paid by 50 families is given in the following table :

House rent (thousand ₹)	0-5	5-10	10-20	20-30	30-50
No. of families	1	7	14	16	12

Find the mode for house rent using the graphical method.

*

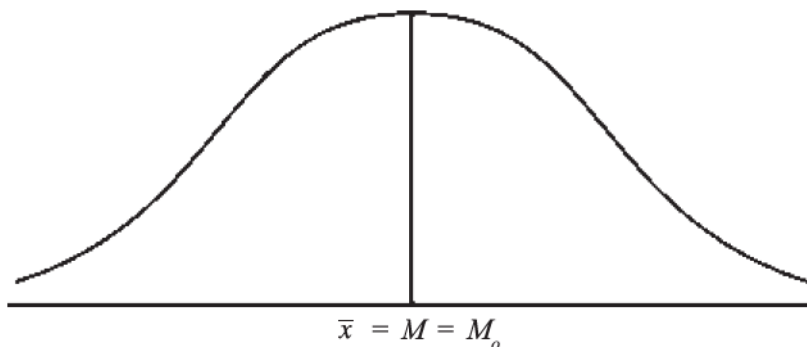
Some results for the measures of central tendency :

- (1) If all the observations in the given data have same value, all the measures of central tendency have the same value.

For example, if the ages (in years) of 5 students selected from a class are 15, 15, 15, 15, 15 years, all the measures of central tendency have the value 15.

Thus, $\bar{x} = M = M_o = G = \bar{x}_w = 15$.

- (2) Mean, median and mode have the same value if the data are evenly distributed around average.



(3) If the variable is multiplied by a non-zero constant b and a constant a is added to it, we get a variable $y = bx + a$

We saw in case of mean that using mean \bar{x} for x , we can find the mean for the variable y as $\bar{y} = b\bar{x} + a$.

Similarly, if we have the median or mode of x , we can find the median or mode of the variable y respectively.

$$\text{Median of } y = b (\text{median } x) + a$$

$$\text{Mode of } y = b (\text{mode of } x) + a$$

Illustration 39 :

- (1) The mean of a variable x is 25. Find the mean of the variable obtained by first subtracting 3 from x and then dividing it by 2.
- (2) The relation between price (p) of an item and its demand (d) is $d = 50 - 2p$. If the median of price is ₹ 11, find the mean of demand.
- (3) The mode of salaries of employees of a company is ₹ 8500. The company has decided to deduct 2% of each employee's salary for welfare fund. Find the mode for the amount of this fund.

(1) Here $y = \frac{x-3}{2}$. As $\bar{x} = 25$,

$$\begin{aligned} \bar{y} &= \frac{\bar{x}-3}{2} \\ &= \frac{25-3}{2} \\ &= \frac{22}{2} = 11 \end{aligned}$$

Thus, the mean of y is 11.

(2) $d = 50 - 2p$ and median of price (p) is 11.

$$\begin{aligned} \therefore \text{median of demand } (d) &= 50 - 2 (\text{median of } p) \\ &= 50 - 2 (11) \\ &= 50 - 22 \\ &= 28 \end{aligned}$$

Thus, the median of demand is 28 items.

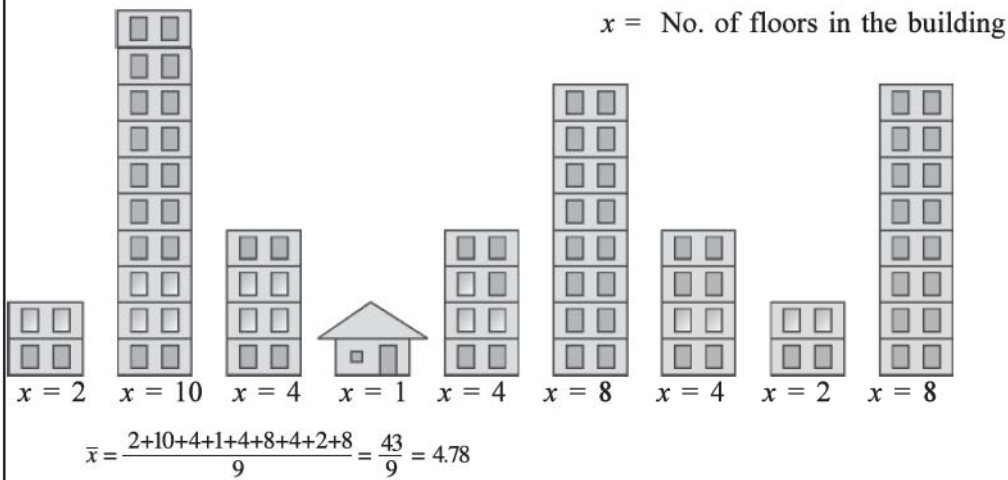
(3) The mode of salaries (x) of employees is ₹ 8500

$$\begin{aligned} \text{Amount of welfare fund } (y) &= 2 \% \times x \\ &= 0.02 x \end{aligned}$$

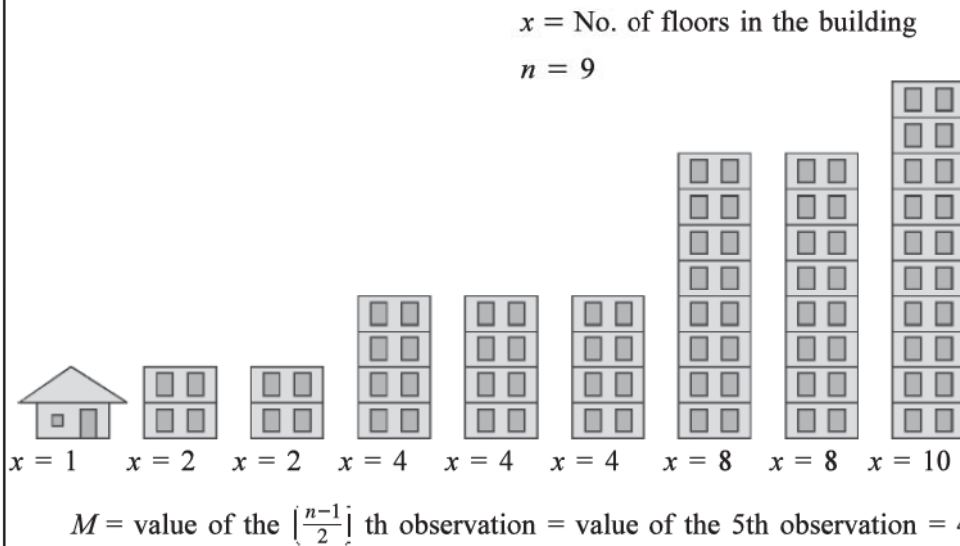
$$\begin{aligned} \therefore \text{mode for the amount of welfare fund } (y) &= 0.02 (\text{mode of } x) \\ &= 0.02 \times 8500 \\ &= 170 \end{aligned}$$

Thus, the mode for the amount deducted for welfare fund is ₹ 170.

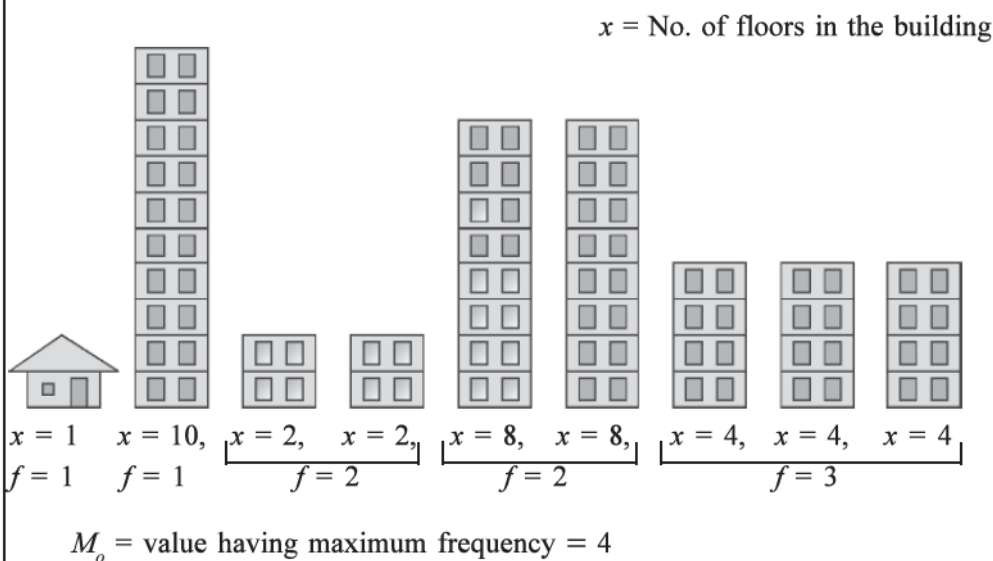
Study the following pictures for a general comparison of mean, median and mode :



There is no need to arrange the observations in any order to find mean.



The observations are arranged in the ascending order of their values to find median.



The observations having same values are arranged together to find mode.

3.6 Comparative study of mean, median and mode

We have discussed advantages and disadvantages of various measures of central tendency. It is obvious from them that any particular average cannot be suitable for all types of practical problems. Each average has some specific applications and also has certain limitations.

Among all the averages, mean satisfies most of the requisites of a good average and hence it is used in most situations of data analysis. The most important property of mean is its compatibility for further algebraic computations. The advanced statistical methods applied for studying various characteristics of a population or for comparing two populations use mean as a representative value for the variable under consideration. These points make mean an optimum measure of central tendency.

However, mean cannot truly represent the entire set of data when the data are not evenly distributed around average. Many variables studied in agriculture, social sciences and in business activities are not found to be evenly distributed. Median is a better measure of central tendency in these situations. Median is used as an average in qualitative data like education, skill and consumer satisfaction.

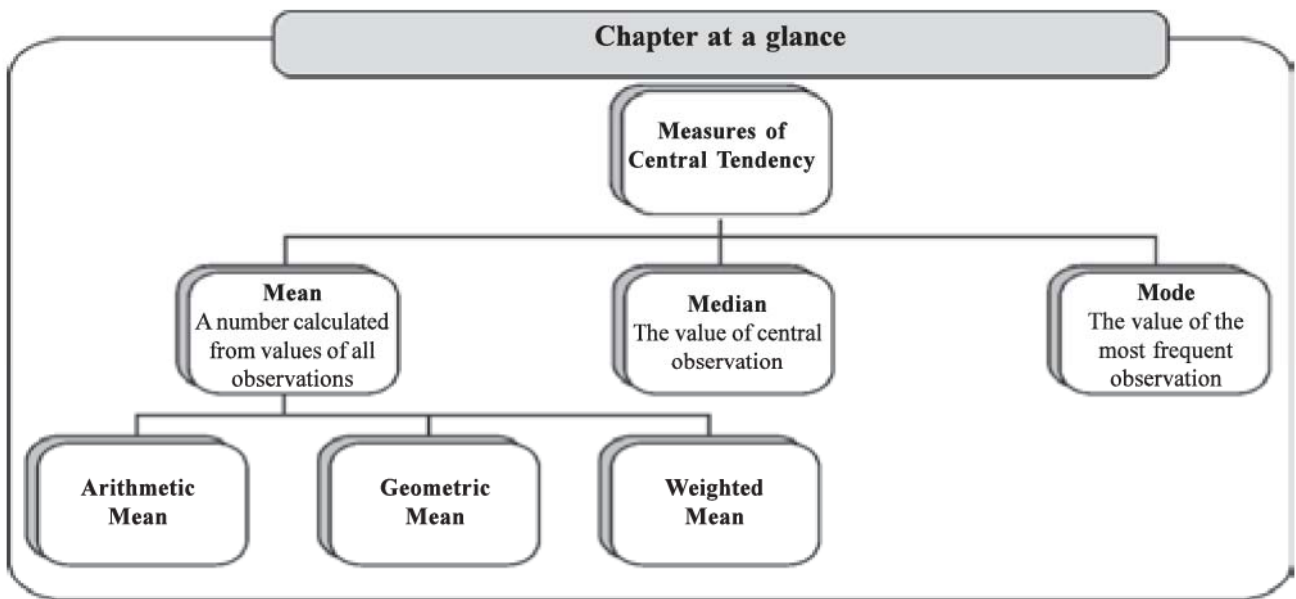
Mode is particularly used in business and commerce fields. For qualitative data also, mode is found to be useful as an average. While deciding the dishes served at a restaurant and their flavour, the choice and taste of maximum customers is taken into consideration which is an example of Mode. Mode is extensively used for finding the average by readymade garment manufacturers and in foot wear industry.

Thus, the selection of average depends upon the following factors :

(1) The nature of data (2) The nature of variable involved (3) The purpose of study (4) The type of classification used (5) The use of average for further statistical analysis

Summary

- The observations of any variable are concentrated around a certain value which is called as a measure of central tendency or average.
- Mean is the most popular average.
- Quartiles, Deciles, Percentiles are called as positional averages.
- Arithmetic mean is the sum of observations divided by the number of observations.
- If arithmetic means of two or more groups of observations are known, the combined mean can be found for the entire data.
- Weighted mean is found by assigning weights to the observations proportional to their importance.
- Geometric mean is the n th root of the product of n positive observations.
- Median is the middle most observation in the ordered data.
- Quartiles, deciles and percentiles divide the data in 4, 10 and 100 parts respectively.
- Mode is the most frequent observation for the given data.
- $M_o = 3M - 2\bar{x}$ is called as the empirical relation between mean, median and mode.



List of formulae :

(1) **Mean :**

Type of data		Short cut method
Raw data	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = A + \frac{\sum d}{n}$
Grouped data	$\bar{x} = \frac{\sum fx}{n}$	$\bar{x} = A + \frac{\sum fd}{n} \times c$

(2) **Combined mean :** $\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \dots + n_k\bar{x}_k}{n_1 + n_2 + \dots + n_k}$

(3) **Weighted mean :** $\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum wx}{\sum w}$

(4) **Geometric Mean :** $G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$

Median and other positional averages :

Positional average	Raw data and Discrete frequency distribution	Continuous frequency distribution
(5) Median	M = value of the $\left(\frac{n+1}{2}\right)$ th observation	$M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$
(6) j th Quartile	Q_j = value of the $j\left(\frac{n+1}{4}\right)$ th observation	$Q_j = L + \frac{j\left(\frac{n}{4}\right) - cf}{f} \times c$
(7) j th Decile	D_j = value of the $j\left(\frac{n+1}{10}\right)$ th observation	$D_j = L + \frac{j\left(\frac{n}{10}\right) - cf}{f} \times c$
(8) j th Percentile	P_j = value of the $j\left(\frac{n+1}{100}\right)$ th observation	$P_j = L + \frac{j\left(\frac{n}{100}\right) - cf}{f} \times c$

9. **Mode :**

Raw data	Discrete frequency distribution	Continuous frequency distribution
M_o = value of observation which gets repeated Maximum number of times	M_o = value of observation with maximum frequency	$M_o = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$

10. Empirical formula : $M_o = 3M - 2\bar{x}$

EXERCISE 3

Section A

Find the correct option for the following multiple choice questions :

- Which average gets most affected by too large or too small values ?
 (a) Arithmetic mean (b) Median (c) Mode (d) Geometric mean
- Which of the following will give us the value of median ?
 (a) D_7 (b) Q_1 (c) P_{45} (d) P_{50}
- In which of the following situations, mean cannot be found ?
 (a) class lengths are unequal, (b) there are open ended class intervals,
 (c) the number of class intervals is more than 5, (d) inclusive type of classes are used
- For any set of observations, which of the following is true ?
 (a) $\bar{x} \leq G$ (b) $\bar{x} = G$ (c) $\bar{x} \geq G$ (d) $\bar{x} > G$
- Which of the following results is true for the data that are evenly distributed around average ?
 (a) $\bar{x} = M = M_o$ (b) $\bar{x} > M > M_o$
 (c) $\bar{x} < M < M_o$ (d) $\bar{x} < M > M_o$
- If the mean of 10 observations is 15, what is the sum of observations ?
 (a) 25 (b) 150 (c) 5 (d) 1.5
- If $\Sigma(x-9)=0$ for data having 5 observations, then what is the value of mean ?
 (a) $\bar{x}=0$ (b) $\bar{x}=5$ (c) $\bar{x}=9$ (d) $\bar{x}=45$
- What is the mode of observations 7, 9, 9, 1, 7, 9, 4, 9, 1 ?
 (a) 1 (b) 4 (c) 7 (d) 9
- In a set of 50 observations, what is the median ?
 (a) value of 25th observation (b) value of 26th observation
 (c) value of 25.5th observation (d) value of 26.5th observation
- What is the geometric mean of 4 and 9 ?
 (a) 4 (b) 6 (c) 6.5 (d) 36
- If mean for a variable is 15 and its median is 20, what is the mode using empirical formula ?
 (a) 30 (b) 5 (c) 35 (d) 17.5

12. The median of 10 observations is 14. What will be the median of observations obtained when each observation gets doubled ?
 (a) 10 (b) 28 (c) 7 (d) 1.4
13. All the observations in a data are of same value 16. What will be their mode ?
 (a) 8 (b) 2 (c) 16 (d) 4
14. Which of the following statements is false ?
 (a) The quartiles divide the data in 4 equal parts.
 (b) The mean divides the data in 2 equal parts.
 (c) The percentiles divide the data in 100 equal parts.
 (d) The deciles divide the data in 10 equal parts.
15. The lengths (in meters) of 6 pipes manufactured by a company are as follows :
 1.05, 1.15, 0.98, 1.12, 0.89, 0.95
 Which of the following statements is true ?
 (a) Mode = 1m (b) Mode = 1.15 m (c) Mode = 0.98 m (d) Mode cannot be found

Section B

Answer the following questions in one sentence:

1. State any one advantage of mean.
2. If observations have varying importance, which average should be used ?
3. Name any two positional averages.
4. State the empirical relation between mean, median and mode.
5. State the condition under which geometric mean cannot be found.
6. Define mode.
7. State the name of the statistician who gave the empirical formula between mean, median and mode.
8. Median of 10 observations is 55. If the value of the largest observation increases from 100 to 110, find the new median.
9. Mean of variable x is 9. What is the mean of the variable $y = x + 4$?
10. Find the modal value of the variable having the following frequency distribution :

x	5	10	15	20	25
f	12	48	23	10	2
11. Arithmetic mean of two numbers is 5. If one number is 6, find the other number.
12. Find first quartile for variable with observations 15, 4, 7, 20, 2, 7, 13.
13. Which average can be obtained if the continuous frequency distribution has open ended classes ?
14. If $Q_3 = 25.75$ for a variable, then find P_{75} .
15. The median of daily demand of a vendor is 15. If he sells each item for ₹ 10, find the median of his revenue ?

Section C

Answer the following questions :

1. Define weighted mean.
2. Explain what is meant by measure of central tendency.
3. State the advantages of mode.
4. Explain the combined mean.
5. Which type of data have median as a better measure of central tendency than mean ?
6. What are the factors to be considered while choosing an appropriate average ?
7. The mean and mode of a variable are 5.5 and 6.4 respectively. Find its median.
8. Geometric mean of two numbers is 8. If one number is 4, find the other number.
9. Mean weekly production (x) of factory is 81units. Find the mean production cost if cost is given by $y = 3x + 50$
10. The median of observations $a - 5, a + 1, a + 2, a - 3, a$ is 10. Find a .
11. The mean of marks in Mathematics of 40 students in class is 76, whereas the same for the other class of 50 students is 85. Find the mean of marks in Mathematics of students in both the classes together.
12. The number of vehicles per family in families residing in a certain area are given in the following table. Find the median for the number of vehicles.

No. of vehicles	0	1	2	3	4	Total
No. of families	2	4	9	7	3	25

13. Find the weighted mean of variable x from the following data.

Variable x	1500	800	200
Weight w	5	4	1

Section D

Answer the following questions :

1. State the characteristics of an ideal average.
2. Define geometric mean and state its advantages.
3. Explain the use of mode as a measure of central tendency.
4. Explain the positional averages briefly.
5. Compare mean and median as the measures of central tendency.
6. Which average is called as optimum average ? why ?
7. The economy growth rates of a state for four consecutive years are 2%, 2.5%, 3%, 4% respectively. Find the average growth rate using an appropriate average.
8. Find D_7 and P_{15} from the following data about daily sales of a mobile phone shop and interpret them.

No. of phones	4	6	7	8	10	12
No. of days	3	9	15	23	8	2

9. The mean perfume content in the bottles filled by a perfume manufacturer's machine should be between 29.6 ml and 30.4 ml. The 7 bottles tested had the following perfume contents (in ml): 30.2, 28.9, 29.2, 30.1, 29.4, 31.3, 31.4
Is the machine working properly ?
10. The mean of marks scored by 34 boys in a class is 57. The mean of marks of all 60 students of the class is 59. Find the mean marks scored by the girls.
11. The mean of 50 observations is 35. Later on, it was known that the value of one observation was taken as 50, which was wrong. Find the mean of remaining observations by excluding the wrong observation.
12. 3 students from a group of 18 students failed in the examination for the subject of Economics. The marks obtained by the 15 students who passed are as follows:
42, 65, 53, 75, 43, 50, 68, 57, 79, 48, 51, 61, 55, 70, 64. Find the median marks of all 18 students.
13. The mean daily sale of a company is 126.2 units. The sales on 10 days after adopting a new advertising strategy are as follows : 156, 125, 162, 153, 130, 124, 127, 142, 149, 121. Can we say that mean sale has increased by new advertising strategy ?

Section E

Solve the following :

1. The following table shows the number of units of electricity consumption of different families :

No. of units	Below 200	200 - 300	300 - 400	400 - 500	500 and above
No. of families	7	13	24	16	10

Find the median consumption.

2. The weekly profit and loss information of a vendor is available as follows. Find the modal profit.

Profit (thousand ₹)	-2 - 0	0 - 2	2 - 4	4 - 6	6 - 8	8 - 10
No. of weeks	4	8	14	6	2	1

3. The number of bags of wheat sold in a grocer's shop each day are shown in the following table :

No. of bags	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 and above
No. of days	9	17	32	24	10	5	3

Find Q_1 and D_4 for the number of bags sold.

4. The heights of students of a college are given in the following table. Find the mean height of the students :

Height (cm)	150 - 155	155 - 160	160 - 165	165 - 170	170 - 175	175 - 180	180 - 185
No. of students	8	10	20	17	15	4	1

5. The monthly income (in thousand ₹) of 130 persons living in a certain area is as follows :

Income (thousand ₹)	Less than 4	4 - 8	8 - 12	12 - 20	20 - 28	28 - 36
No. of persons	6	14	31	35	28	16

Find the median of income.

6. The data about population (in thousands) of 70 villages in a district is given in the following table :

Population (thousands)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No. of villages	6	18	22	15	9

Find the mode for the population using graphical method.

7. The marks obtained by 60 students in an examination are as follows. Find the mean marks of the students.

Marks	0 - 10	10 - 20	20 - 25	25 - 30	30 - 35	35 - 40
No. of students	3	8	20	16	9	4

8. A survey was conducted for 50 employees in an office regarding their computer usage time. The details are shown in the following table :

Time (hours)	5 - 5.5	5.5 - 6	6 - 6.5	6.5 - 7	7 - 7.5	7.5 - 8	8 - 8.5	8.5 - 9
No. of employees	1	3	5	11	15	9	4	2

Find the quartiles Q_1 and Q_3 for the time for the usage of computer.

Section F

Solve the following :

1. The data about marks scored by 55 students from a school are given below.

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of students	4	7	11	14	9	7	3

- (i) If 30% students failed in the examination, what are the passing marks ?
 (ii) If top 5% students are to be selected for scholarship, find the lowest marks among them.

2. Two brands of tyres are to be compared for their mean life. The following data are available.

Life (thousand km.)	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40
No. of tyres of brand A	4	7	10	5	3	1
No. of tyres of brand B	5	8	15	9	6	2

On the basis of mean, which brand of tyres is better ?

3. The distribution of sale of cars of a company on different days is as follows. Find the mode for the number of cars sold using an appropriate formula.

No. of cars	0 - 10	10 - 15	15 - 20	24	26	28
No. of days	8	14	16	11	4	2

4. The wheat crop grown per acre by farmers in different parts of a state is given below :

Wheat crop per acre (quintals)	20 - 25	25 - 30	30 - 40	40 - 50	50 - 60
No. of farmers	12	23	45	29	7

Find the mean and median for the wheat crop per acre.

5. The distribution of age of 150 spectators in a theatre is as follows :

Age (years)	15 - 20	20 - 25	25 - 30	30 - 40	40 - 50	50 - 60	60 - 80
No. of spectators	6	13	19	52	34	18	8

Find the mode for age of spectators using graphical method.

6. A producer believes that the mode of his daily production is 70. The distribution of production from the data obtained after making some changes in the design of the produced units is as follows :

No. of units	60 - 64	65 - 69	70 - 74	75 - 79	80 - 84	85 - 89	90 - 94
No. of days	5	7	10	8	5	3	2

Is there any change in the mode of the number of items produced ?

7. The data for sales of oil tins of two companies sold in a shop are as follows, which show the sales of 40 days.

No. of oil tins		2 - 5	6 - 9	10 - 13	14 - 17	18 - 21	22 - 25
No. of days	Company X	1	3	17	9	6	4
	Company Y	5	9	20	3	2	1

If median is used to compare the sales, which company can be said to have higher sale ?

8. The distribution of age (in complete years) at the time of marriage of 50 married men is as follows :

Age (years)	21 - 23	24 - 26	27 - 29	30 - 32	33 - 35
No. of men	6	21	15	6	2

Find the mode of their age at the time of marriage using graphical method.



C. G. Khatri
(1931 - 1989)

Prof. C.G. Khatri obtained his Ph.D. degree (1960) in Statistics from the MS University of Baroda. He was a Professor and head of the department of Statistics, Gujarat University, Ahmedabad.

Dr. Khatri did original work on multivariate distribution theory, matrix algebra, especially on g-inverses, linear models, estimation of variance components and location parameters in linear models, design of experiments, characterization of distributions and optimality of certain functions of matrix arguments.

He has authored or co-authored several books and about two hundred research publications in prestigious journals.

•

“ Uncontrolled variation is the enemy of quality.”

– Edward Deming

4

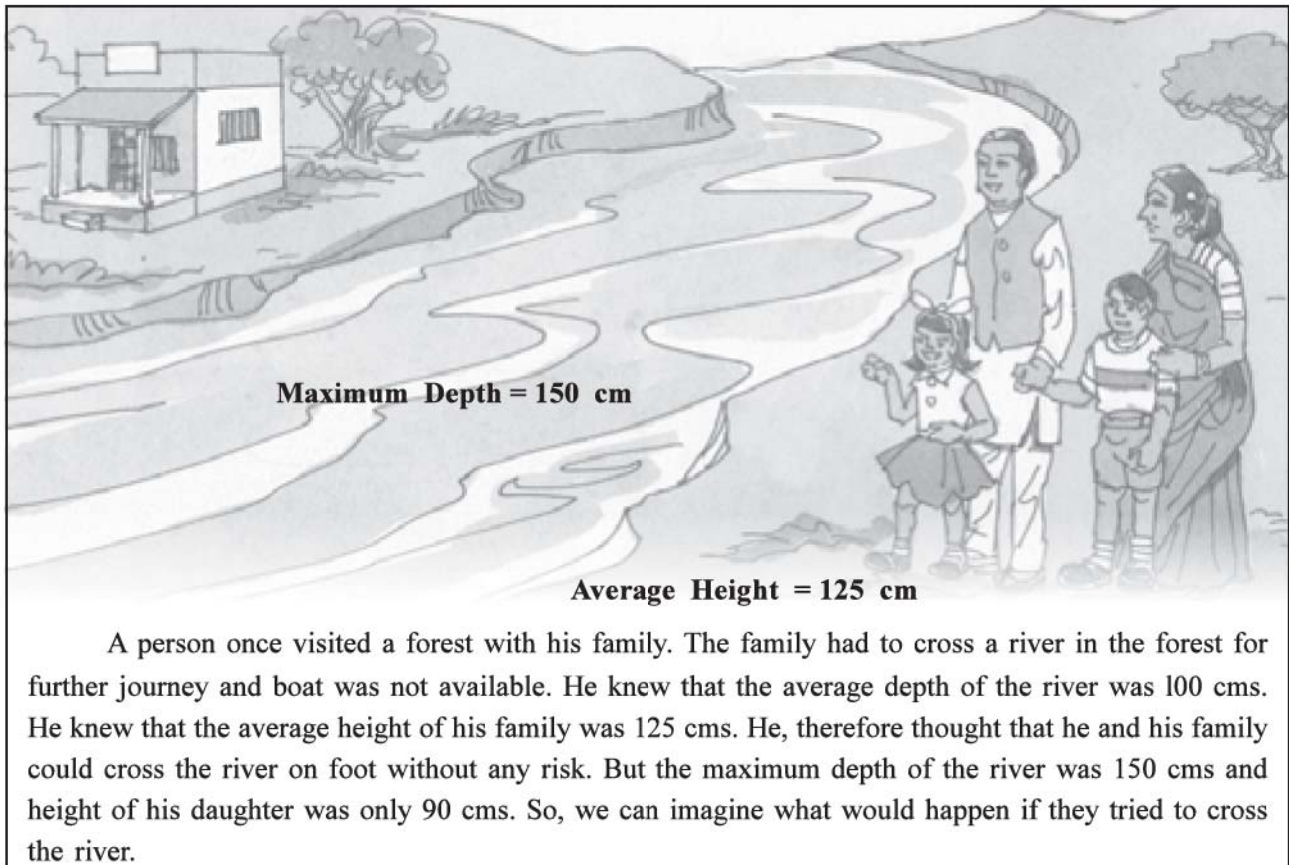
Measures of Dispersion

Contents :

- 4.1 Meaning and characteristics of dispersion**
- 4.2 Concept of absolute and relative measures of dispersion**
- 4.3 Measures of dispersion: Absolute and Relative Measures**
 - 4.3.1 Range : Meaning, advantages and disadvantages
 - 4.3.2 Quartile Deviation : Meaning, advantages and disadvantages
 - 4.3.3 Average Deviation : Meaning, advantages and disadvantages
 - 4.3.4 Standard Deviation : Meaning, advantages and disadvantages
- 4.4 Combined Standard Deviation : Meaning**

4.1 Meaning of Dispersion

We have studied classification, tabulation and the measures of central tendency of the collected data in the preceding three chapters. As we know now, any measures of central tendency or averages of data represent summary or central value of the data. It may happen that some observations of the data are very near to the central value whereas some are very far. So, it is useful to know how the observations are scattered from the central values of the population. Though measures of central tendency are very useful in statistical analysis, only these measures are not sufficient. Let us understand it by the following figure and its details.



Thus, it is clear that without the knowledge of variability, only average (measures of central tendency) does not serve the purpose.

Similarly, let us take another example of average income of the people in a country, popularly known as per capita income. It is one of the key indicators, showing economic condition of the people of that country. But it can be easily understood that it does not throw any light on the distribution of income among various groups of people of the country. So, it is not possible to know inequality of income between the rich and the poor by just having average income (per capita income).

So, to study the data, we should know various characteristics of it. The measures of central tendency tell us only a part of what we need to know, but we should measure the spread or variability of the data for its better understanding. The measure of variability, along with the measures of central tendency gives such information. In this chapter, we shall consider the different measures of variation based on internal variation of the observations and the scattering of the observations of data from the mean as measure of average.

We have experienced that two or more groups having identical average may differ from each other in certain aspects. The spread or scattering of the observations from the average and the internal variations among individual observations could be different. Thus, instead of comparing groups only on the basis of averages, it is advisable to consider the variation within observations of each group. We now illustrate this fact by an example.

Suppose a financial analyst wants to study the performance of three companies A, B and C. He gets the following information of the profits of these three companies for last 5 years :

Year	1	2	3	4	5	Total
Profit of A (lakhs ₹)	30	30	30	30	30	150
Profit of B (lakhs ₹)	15	30	30	30	45	150
Profit of C (lakhs ₹)	-5	30	70	30	25	150

Now it is very obvious that financial analyst will first try to know the average annual profits of these three companies and then will try to know the fluctuations in them.

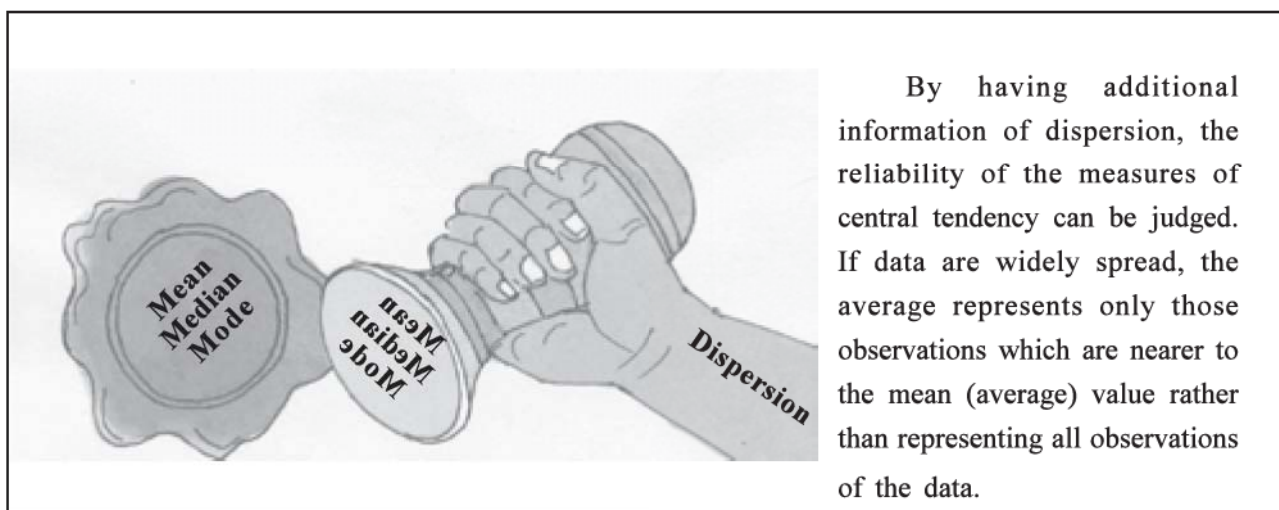
It is very clear from the data of profits of the three companies that for all three companies A, B and C their mean = median = mode = 30 (lakhs ₹). Now, observing individual yearly profits of three companies A, B and C, we can see that the profits of company A remain same in last five years, so the spread in its profit is $30 - 30 = 0$; the spread in the profits of company B is $45 - 15 = 30$ (lakhs ₹), whereas spread of the profits of company C is $70 - (-5) = 75$ (lakhs ₹). Here, there is no variation in the profits for company A, as it remained same for all five years. The yearly profits of company B for last 5 years are nearer to its average of 30 (lakhs ₹); but for company C, its yearly profits are very far from its average profit of 30 (lakhs ₹). Hence though the mean, the median and the mode of these three companies are same, their profits differ significantly in terms of the variation. Thus, analyzing these three companies simply on the basis of equality of their averages may lead to erroneous and misleading conclusion that the three companies are same in terms of their profits.

So, the information about scatter or spread of the observations of a population is necessary not only for study of characteristics of that population but it is also necessary for comparative statistical study of two or more populations.

A measure which shows how far the observations of the data are scattered from the measure of average is termed as **dispersion**.

However, the term dispersion not only gives a general impression about the variability of a population but also a precise measure of this variation. There are several definitions available which are stated by different statisticians. One of the definitions given by **Spiegel** is as follows :

“The degree to which numerical data tend to spread about an average value is called variation or dispersion of the data”.



Desirable Characteristics of a measure of dispersion :

The following are some desirable characteristics of a measure of dispersion.

- (1) The definition of a measure of dispersion should be clear and unambiguous.
- (2) It should be simple to understand and easy to compute.
- (3) It should be based on all observations of the data.
- (4) It should be suitable for further algebraic and statistical calculations.
- (5) It should be a stable measure in the sense that if different samples of equal size are drawn and measures of dispersion are obtained, the values of these measures should be almost same for all samples.
- (6) It should not be unduly affected by very small or very large observations of the data.

4.2 Concept of Absolute and Relative measures of dispersion

Absolute Measure :

A measure of dispersion which is expressed in the same (Statistical) units in which the observations of the data are expressed is called an **absolute measure** of dispersion. For example, if the original data are in kilogram, an absolute measure will also be in kilogram.

An absolute measure is not useful for comparison purpose for variability of two or more sets of data having different units of the measurement. Let us understand this from the following example :

Suppose weights (in kg) and heights (in cms) of students of a school are given. To know, which characteristic is having more variation, we obtain absolute measure of dispersion. Now, the unit of absolute measure of dispersion of weight is in kg, which is different from that of height which is in cms. So, it is not possible to compare their variability using absolute measures.

Relative Measure :

A measure of dispersion which is free from the unit of measurement is called **relative measure** of dispersion. The variability of two or more sets of data having different units of measurement can be compared only by relative measure of dispersion.

Generally, a measure of relative dispersion is the ratio of a measure of an absolute dispersion of observations of a data set to an appropriate average of the observations. The relative measure of dispersion is known as coefficient of dispersion as it is independent of units of measurement of the observations of data.

4.3 Measures of Dispersion

We shall study the following absolute and relative measures of dispersion :

- | | |
|--------------------|------------------------|
| (1) Range | (2) Quartile Deviation |
| (3) Mean Deviation | (4) Standard Deviation |

From the above measures, range and quartile deviation are known as positional measures of dispersion as these measures depend on the positions of observations of the data arranged in increasing order of their magnitudes, while the mean deviation and standard deviation are known as the measures of dispersion representing the summary of deviations of observations from the measure of central tendency.

4.3.1 Range

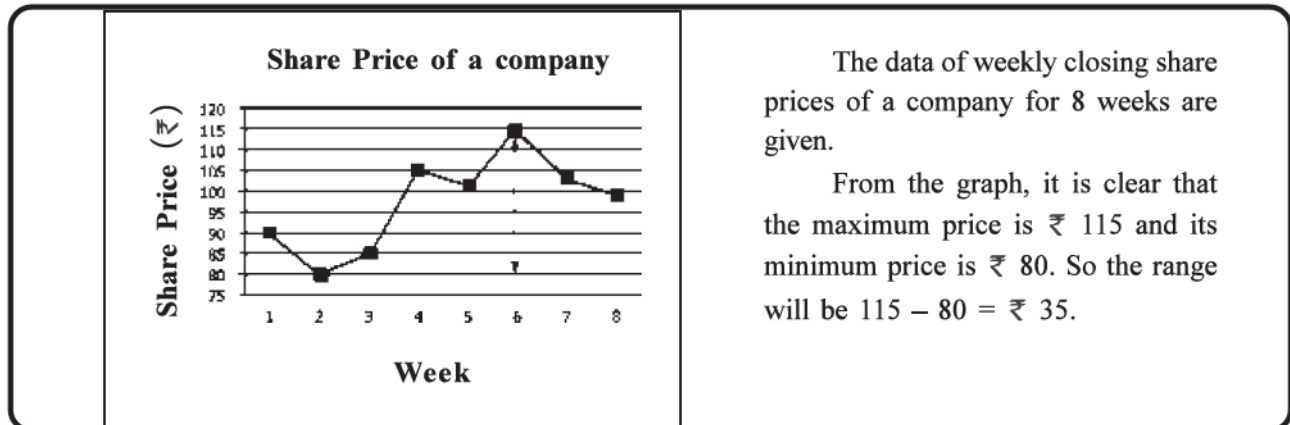
The difference between the highest and lowest observation of the data is called the Range and it is denoted by the symbol R .

$$\therefore \text{Range } R = x_H - x_L$$

where x_H = the highest observation

x_L = the lowest observation

Range R is an absolute measure of dispersion having same unit of measurement as that of the observations.



It is obvious from the definition of range that frequency does not play any role in determining the range, even for grouped data. For any grouped frequency distribution, the range can be obtained as the difference between the upper limit of the highest class interval and the lower limit of the lowest class interval.

If we divide the range R of the data by the sum $x_H + x_L$, we get the relative range.

$$\therefore \text{Relative Range} = \frac{R}{x_H + x_L} = \frac{x_H - x_L}{x_H + x_L}$$

The relative range is also known as **coefficient of range**. It is free from the unit of measurement.

If the coefficient of range for a population is small, then it can be said that variability is less in the observations of the population i.e. the values of the observations are not far from each other. But if the coefficient of range is high then it can be said that variability is more in the observations of the population, i.e. the values of the observations are very far from each other.

Illustration 1 : The runs scored by a batsman in his last 10 innings of cricket matches are 48, 75, 37, 52, 93, 81, 25, 72, 18 and 60. Find the range and the coefficient of range of his runs.

Here, $x_H = 93$, $x_L = 18$

Hence Range = $x_H - x_L = 93 - 18 = 75$

$\therefore R = 75$ runs

Thus, the range of runs scored in last ten innings is 75 runs.

$$\begin{aligned} \text{Coefficient of range} &= \frac{R}{x_H + x_L} \\ &= \frac{75}{93+18} = \frac{75}{111} = 0.6757 \end{aligned}$$

\therefore Coefficient of range ≈ 0.68

Thus, the coefficient of range of runs is 0.68.

Illustration 2 : From the following information of monthly salary (in ₹) of workers of a factory, find the range and the coefficient of range of the salary.

Monthly Salary (₹)	3500	4000	5000	7500	10,000	12,000
No. of workers	3	21	30	19	6	5

The variable (monthly salary) of the frequency distribution is discrete and from the observations of salary it is clear that $x_H = 12,000$ and $x_L = 3500$

$$\begin{aligned} \text{Range} &= x_H - x_L \\ &= 12,000 - 3500 \\ &= 8500 \end{aligned}$$

$$\therefore R = ₹ 8500$$

Thus, the range of the monthly salary of workers is ₹ 8500.

$$\begin{aligned} \text{Coefficient of range} &= \frac{R}{x_H + x_L} \\ &= \frac{8500}{12000 + 3500} \\ &= \frac{8500}{15500} \\ &= 0.5484 \end{aligned}$$

$$\therefore \text{Coefficient of range} \approx 0.55$$

Thus, the coefficient of range of the monthly salary of workers is 0.55.

Illustration 3 : The items produced in a factory are packed into different boxes according to their weight. Using the following information, find the range and the relative range of weight of boxes :

Weight (kg)	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35
No. of boxes	8	15	26	47	4

The given frequency distribution is continuous. The upper limit of the last class and lower limit of the first class will be the highest and lowest observations of the data respectively.

$$\text{Hence } x_H = 35 \text{ and } x_L = 10$$

$$\begin{aligned} \text{Range} &= x_H - x_L \\ &= 35 - 10 \\ &= 25 \end{aligned}$$

$$\therefore R = 25 \text{ kgs.}$$

Thus, range of the weight of the boxes is 25 kgs.

$$\begin{aligned} \text{Relative range} &= \frac{R}{x_H + x_L} \\ &= \frac{25}{35 + 10} \\ &= \frac{25}{45} \\ &= 0.5556 \end{aligned}$$

$$\therefore \text{Relative range} \approx 0.56$$

Thus, relative range of the weight of the boxes is 0.56.

Illustration 4 : Find the range and the coefficient of range of the marks from the following frequency distribution of marks scored by 50 students of a school in a certain examination.

Marks	50 - 59	60 - 69	70 - 79	80 - 89	90 - 99
No. of Students	2	15	23	6	4

The upper limit of the last class is 99 and the lower limit of the first class is 50.

Hence, $x_H = 99$ and $x_L = 50$

$$\text{Range} = x_H - x_L = 99 - 50 = 49$$

$\therefore R = 49$ Marks

Thus, range of the marks of students is 49 marks.

$$\begin{aligned} \text{Coefficient of range} &= \frac{R}{x_H + x_L} \\ &= \frac{49}{99 + 50} \\ &= \frac{49}{149} \\ &= 0.3289 \end{aligned}$$

\therefore Coefficient of range ≈ 0.33

Thus, coefficient of range of marks obtained by students is 0.33.

Advantages and Disadvantages of Range

Advantages :

- (1) The range is very clearly defined.
- (2) Its computation is simple.
- (3) Range is a useful measure especially when variability among the observations of the data is less.

Disadvantages :

- (1) All observations of the data are not used in the computation of range.
- (2) Range is very sensitive about sampling fluctuations.
- (3) It is not a suitable measure for algebraic operations.
- (4) It cannot be calculated for the frequency distribution having open-ended classes.

Note :

Range is useful in statistical quality control for the construction of control charts which measure the variability within the samples taken from the production. If variation is not very large then range is useful in measuring variation in money rates, exchange rates, share prices etc. In our day-to-day problems like 'daily sales in a supermarket', 'temperature of a city', 'expense of petrol by two wheeler or car', etc. are generally expressed in the form of the interval in which it lies, and range of the data can be known from it.

Activity

Collect the information about height and weight of boys and girls whose age is from 15 to 25 years and obtain the interval of height and weight from it and find range. Find the relative range for height and weight and compare them.

EXERCISE 4.1

- The following data refer to the heights in cms. of 10 students of a class. Find range and coefficient of range of height of the students :
162, 145, 170, 181, 167, 151, 175, 185, 169, 156
- A bus company has 77 buses for travelling in the city. The information of number of passengers in bus on a particular day at a particular time is given below. Find the range and coefficient of range of number of passengers.

No. of Passengers	2	7	10	18	25	30	37
No. of Buses	1	4	11	17	23	16	5

- Using the following frequency distribution of marks of students of a school, find range and relative range of the marks.

Marks	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
No. of Students	8	20	25	60	45	10

- The frequency distribution of daily income (in thousand ₹) of 80 shops of an area is as follows. Find the absolute and the relative measure of range of daily income from it.

Daily income (thousand ₹)	5-9	10-14	15-19	20-24	25-29	30-34
No. of shops	11	20	17	13	12	7

*

4.3.2 Quartile Deviation

We know that only extreme observations i.e. the largest and the smallest observations are used in the computation of range. Similarly by using positional measures known as first quartile Q_1 and third quartile Q_3 , another measure of dispersion namely quartile deviation is obtained. The measure of variation defined by using the middle 50% of the observations arranged in increasing order of magnitudes is called the **quartile deviation**.

The measure of quartile deviation is obtained by dividing the difference between Q_3 and Q_1 by 2. Symbolically, it is denoted by Q_d . We can write the formula of quartile deviation as follows.

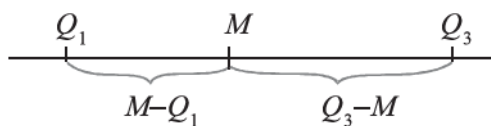
$$Q_d = \frac{Q_3 - Q_1}{2}$$

The quartile deviation is also known as **semi-inter-quartile range**.

Additional Information for Understanding

$(Q_3 - Q_1)$ is called inter-quartile range, but mostly it is reduced to the semi-inter-quartile range which is nothing but the mid-point of inter-quartile range.

Quartile deviation shows the average value by which two quartile (i.e. Q_1 and Q_3) differ from the median. It is clear from the following figure.



$$\text{Quartile Deviation} = \frac{(Q_3 - M) + (M - Q_1)}{2} = \frac{Q_3 - Q_1}{2}$$

If Q_d is divided by the mean of Q_1 and Q_3 we get the relative measure of quartile deviation.

$$\therefore \text{Relative Quartile Deviation} = \frac{(Q_3 - Q_1)/2}{(Q_3 + Q_1)/2} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

The relative quartile deviation is also known as **coefficient of quartile deviation**. Note that Q_d has unit of measurement in which the observations are expressed but coefficient of quartile deviation is independent of such unit i.e coefficient of quartile deviation is unit free measure.

Illustration 5 : A bus operator gets the following number of passengers in 10 trips on a day. Find the quartile deviation and the coefficient of quartile deviation of number of passengers from the following data :

19, 25, 35, 10, 24, 8, 12, 5, 20, 30

Arranging the observations of given data in increasing order, we get

5, 8, 10, 12, 19, 20, 24, 25, 30, 35

Here, $n = 10$, $\frac{n-1}{4} = 2.75$ and $3\left(\frac{n+1}{4}\right) = 8.25$

Here, Quartile Q_1 = Value of the $\left\lceil \frac{n+1}{4} \right\rceil$ th observation
 = Value of the 2.75th observation

$$\begin{aligned} \text{Hence, } Q_1 &= 8 + 0.75(10 - 8) \\ &= 8 + 1.5 \end{aligned}$$

$$\therefore Q_1 = 9.5 \text{ passengers}$$

Third Quartile Q_3 = Value of the $3\left\lceil \frac{n+1}{4} \right\rceil$ th observation
 = Value of the 8.25th observation

$$\begin{aligned} \text{Hence, } Q_3 &= 25 + 0.25(30 - 25) \\ &= 25 + 1.25 \end{aligned}$$

$$\therefore Q_3 = 26.25 \text{ passengers}$$

$$\begin{aligned} \text{Quartile deviation } Q_d &= \frac{Q_3 - Q_1}{2} \\ &= \frac{26.25 - 9.5}{2} \\ &= 8.38 \end{aligned}$$

$$\therefore Q_d = 8.38 \text{ passengers}$$

The quartile deviation of the passengers is 8.38 passengers.

$$\begin{aligned} \text{Coefficient of quartile deviation} &= \frac{Q_3 - Q_1}{Q_3 + Q_1} \\ &= \frac{16.75}{26.25 + 9.5} \\ &= \frac{16.75}{35.75} \\ &= 0.4685 \end{aligned}$$

$$\therefore \text{Coefficient of quartile deviation} = 0.47$$

The coefficient of quartile deviation for the passengers is 0.47.

Illustration 6 : The information regarding time (in minutes) taken to solve a puzzle by 50 students is given below. Find the quartile deviation and the coefficient of quartile deviation of the time taken to solve a puzzle by the children from it.

Times (minutes)	2	4	6	8	10
No. of children	3	12	18	12	5

Time (minutes) x	No. of students f	Cumulative frequency cf
2	3	3
4	12	15
6	18	33
8	12	45
10	5	50
Total	$n = 50$	

Here, $n = 50$, $\frac{n+1}{4} = 12.75$, $3\left(\frac{n+1}{4}\right) = 38.25$

First Quartile Q_1 = Value of the $\left(\frac{n+1}{4}\right)$ th observation
 = Value of the 12.75th observation

$\therefore Q_1 = 4$ minutes

Third Quartile Q_3 = Value of the $3\left(\frac{n+1}{4}\right)$ th observation
 = Value of the 38.25 th observation

$\therefore Q_3 = 8$ minutes

Quartile deviation $Q_d = \frac{Q_3 - Q_1}{2}$

$$= \frac{8-4}{2}$$

$$= 2$$

$\therefore Q_d = 2$ minutes

The quartile deviation of the time taken to solve a puzzle by the student is 2 minutes.

Coefficient of quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{4}{8+4}$$

$$= \frac{4}{12}$$

$$= 0.3333$$

\therefore Coefficient of quartile deviation ≈ 0.33

Thus, the coefficient of quartile deviation of the time taken to solve a puzzle is 0.33.

Illustration 7 : Using the following distribution of income of 1000 persons of a city, calculate the quartile deviation and coefficient of quartile deviation of income of the persons :

Income (thousand ₹)	Less than 50	50 - 70	70 - 90	90 - 110	110 - 130	130 - 150	above 150
No. of persons	54	100	140	300	230	125	51

Income (Thousand ₹)	No. of Persons f	Cumulative frequency cf
Less than 50	54	54
50 - 70	100	154
70 - 90	140	294
90 - 110	300	594
110 - 130	230	824
130 - 150	125	949
above 150	51	1000
Total	$n = 1000$	—

Here, $n = 1000$, $\frac{n}{4} = 250$ and $3\left(\frac{n}{4}\right) = 750$

First Quartile $Q_1 =$ Value of the $\left(\frac{n}{4}\right)$ th observation
 $=$ Value of the 250 th observation

Referring to the column of cumulative frequencies (cf), we see that the 250 th observation lies in the class 70 - 90. Hence, the Q_1 -class is 70 - 90.

$$\text{First Quartile } Q_1 = L + \frac{\frac{n}{4} - cf}{f} \times c$$

Here, $L = 70$, $\frac{n}{4} = 250$, $cf = 154$, $f = 140$, $c = 20$

$$\begin{aligned} \text{Hence, } Q_1 &= 70 + \frac{250 - 154}{140} \times 20 \\ &= 70 + \frac{1920}{140} \\ &= 70 + 13.7143 \\ &= 83.7143 \end{aligned}$$

$\therefore Q_1 = 83.71$ (thousand ₹)

Third deviation $Q_3 =$ Value of the $3\left[\frac{n}{4}\right]$ th observation
 $=$ Value of the 750th observation

Referring to the column of cumulative frequencies (cf), we see that the 750th observation lies in the class 110 - 130. Hence the Q_3 class is 110 - 130.

$$\text{Third deviation } Q_3 = L + \frac{3\left[\frac{n}{4}\right] - cf}{f} \times c$$

Here, $L = 110$, $3\left[\frac{n}{4}\right] = 750$, $cf = 594$, $f = 230$, $c = 20$

$$\begin{aligned} \text{Hence, } Q_3 &= 110 + \frac{750 - 594}{230} \times 20 \\ &= 110 + \frac{3120}{230} \\ &= 110 + 13.5652 \\ &= 123.5652 \end{aligned}$$

$\therefore Q_3 \approx 123.57$ (thousand ₹)

$$\begin{aligned} \text{Quartile Deviation } Q_d &= \frac{Q_3 - Q_1}{2} \\ &= \frac{123.57 - 83.71}{2} \\ &= \frac{39.86}{2} \end{aligned}$$

$\therefore Q_d = 19.93$

Thus, the quartile deviation of income of the persons is 19.93 (thousand ₹).

$$\begin{aligned} \text{Coefficient of quartile deviation} &= \frac{Q_3 - Q_1}{Q_3 + Q_1} \\ &= \frac{123.57 - 83.71}{123.57 + 83.71} \\ &= \frac{39.86}{207.28} \\ &= 0.1923 \end{aligned}$$

\therefore Coefficient of quartile deviation ≈ 0.19

Thus, coefficient of quartile deviation of income of the persons is 0.19.

Advantages and Disadvantages of Quartile Deviation

Advantages :

- (1) The quartile deviation is a clearly defined measure of dispersion.
- (2) Its computation is simple.
- (3) Its value is not affected by unusually small or large observations of the data, as only middle 50% observations are taken into account for the computation of the quartile deviation.
- (4) It is the only measure of dispersion which can be computed if frequency distribution has open-ended classes.

Disadvantages :

- (1) The first 25% and the last 25% observations are ignored for obtaining quartile deviation. Thus, all the observations are not used in the computation of this measure.
- (2) It is not a suitable measure for algebraic operations.
- (3) This is not a stable measure with respect to sampling.
- (4) This measure is less applicable in the advanced study of statistics.

EXERCISE 4.2

1. A shooter missed his target in the last 10 trials by the following distance (mm) during the practice session.
20, 32, 24, 41, 18, 27, 15, 36, 35, 25

Find the quartile deviation and coefficient of quartile deviation of such distance missed by the shooter.

2. Find the quartile deviation and coefficient of quartile deviation of the marks from the following frequency distribution of marks of 43 students of a school.

Marks	10	20	30	40	50	60
No. of students	4	7	15	8	7	2

3. The distribution of amount paid by 200 customers coming for snacks at a restaurant on a particular day is given below :

Amount (₹)	0-50	50-100	100-150	150-200	200-250
No. of customers	25	40	80	30	25

Find the quartile deviation and coefficient of quartile deviation of the amount paid by customers on the day.

*

4.3.3 Average Deviation

The range and quartile deviation are two measures of dispersion in which all observations of the data are not used and they do not show variation of the observations from any average. This limitation can be overcome by a measure of dispersion in which all observations are considered and variation of each observation from the average is also taken into account. These requirements are fulfilled by average deviation (also known as mean deviation). The difference between value of the observation and its mean is known as **deviation**. These deviations can be negative, zero or positive. The sum of all such deviations is zero as seen in chapter 3. To overcome this situation, the absolute values of these deviations are taken. It means that the negative signs of the negative deviations are ignored and the measure based on absolute deviations of the data is defined.

Thus, the mean of the absolute deviations of the observations of a data from its mean is called Mean Deviation and it is denoted by MD .

The relative measure of mean deviation obtained by dividing MD by the mean \bar{x} i.e. $\frac{MD}{\bar{x}}$ is known as coefficient of mean deviation.

$$\therefore \text{Coefficient of mean deviation} = \frac{MD}{\bar{x}}$$

Additional Information for Understanding

Since mean is the most widely used measure of central tendency in statistics, we take deviations only from mean in computation of average deviation. The measure of average deviation can also be obtained by taking absolute deviations from the median or the mode as per the need of the data.

Method of computing Mean Deviation

We shall now discuss the method and the formulae of computing mean deviation for ungrouped (raw) and grouped (classified) data.

Ungrouped (raw) Data

Suppose x_1, x_2, \dots, x_n are observations of ungrouped data and \bar{x} is their mean. First, the absolute differences of each observation x_i from its mean \bar{x} (i.e. $|x_i - \bar{x}|$) are obtained. Now, the sum of all such absolute deviations is obtained and dividing it by the total number of observations, we get mean deviation. Thus, we define the mean deviation MD as

$$MD = \frac{\sum |x_i - \bar{x}|}{n}$$

Where, $\bar{x} = \frac{\sum x_i}{n}$

$|x_i - \bar{x}|$ = absolute value of $(x_i - \bar{x})$ which is the deviation of observation x_i from mean.

n = Total number of observations

Note : For simplicity at the time of computation in examples, we shall not use suffix 'i'. We shall write x in place of x_i , d in place of d_i , and f in place of f_i .

Grouped Data

Discrete Frequency Distribution :

Suppose x_1, x_2, \dots, x_k are the values assumed by the discrete variable x of the discrete frequency distribution with corresponding frequencies f_1, f_2, \dots, f_k . The mean deviation of the discrete frequency distribution is computed by using the following formula.

$$MD = \frac{\sum f_i |x_i - \bar{x}|}{n}$$

Where, x_i = i -th value of the variable x

f_i = frequency corresponding to x_i

$n = \sum f_i$ = total frequency or sum of all frequencies

$$\bar{x} = \frac{\sum f_i x_i}{n} = \text{Mean}$$

Continuous Frequency Distribution

Suppose the mid-values of k classes of the continuous frequency distribution are x_1, x_2, \dots, x_k with corresponding frequencies f_1, f_2, \dots, f_k then the mean deviation of the continuous frequency distribution is computed using the following formula :

$$MD = \frac{\sum f_i |x_i - \bar{x}|}{n}$$

Where x_i = mid value of the i -th class

f_i = frequency of the i -th class

$n = \sum f_i$ = total frequency or sum of all frequencies

$$\bar{x} = \frac{\sum f_i x_i}{n} = \text{mean}$$

After obtaining the mean deviation (MD) by any of the above formula, the coefficient of mean deviation can be obtained as follows :

$$\text{Coefficient of mean deviation} = \frac{MD}{\bar{x}}$$

Illustration 8 : The measurements of weight (in kg) of 8 students of a class of a school are given below:

46, 58, 60, 43, 75, 66, 51, 81

Find the mean deviation and coefficient of mean deviation of weights of the students.

Weight (kg) x	Deviation $x - \bar{x}$ $\bar{x} = 60$	Absolute deviation $ x - \bar{x} $
46	-14	14
58	- 2	2
60	0	0
43	- 17	17
75	15	15
66	6	6
51	- 9	9
81	21	21
Total	480	84

$$\bar{x} = \frac{\sum x}{n} = \frac{480}{8} = 60 \text{ kg}$$

$$\begin{aligned} \text{Mean Deviation } MD &= \frac{\sum |x - \bar{x}|}{n} \\ &= \frac{84}{8} \\ &= 10.5 \end{aligned}$$

$$\therefore MD = 10.5 \text{ kg}$$

Thus, mean deviation of the weights of students is 10.5 kg.

$$\begin{aligned} \text{Coefficient of mean Deviation} &= \frac{MD}{\bar{x}} \\ &= \frac{10.5}{60} \\ &= 0.175 \end{aligned}$$

$$\therefore \text{Coefficient of mean Deviation} = 0.18$$

Thus, coefficient of mean deviation of the students is 0.18.

Illustration 9 : Calculate mean deviation and coefficient of mean deviation of typing time from the following information of time (in minutes) taken to type a report by 32 typists.

Typing time (minutes)	10	11	12	13	14
No. of typists	2	8	12	8	2

Typing time (minutes) x	No. of typist f	fx	Deviation $x - \bar{x}$ $\bar{x} = 12$	Absolute Deviation $ x - \bar{x} $	$f x - \bar{x} $
10	2	20	- 2	2	4
11	8	88	- 1	1	8
12	12	144	0	0	0
13	8	104	1	1	8
14	2	28	2	2	4
Total	32	384	-	-	24

$$\begin{aligned} \text{Mean } \bar{x} &= \frac{\sum fx}{n} \\ &= \frac{384}{32} \\ &= 12 \end{aligned}$$

$$\therefore \bar{x} = 12 \text{ minutes}$$

$$\begin{aligned} \text{Mean deviation } MD &= \frac{\sum f|x-\bar{x}|}{n} \\ &= \frac{24}{32} \\ &= 0.75 \end{aligned}$$

∴ $MD = 0.75$ minutes

Thus, mean deviation of time taken to type a report is 0.75 minutes.

$$\begin{aligned} \text{Coefficient of mean deviation} &= \frac{MD}{\bar{x}} \\ &= \frac{0.75}{12} \\ &= 0.0625 \end{aligned}$$

∴ Coefficient of mean deviation = 0.06

Thus, the coefficient of mean deviation of time taken to type a report is 0.06.

Illustration 10 : 20 children are selected for a district level spelling test. The distribution of their marks out of 50 marks is given below :

Marks	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49
No. of Children	1	3	8	6	2

Find the mean deviation of the marks obtained by the children from this data.

Marks	No. of children f	Mid-value x	fx	$x - \bar{x}$ $\bar{x} = 27$	$x - \bar{x}$	$f x - \bar{x}$
0 - 9	1	4.5	4.5	- 22.5	22.5	22.5
10 - 19	3	14.5	43.5	- 12.5	12.5	37.5
20 - 29	8	24.5	196	- 2.5	2.5	20
30 - 39	6	34.5	207	7.5	7.5	45
40 - 49	2	44.5	89	17.5	17.5	35
Total	20	-	540	-	-	160

$$\begin{aligned} \text{Mean } \bar{x} &= \frac{\sum fx}{n} \\ &= \frac{540}{20} = 27 \end{aligned}$$

∴ $\bar{x} = 27$ Marks

$$\begin{aligned}\text{Mean } MD &= \frac{\sum f|x-\bar{x}|}{n} \\ &= \frac{160}{20} \\ &= 8\end{aligned}$$

$\therefore MD = 8$ marks

Thus, mean deviation of the marks obtained by children is 8 marks.

Illustration 11 : From the following information of fortnight sale of two wheelers by 30 dealers of a city, find the mean deviation of 'number of two wheelers sold'.

No. of two wheelers	12 - 16	17 - 21	22 - 26	27 - 31	32 - 36
No. of dealers	2	3	14	8	3

No. of two wheelers	No. of dealers <i>f</i>	Mid-value <i>x</i>	<i>fx</i>	$x - \bar{x}$ $\bar{x} = 25.17$	$x - \bar{x}$	$f x - \bar{x}$
12 - 16	2	14	28	- 11.17	11.17	22.34
17 - 21	3	19	57	- 6.17	6.17	18.51
22 - 26	14	24	336	- 1.17	1.17	16.38
27 - 31	8	29	232	3.83	3.83	30.64
32 - 36	3	34	102	8.83	8.83	26.49
Total	30	-	755	-	-	114.36

$$\begin{aligned}\text{Mean } \bar{x} &= \frac{\sum fx}{n} \\ &= \frac{755}{30} \\ &= 25.1667\end{aligned}$$

$\therefore \bar{x} \approx 25.17$ two wheelers

$$\begin{aligned}\text{Mean deviation } MD &= \frac{\sum f|x-\bar{x}|}{n} \\ &= \frac{114.36}{30} \\ &= 3.812\end{aligned}$$

$\therefore MD \approx 3.81$ two wheelers

Thus, mean deviation of 'number of two wheelers sold' is 3.81.

Advantages and Disadvantages of Mean Deviation

Advantages :

- (1) The mean deviation is a clearly defined measure of dispersion.
- (2) It is superior measure to the range and the quartile deviation as all the observations are used in its computation.
- (3) Its value is less affected by the extreme values (i.e. unduly the large and the small values) as compared to some other measures of dispersion.
- (4) The absolute value of the difference between observation and the mean is used to measure the distance between an observation from the mean, which is an appropriate measure of distance.

Disadvantages :

- (1) The computation of mean deviation is complicated as compared to the range and the quartile deviation.
- (2) This measure is not suitable for algebraic operations.
- (3) This measure is less used in advanced study of statistics as its definition is based on absolute value.
- (4) It cannot be computed if the frequency distributions has open-ended classes.

Note : Mean deviation is frequently used to study the problems occurring in social sciences in general. It is also useful in Economics to determine economic inequality, in computing the distribution of personal wealth in the community or country, in forecasting weather and business cycles, etc.

EXERCISE 4.3

1. The measurements of height (in centimeters) of 10 soldiers are given below :

160, 175, 158, 165, 170, 166, 173, 176, 163, 168

Find the mean deviation of the heights of the soldiers.

2. The distribution of number of ball bearings used in machines of a factory is given below. Calculate the mean deviation and coefficient of mean deviation of number of ball bearings per machine.

No. of ball bearings	2	4	6	8	10	12	14	16
No. of machines	2	2	4	5	3	2	1	1

3. Find the mean deviation and the coefficient of mean deviation of the distribution of talk time (in minutes) per call :

Talk time (minutes)	3	5	10	12	15
No. of calls	4	7	6	2	1

4. Find the mean deviation and coefficient of mean deviation of number of TV sets using the following frequency distribution of TV sets sold in last 16 months in a town.

No. of TV sets	10 - 30	30 - 50	50 - 70	70 - 90	90 - 110
No. of months	1	4	6	4	1

5. There are 50 boxes containing different number of units of an item in a factory. Find the mean deviation of number of units per box using the following distribution of the units.

No. of units	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of boxes	6	5	8	15	7	6	3

4.3.4 Standard Deviation

We have seen that the definition of mean deviation is based on absolute values of the deviations of observations of the data from the mean. Since the algebraic signs of the deviations are ignored, mean deviation is less used in advanced study of statistics. This limitation of mean deviation is overcome by an important measure of dispersion known as Standard Deviation. Instead of taking the absolute value of deviation of each observation from the mean, the square of the deviation is taken. If the sum of squares of these deviations is divided by the total number of observations, we get an important measure of dispersion known as Variance. It is denoted by s^2 . The positive square root of the variance is called the Standard Deviation. It is denoted by s .

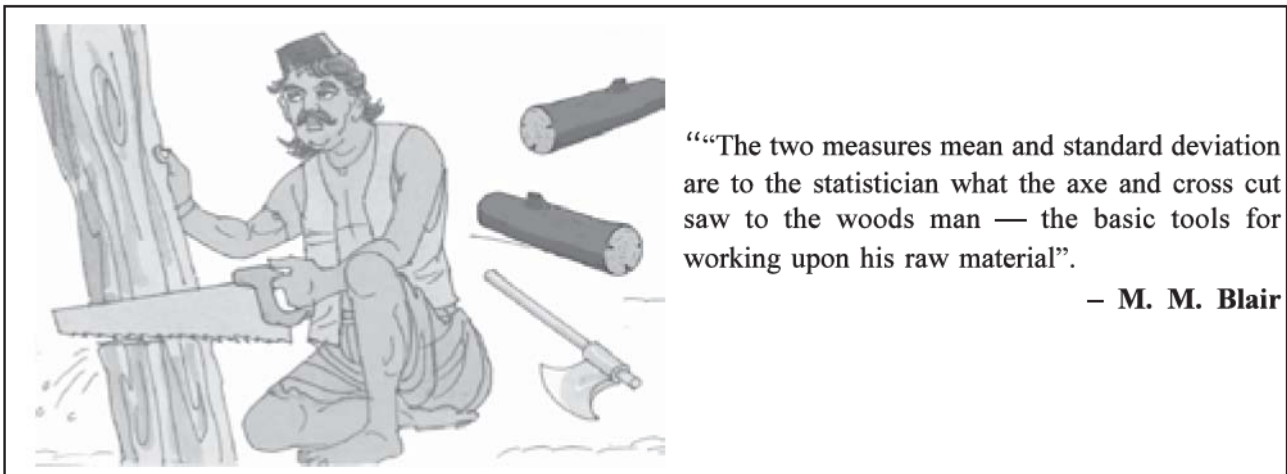
Well known statistician Karl Pearson defined the Standard Deviation as, “Standard Deviation is the positive square root of the mean of the squares of the deviations measured from the mean.”

After mean, standard deviation is another very useful measure which gives information about values of the observations of a population.

Note that the standard deviation is an absolute measure of dispersion. If the standard deviation is divided by the mean of the data, we get its relative measure of dispersion. It is called the **coefficient of standard deviation**.

$$\therefore \text{Coefficient of standard deviation} = \frac{s}{\bar{x}}$$

Note : Among all the measures of dispersion, standard deviation is the most important and widely used measure. The variance and the standard deviation are widely used in experimental research in applied fields such as physics, agricultural science and medicine. They are also useful and an important measure in the study of statistical inference, correlation analysis, sampling and other areas of study.



Computation of Standard Deviation

Computation of Standard Deviation from ungrouped data :

If x_1, x_2, \dots, x_n are the observations of ungrouped data and \bar{x} is its mean, then first the deviation $x_i - \bar{x}$ of the i th observation (where $i = 1, 2, 3, \dots, n$) is obtained as discussed in the definition of standard deviation. The sum of squares of such deviations $\sum(x_i - \bar{x})^2$ is obtained. Dividing the sum by the total number of observations, we get the variance s^2 .

$$\therefore \text{variance } s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

The positive square root of the variance gives the standard deviation and its formula is as follows :

$$\text{Standard Deviation } s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

Illustration 12 : The runs scored by a batsman in his last 7 matches are given below :

52, 58, 40, 60, 54, 38, 48

Find the variance of the runs of the batsman. Also find the standard deviation.

Runs x	$x - \bar{x}$ $\bar{x} = 50$	$(x - \bar{x})^2$
52	2	4
58	8	64
40	- 10	100
60	10	100
54	4	16
38	- 12	144
48	- 2	4
Total	350	432

$$\text{Mean } \bar{x} = \frac{\Sigma x}{n}$$

$$= \frac{350}{7}$$

$$= 50 \text{ runs}$$

$$\text{Variance } s^2 = \frac{\Sigma(x - \bar{x})^2}{n}$$

$$= \frac{432}{7}$$

$$= 61.7143$$

$$\therefore s^2 \approx 61.71 \text{ (runs)}^2$$

$$\text{Standard deviations } s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$$

$$= \sqrt{61.7143}$$

$$= 7.8558$$

$$\therefore s \approx 7.86 \text{ runs}$$

Thus, the standard deviation of runs scored by batsman is 7.86 runs.

Note : The standard deviation is expressed in the units of the observations. We know that variance is square of the standard deviation, hence the unit of variance is square of the unit of the standard deviation.

e.g. : If the unit of the observations is kg., the unit of its standard deviation is also kg, whereas the unit of its variance is (kg.)².

Note : When the value of mean \bar{x} is a fractional number and observations are not numerically large, the computation of s can be made simpler by the following formula for the ungrouped data.

$$s = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

Illustration 13 : The time (in minutes) taken to solve a puzzle by 5 students are 5, 8, 3, 6, 10. Compute the standard deviation of the time taken to solve the puzzle.

	Time (minutes) x	x^2
	5	25
	8	64
	3	9
	6	36
	10	100
Total	32	234

$$\begin{aligned} \text{Mean } \bar{x} &= \frac{\sum x}{n} \\ &= \frac{32}{5} \\ &= 6.4 \text{ minutes} \end{aligned}$$

Since the value of \bar{x} is fractional, we shall use the following alternative formula to obtain the standard deviation.

$$\begin{aligned} s &= \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} \\ &= \sqrt{\frac{234}{5} - (6.4)^2} \\ &= \sqrt{46.8 - 40.96} \\ &= \sqrt{5.84} \\ &= 2.4166 \end{aligned}$$

$$\therefore s \approx 2.42 \text{ minutes}$$

Thus the standard deviation of time taken by students to solve a puzzle is 2.42 minutes.

Short-cut Method :

To make the computation of standard deviation simpler, the following short-cut method can be used.

$$s = \sqrt{\frac{\sum d_i^2}{n} - \left[\frac{\sum d_i}{n}\right]^2}$$

where $d_i = x_i - A$

A = Assumed mean

n = total number of observations

Illustration 14 : The following are closing prices (in ₹) of 5 shares :

132, 147, 120, 152, 125

Find the standard deviation by short-cut method.

We take assumed mean $A = 135$

Price (₹) x	$d = x - A$ $A = 135$	d^2
132	- 3	9
147	12	144
120	- 15	225
152	17	289
125	- 10	100
Total	1	767

Standard Deviation

$$\begin{aligned}
 s &= \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} \\
 &= \sqrt{\frac{767}{5} - \left(\frac{1}{5}\right)^2} \\
 &= \sqrt{153.4 - 0.04} \\
 &= \sqrt{153.36} \\
 &= 12.3839
 \end{aligned}$$

$\therefore s = ₹ 12.38$

Thus, the standard deviation of price of share is ₹ 12.38.

Computation of Standard Deviation For Grouped Data

For Discrete Frequency Distribution :

Suppose the values of variable x of a discrete frequency distribution are x_1, x_2, \dots, x_k with frequencies f_1, f_2, \dots, f_k respectively, then the formula for the standard deviation of the frequency distribution is as follows :

$$\begin{aligned}
 s &= \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n}} \\
 &= \sqrt{\frac{\sum f_i x_i^2}{n} - \bar{x}^2}
 \end{aligned}$$

where f_i = frequency of the i -th value x_i

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

$x_i - \bar{x}$ = deviation of x_i from the mean \bar{x}

$n = \sum f_i$ = total number of observations

Short-cut Method :

The standard deviation for the discrete frequency distribution is obtained by short cut method using the following formula :

$$s = \sqrt{\frac{\sum f_i d_i^2}{n} - \left(\frac{\sum f_i d_i}{n}\right)^2}$$

where f_i = frequency of the i value x_i of the variable

A = assumed mean

$d_i = x_i - A$ = deviation of x_i from assumed mean A

$n = \sum f_i$ = total number of observations

Note : The value of assumed mean A can be taken as any one of the observations x_1, x_2, \dots, x_k or any other suitable value.

Illustration 15 : The distribution of number of absent days of 15 students of a class in the month of January is given below. Find the standard deviation and coefficient of standard deviation of their number of absent days.

No. of absent days	0	1	2	3	4
No. of Students	1	3	7	3	1

x	f	fx	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$	fx^2
0	1	0	-2	4	4	0
1	3	3	-1	1	3	3
2	7	14	0	0	0	28
3	3	9	1	1	3	27
4	1	4	2	4	4	16
Total	$n = 15$	30	0	-	14	74

$$\bar{x} = \frac{\sum fx}{n} = \frac{30}{15} = 2 \text{ days}$$

$$s = \sqrt{\frac{\sum f(x - \bar{x})^2}{n}}$$

$$s = \sqrt{\frac{14}{15}}$$

$$s = \sqrt{0.9333}$$

$$= 0.9661$$

$$\therefore s \approx 0.97 \text{ days}$$

The value of standard deviation can be computed by the alternative formula as follows :

$$\begin{aligned}
 s &= \sqrt{\frac{\sum fx^2}{n} - \bar{x}^2} \\
 &= \sqrt{\frac{74}{15} - (2)^2} \\
 &= \sqrt{4.9333 - 4} \\
 &= \sqrt{0.9333} \\
 &= 0.9661
 \end{aligned}$$

$$\therefore s \approx 0.97 \text{ days}$$

Thus, the standard deviation of number of 'absent days' is 0.97 day.

$$\text{Co-efficient of standard deviation } \frac{s}{\bar{x}} = \frac{0.97}{2}$$

$$= 0.485$$

$$\approx 0.49$$

Thus, coefficient of standard deviation of 'absent days' is 0.49

Illustration 16 : The information of number of mobiles phones sold in last 35 days in a mobile shop is given below. Find the co-efficient of standard deviation of number of mobile phones sold. (Use short-cut method).

No. of mobile phones sold	5	6	7	8	9	10
No. of Days	2	5	8	12	7	1

Assumed mean $A = 8$

x	f	$d = x - A$ $A = 8$	fd	fd^2
5	2	- 3	- 6	18
6	5	- 2	- 10	20
7	8	- 1	- 8	8
8	12	0	0	0
9	7	1	7	7
10	1	2	2	4
Total	$n = 35$	-	- 15	57

$$\bar{x} = A + \frac{\sum fd}{n}$$

$$= 8 + \frac{(-15)}{35}$$

$$= 8 - 0.4286$$

$$= 7.5714$$

$$\therefore \bar{x} \approx 7.57 \text{ mobile phones}$$

$$\begin{aligned}
s &= \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \\
&= \sqrt{\frac{57}{35} - \left(\frac{-15}{35}\right)^2} \\
&= \sqrt{1.6286 - 0.1837} \\
&= \sqrt{1.4449} \\
&= 1.2020
\end{aligned}$$

∴ $s \approx 1.20$ mobiles

Thus, the standard deviation of number of mobile phones sold is 1.20.

$$\begin{aligned}
\text{Coefficient of standard deviation} &= \frac{s}{\bar{x}} \\
&= \frac{1.20}{7.57} \\
&= 0.1585
\end{aligned}$$

∴ Coefficient of standard deviation ≈ 0.16

Thus, coefficient of standard deviation of 'number of mobile phones sold' is 0.16.

For Continuous Frequency Distribution :

Suppose the mid-values of k classes of a continuous frequency distribution are x_1, x_2, \dots, x_k and the respective frequencies of k classes are f_1, f_2, \dots, f_k . Then the formula for computing the standard deviation of frequencies distribution is given by

$$\begin{aligned}
s &= \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n}} \\
s &= \sqrt{\frac{\sum f_i x_i^2}{n} - \bar{x}^2}
\end{aligned}$$

Where f_i = frequency of the i th class

x_i = mid-value of the i th class

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

$x_i - \bar{x}$ = deviation of the mid-value x_i from the mean \bar{x}

$n = \sum f_i$ = total number of observations

Short-cut Method :

When a continuous frequency distribution with same class length is given, the following formula is used to compute the standard deviation.

$$s = \sqrt{\frac{\sum f_i d_i^2}{n} - \left(\frac{\sum f_i d_i}{n}\right)^2} \times c$$

Where, x_i = mid-value of the i th class

A = assumed mean

f_i = frequency of the i -th class

c = class length

$$d_i = \frac{x_i - A}{c}$$

$n = \sum f_i$ = total number of observations.

- Note :**
- The value of the assumed mean A can be taken as any of the mid-values or any other suitable value.
 - The value of the standard deviation remains same, if it is obtained by any suitable form of the formula.

Illustration 17 : Calculate the standard deviation from the following frequency distribution of marks obtained by 200 students of a school in an examination :

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of students	5	12	30	45	50	37	21

We do not need the value of the mean \bar{x} as only the standard deviation is to be obtained. In such a situation, generally, short-cut method is preferred.

Here assumed mean $A = 35$ and class length $c = 10$

Marks	No. of students f	Mid-value x	$d = \frac{x - A}{c}$ $A = 35, c = 10$	fd	fd^2
0 - 10	5	5	- 3	- 15	45
10 - 20	12	15	- 2	- 24	48
20 - 30	30	25	- 1	- 30	30
30 - 40	45	35	0	0	0
40 - 50	50	45	1	50	50
50 - 60	37	55	2	74	148
60 - 70	21	65	3	63	189
Total	$n = 200$	-	-	118	510

Standard Deviation

$$\begin{aligned}
 s &= \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times c \\
 &= \sqrt{\frac{510}{200} - \left(\frac{118}{200}\right)^2} \times 10 \\
 &= \sqrt{2.55 - 0.3481} \times 10 \\
 &= \sqrt{2.2019} \times 10 \\
 &= 14.8388
 \end{aligned}$$

$$\therefore s = 14.84 \text{ Marks}$$

Thus, standard deviation of the marks of the students is 14.84 marks.

Illustration 18 : Find the standard deviation of the daily wages from following information of wages (in ₹) of workers of a factory.

Daily wages (₹)	More than 130	More than 150	More than 170	More than 190	More than 210	More than 230
No. of persons	150	142	116	57	14	0

Here, the 'more than' cumulative frequency distribution is given. Converting it into frequency distribution, we get the following frequency distribution.

Daily wages (₹)	130 - 150	150 - 170	170 - 190	190 - 210	210 - 230
No. of persons	150 - 142 = 8	142 - 116 = 26	116 - 57 = 59	57 - 14 = 43	14 - 0 = 14

Daily wages(₹)	No. of persons f	Mid-value x	$d = \frac{x-A}{c}$ $A = 180, c = 20$	fd	fd^2
130 - 150	8	140	- 2	- 16	32
150 - 170	26	160	- 1	- 26	26
170 - 190	59	180	0	0	0
190 - 210	43	200	1	43	43
210 - 230	14	220	2	28	56
Total	$n = 150$	-	-	29	157

Standard Deviation

$$\begin{aligned}
 s &= \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times c \\
 &= \sqrt{\frac{157}{150} - \left(\frac{29}{150}\right)^2} \times 20 \\
 &= \sqrt{1.0467 - (0.1933)^2} \times 20 \\
 &= \sqrt{1.0467 - 0.0374} \times 20 \\
 &= \sqrt{1.0093} \times 20 \\
 &= 20.0928
 \end{aligned}$$

$$\therefore s \approx 20.09 \text{ ₹}$$

Thus standard deviation of the daily wages of workers of the factory is 20.09 ₹

Note : If all observations under the study are same i.e. $x_1 = x_2 = x_3 = \dots x_n = k$; where $k =$ some constant value, then the value of any measure of dispersion is zero.

Exercise 4.4

- The marks obtained by 9 students in a test of 100 marks in Mathematics are given below :
64, 63, 72, 65, 68, 69, 66, 67, 69
Find the standard deviation of marks obtained by the students.
- The numbers of cars coming for service in five service stations of a company on a particular day are 7, 3, 11, 8, 9. Calculate the standard deviation of number of cars coming at the service station.
- The following frequency distribution represents the amounts of deposits and the number of depositors in a bank. Find the coefficient of standard deviation of the deposits.

Deposits (thousand ₹)	5	10	15	20	25	30	35
No. of depositors	2	7	11	15	10	4	1

- The information of profits (in lakhs ₹) of 50 firms in the last year is given below. Find the standard deviation of the profit of the firms.

Profit (lakh ₹)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No. of firms	7	6	15	12	10

- Find the standard deviation of age of the persons from the following distribution of 125 persons living in a society. Also find the coefficient of standard deviation.

Age (years)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
No. of persons	15	15	23	22	25	10	5	10

*

Coefficient of Variation :

We have seen that the standard deviation is an absolute measure and it is expressed in terms of unit of the given observations of the data. Therefore, for the comparison of variability of two or more groups, their absolute measures cannot be used. For such a comparison, its relative measure, coefficient of standard deviation $\left(\frac{s}{\bar{x}}\right)$, should be used. Often, the value of coefficient of standard deviation $\left(\frac{s}{\bar{x}}\right)$ comes in fractional form, so Karl Pearson has suggested “Coefficient of Variation” as a relative measure which can be easily understood by common people. The coefficient of variation is obtained by multiplying coefficient of standard deviation by 100.

$$\therefore \text{Coefficient of Variation} = \frac{s}{\bar{x}} \times 100$$

The coefficient of variation is measured in terms of percentage. i.e. coefficient of variation is percentage measure of standard deviation with respect to mean.

It is a very useful measures for comparing the dispersion of two or more data sets. A group of observations which has smaller value of coefficient of variation is said to be more stable and having less dispersion. Such a sequence is also said to be consistent from the point of view of variability. A sequence of observations which has larger value of coefficient of variation is said to be less stable and having more dispersion.

Illustration 19 : From the following data of runs scored by two batsmen A and B in last 10 innings, decide who is more consistent.

Runs by batsman A	25	50	45	30	70	42	36	48	34	60
Runs by batsman B	10	70	50	20	95	55	42	60	48	80

In order to know who is a more consistent batsman in terms of runs, we shall obtain the coefficient of variation of runs for both A and B.

Batsman A

Runs <i>x</i>	$x - \bar{x}$ $\bar{x} = 44$	$(x - \bar{x})^2$
25	- 19	361
50	6	36
45	1	1
30	- 14	196
70	26	676
42	- 2	4
36	- 8	64
48	4	16
34	- 10	100
60	16	256
Total	440	0
		1710

Batsman B

Runs <i>x</i>	$x - \bar{x}$ $\bar{x} = 53$	$(x - \bar{x})^2$
10	- 43	1849
70	17	289
50	- 3	9
20	- 33	1089
95	42	1764
55	2	4
42	- 11	121
60	7	49
48	- 5	25
80	27	729
Total	530	0
		5928

For Batsman A

$$\begin{aligned}\text{Mean } \bar{x} &= \frac{\sum x}{n} \\ &= \frac{440}{10} = 44\end{aligned}$$

$$\therefore \bar{x} = 44 \text{ runs}$$

$$\begin{aligned}\text{Standard deviation } s &= \sqrt{\frac{\sum(x-\bar{x})^2}{n}} \\ &= \sqrt{\frac{1710}{10}} \\ &= \sqrt{171} \\ &= 13.0767\end{aligned}$$

$$\therefore s \approx 13.08 \text{ runs}$$

$$\begin{aligned}\text{Coefficient of Variation} &= \frac{s}{\bar{x}} \times 100 \\ &= \frac{13.08}{44} \times 100 \\ &= \frac{1308}{44} \\ &= 29.7272 \%\end{aligned}$$

$$\therefore \text{Coefficient of Variation} \approx 29.73 \%$$

For Batsman B

$$\begin{aligned}\text{Mean } \bar{x} &= \frac{\sum x}{n} \\ &= \frac{530}{10} = 53\end{aligned}$$

$$\therefore \bar{x} = 53 \text{ runs}$$

$$\begin{aligned}\text{Standard deviation } s &= \sqrt{\frac{\sum(x-\bar{x})^2}{n}} \\ &= \sqrt{\frac{5928}{10}} \\ &= \sqrt{592.8} \\ &= 24.3475\end{aligned}$$

$$\therefore s \approx 24.35 \text{ runs}$$

$$\begin{aligned}\text{Coefficient of Variation} &= \frac{s}{\bar{x}} \times 100 \\ &= \frac{24.35}{53} \times 100 \\ &= \frac{2435}{53} \\ &= 45.9434 \%\end{aligned}$$

$$\therefore \text{Coefficient of Variation} \approx 45.94 \%$$

Since the Coefficient of variation for batsman A is less, batsman A is more consistent.

Additional information for understanding

When the means of runs by batsmen are same or approximately same, then a consistent batsman is a better batsman. But the same cannot be said if the means of two batsmen are different.

Illustration 20 : The following information is available for two workers of a factory :

	Workers A	Workers B
Mean time of completing job (minutes)	30	25
Standard Deviation (minutes)	6	4

Which worker has more relative variation or fluctuation in the time taken to complete the job ?

For the decision, coefficients of variation of both workers are to be compared.

worker A	worker B
$\bar{x} = 30$ minutes, $s = 6$ minute	$\bar{x} = 25$ minutes, $s = 4$ minutes
Coefficient of Variation = $\frac{s}{\bar{x}} \times 100$	Coefficient of Variation = $\frac{s}{\bar{x}} \times 100$
= $\frac{6}{30} \times 100$	= $\frac{4}{25} \times 100$
= 20 %	= 16 %

Since the coefficient of variation of worker A is more, there is more variation in the time taken by worker A to complete the job.

Illustration 21 : The means and standard deviations of heights and weights of 50 students of a class are as follows :

	Weight	Height
Mean	56.2 kg	62.5 inch
Standard Deviation	4.8 kg	9.3 inch

Where do you find more variation, among heights or weights?

The units of weights and heights are different here. So, a relative measure of dispersion must be used for comparison and considering the given data, coefficient of variation is an appropriate measure.

Weight	Height
$\bar{x} = 56.2$ kg $s = 4.8$ kg	$\bar{x} = 62.5$ kg, $s = 9.3$ inch
Coefficient of Variation = $\frac{s}{\bar{x}} \times 100$	Coefficient of Variation = $\frac{s}{\bar{x}} \times 100$
= $\frac{4.8}{56.2} \times 100$	= $\frac{9.3}{62.5} \times 100$
= 8.54 %	= 14.88 %

Since the coefficient of variation is more for the heights, we say that height shows more variation.

Advantages and Disadvantages of Standard Deviation

Advantages :

- (1) Its definition is clear and precise.
- (2) All the observations are used in its computation.
- (3) Standard deviation is the most efficient measure of dispersion among all the measures of dispersion.
- (4) Standard deviation is a suitable measure for algebraic calculations. For example, if the means and standard deviations of two data sets are given, the combined standard deviation of a new data set formed by combining the observations of two given data sets can be obtained. It is not possible to obtain a combined measure in case of other measures of dispersion by such an algebraic manipulation.
- (5) Standard deviation is the most widely used measure of dispersion among all the measures of dispersion.

Disadvantages :

- (1) The computation of standard deviation is more complicated as compared to computation of other measures of dispersion.
- (2) The extreme observations get undue importance in the value this measure.
- (3) It cannot be obtained if the frequency distributions have open-ended classes.

EXERCISE 4.5

1. Price fluctuations of two shares A and B are given below, which type of share has more relative variation in its price ?

Price (₹) share A	321	322	325	322	324	320	323	316	319	318
Price (₹) share B	141	146	130	146	142	145	132	134	132	152

2. The daily salary of administrative staff of two companies yielded the following results :

	Company A	Company B
Mean Salary (₹)	600	2100
Standard Deviation (₹)	30	84

Which company has more stable salary ?

3. The Coefficients of variation of two series are 30% and 25% and their standard deviations are 15 and 9 respectively. Find their means.

*

4.4 Combined Standard Deviation

Suppose we have two groups G_1 and G_2 of the data obtained from population and the following information is obtained for the two groups.

Details	For Group G_1	For Group G_2
No. of observations	n_1	n_2
Mean	\bar{x}_1	\bar{x}_2
Standard Deviation	s_1	s_2

A new group G is obtained by combining the observations of two groups G_1 and G_2 . Then the mean and standard deviation of this combined group are known as combined mean \bar{x}_c and combined standard deviation s_c respectively and their formulae are as follows :

$$\text{Combined Mean } \bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$\text{Combined Standard Deviation } s_c = \sqrt{\frac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}}$$

Where, n_1 = No. of observations in group G_1

n_2 = No. of observations in group G_2

s_1 = Standard deviation of group G_1

s_2 = Standard deviation of group G_2

$d_1 = \bar{x}_1 - \bar{x}_c$

$d_2 = \bar{x}_2 - \bar{x}_c$

Illustration 22 : Five observations in each of two groups G_1 and G_2 are given below :

Group G_1 : 1, 3, 5, 7, 9

Group G_2 : 2, 4, 6, 8, 10

Find the mean and variance of both the groups. Hence obtain the combined standard deviation from it.

For Group G_1

	x	$(x - \bar{x})$ $\bar{x} = 5$	$(x - \bar{x})^2$
	1	-4	16
	3	-2	4
	5	0	0
	7	2	4
	9	4	16
Total	25	0	40

Mean of Group G_1

$$\begin{aligned}\bar{x}_1 &= \frac{\sum x}{n_1} \\ &= \frac{25}{5} \\ &= 5\end{aligned}$$

$$\therefore \bar{x}_1 = 5$$

Variance of Group G_1

$$\begin{aligned}s_1^2 &= \frac{\sum (x - \bar{x}_1)^2}{n_1} \\ &= \frac{40}{5} \\ &= 8\end{aligned}$$

$$\therefore s_1^2 = 8$$

$$\begin{aligned}\text{Combined Mean } \bar{x}_c &= \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \\ &= \frac{5(5) + 5(6)}{5 + 5} \\ &= \frac{25 + 30}{10} \\ &= \frac{55}{10} \\ &= 5.5 \\ \therefore \bar{x}_c &= 5.5\end{aligned}$$

For Group G_2

	x	$(x - \bar{x})$ $\bar{x} = 6$	$(x - \bar{x})^2$
	2	-4	16
	4	-2	4
	6	0	0
	8	2	4
	10	4	16
Total	30	0	40

Mean of Group G_2

$$\begin{aligned}\bar{x}_2 &= \frac{\sum x}{n_2} \\ &= \frac{30}{5} \\ &= 6\end{aligned}$$

$$\therefore \bar{x}_2 = 6$$

Variance of Group G_2

$$\begin{aligned}s_2^2 &= \frac{\sum (x - \bar{x}_2)^2}{n_2} \\ &= \frac{40}{5} \\ &= 8\end{aligned}$$

$$\therefore s_2^2 = 8$$

$$d_1 = \bar{x}_1 - \bar{x}_c = 5 - 5.5 = -0.5$$

$$d_2 = \bar{x}_2 - \bar{x}_c = 6 - 5.5 = 0.5$$

Combined Standard Deviation

$$\begin{aligned} s_c &= \sqrt{\frac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}} \\ &= \sqrt{\frac{5(8 + (-0.5)^2) + 5(8 + (0.5)^2)}{5 + 5}} \\ &= \sqrt{\frac{5(8.25) + 5(8.25)}{10}} \\ &= \sqrt{\frac{41.25 + 41.25}{10}} \\ &= \sqrt{\frac{82.5}{10}} \\ &= \sqrt{8.25} \\ &= 2.8723 \end{aligned}$$

$$\therefore s_c = 2.87$$

Activity

Combine the observations of group G_1 and group G_2 from example 22. So, you will have 10 observations 1, 3, 5, 7, 9, 2, 4, 6, 8, 10. Now find mean and standard deviation of these 10 observations and you will see that the mean and the standard deviation will be same as \bar{x}_c and s_c of the example 22.

Illustration 23 : A factory manufactures certain items in two shifts. The information regarding the time taken by workers to manufacture the items is given below. Using the following information, find the combined standard deviation :

	Shift I	Shift II
No. of workers	60	40
Mean manufacturing time (minutes)	25	20
Standard deviation (minutes)	5	3

We shall first find the combined mean \bar{x}_c

$$\begin{aligned} \bar{x}_c &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \\ &= \frac{60(25) + 40(20)}{60 + 40} \\ &= \frac{1500 + 800}{100} \\ &= \frac{2300}{100} \\ &= 23 \text{ minutes} \end{aligned}$$

Thus, it can be said that the mean time taken by all workers of the factory is 23 minutes.

$$d_1 = \bar{x}_1 - \bar{x}_c = 25 - 23 = 2$$

$$d_2 = \bar{x}_2 - \bar{x}_c = 20 - 23 = -3$$

Combined Standard deviation

$$\begin{aligned}
 s_c &= \sqrt{\frac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}} \\
 &= \sqrt{\frac{60(5^2 + 2^2) + 40(3^2 + (-3)^2)}{60 + 40}} \\
 &= \sqrt{\frac{60(25 + 4) + 40(9 + 9)}{100}} \\
 &= \sqrt{\frac{60(29) + 40(18)}{100}} \\
 &= \sqrt{\frac{1740 + 720}{100}} \\
 &= \sqrt{\frac{2460}{100}} \\
 &= \sqrt{24.6} \\
 &= 4.9598
 \end{aligned}$$

$$\therefore s_c = 4.96 \text{ minutes}$$

Thus the standard deviation of the time taken by all workers of both the shifts is 4.96 minutes.

Additional information for understanding

Mean, median and mode are known as “first order averages” where as measures of dispersion are known as “second order averages.”

EXERCISE 4.6

1. The information regarding marks of the students of two classes of a school is given below. Find the combined standard deviation of the marks obtained by the students.

	Division A	Division B
No. of students	50	60
Mean marks	60	48
Standard deviation	10	12

2. The following information is available for two sections of a factory. Obtain the combined standard deviation of the production time.

	Section A	Section B
No. of workers	10	40
Mean production time per unit (minute)	25	20
Variance	16	25

*

Illustration 24 : The information of number of accidents in 10 days on a particular road is given below. Find the mean and standard deviation of number of accidents per day. Find the percentage of days having the number of accidents lying between the limits $\bar{x} \pm s$.

No. of accidents	1	2	3	4	5
No. of days	2	3	3	1	1

No. of accidents x	No. of days f	fx	fx^2
1	2	2	2
2	3	6	12
3	3	9	27
4	1	4	16
5	1	5	25
Total	$n = 10$	26	82

Mean

$$\begin{aligned}\bar{x} &= \frac{\Sigma fx}{n} \\ &= \frac{26}{10} \\ &= 2.6\end{aligned}$$

$$\therefore \bar{x} = 2.6 \text{ accidents}$$

Standard Deviation

$$\begin{aligned}s &= \sqrt{\frac{\Sigma fx^2}{n} - \bar{x}^2} \\ &= \sqrt{\frac{82}{10} - (2.6)^2} \\ &= \sqrt{8.2 - 6.76} \\ &= \sqrt{1.44} \\ &= 1.2\end{aligned}$$

$$\therefore s = 1.2 \text{ accidents}$$

$$\text{Now, } \bar{x} - s = 2.6 - 1.2 = 1.4 \text{ accidents}$$

$$\bar{x} + s = 2.6 + 1.2 = 3.8 \text{ accidents}$$

It can be seen from the frequency distribution that the number of accidents lying within limits 1.4 and 3.8 are 2 and 3 and the number of days having 2 and 3 accidents are 3 and 3 respectively. Hence the number of days having number of accidents within the limits $\bar{x} - s = 1.4$ and $\bar{x} + s = 3.8$ are $3 + 3 = 6$. Since the total number of days is 10, the percentage of days having accidents within the limits is $\frac{6}{10} \times 100 = 60$

Illustration 25 : The mean number of units produced by 100 workers of a factory per day is 60 units and its standard deviation is 10 units. Later on, it was noticed that two workers have actually produced 30 and 20 units respectively but it was registered as 5 and 45 units respectively. By considering this, find corrected mean and corrected standard deviation of the number of units produced by the workers.

We are given $n = 100$, $\bar{x} = 60$, $s = 10$

$$\bar{x} = \frac{\Sigma x}{n} \qquad s = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$$60 = \frac{\Sigma x}{100} \qquad 10 = \sqrt{\frac{\Sigma x^2}{100} - (60)^2}$$

$$\therefore \Sigma x = 6000 \qquad 100 = \frac{\Sigma x^2}{100} - 3600$$

$$3700 = \frac{\Sigma x^2}{100}$$

$$\therefore \Sigma x^2 = 3,70,000$$

But, these values of Σx and Σx^2 are not correct. Now, replacing wrong number of units by the correct number of units produced by the workers, we get,

$$\text{Corrected } \Sigma x = 6000 - 5 - 45 + 30 + 20 = 6000$$

$$\begin{aligned} \text{Corrected } \Sigma x^2 &= 3,70,000 - (5)^2 - (45)^2 + (30)^2 + (20)^2 \\ &= 3,70,000 - 25 - 2025 + 900 + 400 \\ &= 3,69,250 \end{aligned}$$

$$\therefore \text{Corrected Mean } \bar{x} = \frac{\text{Corrected } \Sigma x}{n}$$

$$= \frac{6000}{100}$$

$$= 60 \text{ units}$$

$$\text{Corrected Standard Deviation } s = \sqrt{\frac{\text{corrected } \Sigma x^2}{n} - (\text{corrected } \bar{x})^2}$$

$$= \sqrt{\frac{369250}{100} - (60)^2}$$

$$= \sqrt{3692.5 - 3600}$$

$$= \sqrt{92.5}$$

$$= 9.6177$$

$$\approx 9.62 \text{ units}$$

Illustration 26 : The following results are obtained on the basis of daily wages (in ₹) paid to workers of two firms A and B :

	Firm A	Firm B
No. of workers	20	30
Mean daily wages (₹)	250	400
Standard deviation of wages (₹)	10	12

Answer the following questions using the above information.

1. Which firm is paying more total daily wages to their workers ?
2. Determine which firm shows more relative variation in wages paid to its workers.
3. Find the combined mean and combined standard deviation for the two firms.

(1) Firm A

$$(n_1 = 20, \bar{x}_1 = 250)$$

$$\begin{aligned} \text{Total daily wages} &= n_1 \bar{x}_1 \\ &= 20 (250) \\ &= 5000 \end{aligned}$$

Firm B

$$(n_2 = 30, \bar{x}_2 = 400)$$

$$\begin{aligned} \text{Total daily wages} &= n_2 \bar{x}_2 \\ &= 30 (400) \\ &= 12000 \end{aligned}$$

Hence, Firm B pays more daily wages.

(2) Firm A

$$\begin{aligned} \text{Coefficient of variation} &= \frac{s_1}{\bar{x}_1} \times 100 \\ &= \frac{10}{250} \times 100 \\ &= 4 \% \end{aligned}$$

Firm B

$$\begin{aligned} \text{Coefficient of variation} &= \frac{s_2}{\bar{x}_2} \times 100 \\ &= \frac{12}{400} \times 100 \\ &= 3 \% \end{aligned}$$

The Coefficient of variation for firm A is more. Hence, there is more variation in the daily wages of firm A.

$$\begin{aligned} \text{(3) Combined Mean } \bar{x}_c &= \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \\ &= \frac{20(250) + 30(400)}{20 + 30} \\ &= \frac{5000 + 12000}{50} \end{aligned}$$

$$\therefore \bar{x}_c = \frac{17000}{50} = ₹ 340$$

Thus, after combining the workers of both the firms, their mean daily wages becomes ₹ 340.

$$d_1 = \bar{x}_1 - \bar{x}_c = 250 - 340 = -90$$

$$d_2 = \bar{x}_2 - \bar{x}_c = 400 - 340 = 60$$

Combined Standard Deviation

$$\begin{aligned} s_c &= \sqrt{\frac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}} \\ &= \sqrt{\frac{20\{(10)^2 + (-90)^2\} + \{30(12)^2 + (60)^2\}}{20 + 30}} \\ &= \sqrt{\frac{20(100 + 8100) + 30(144 + 3600)}{50}} \\ &= \sqrt{\frac{164000 + 112320}{50}} \\ &= \sqrt{\frac{276320}{50}} \\ &= \sqrt{55264} \\ &= 74.3398 \end{aligned}$$

$$\therefore s_c = ₹ 74.34$$

Thus, after combining the workers of both the firms their standard deviation is ₹ 74.34.

Illustration 27 : The information of number of roses on 30 rose plants in a nursery is given below. Find range, coefficient of range, quartile deviation, coefficient of quartile deviation, mean deviation and coefficient of mean deviation of the number of roses from it.

No. of roses	1	3	5	6 - 8	8 - 12	12 - 16	16 - 22
No. of plants	1	2	5	10	8	3	1

No. of roses	No. of plants f	cf	Mid-value x	fx	$ x - \bar{x} $ $\bar{x} = 8.1$	$f x - \bar{x} $
1	1	1	1	1	7.1	7.1
3	2	3	3	6	5.1	10.2
5	5	8	5	25	3.1	15.5
6-8	10	18	7	70	1.1	11
8-12	8	26	10	80	1.9	15.2
12-16	3	29	14	42	5.9	17.7
16-22	1	30	19	19	10.9	10.9
Total	$n = 30$	—	—	243	35.1	87.6

$$\text{Mean } \bar{x} = \frac{\sum fx}{n}$$

$$= \frac{243}{30}$$

$$= 8.1 \text{ roses}$$

$$\text{Here, } x_H = 22 \text{ and } x_L = 1$$

$$\therefore \text{Range} = x_H - x_L$$

$$= 22 - 1$$

$$= 21 \text{ roses}$$

$$\text{Coefficient of range} = \frac{x_H - x_L}{x_H + x_L}$$

$$= \frac{21}{22+1}$$

$$= \frac{21}{23}$$

$$= 0.9130$$

$$\therefore \text{Coefficient of range} \approx 0.91$$

$$Q_1 = \text{Value of the } \left(\frac{n}{4}\right)\text{th observation}$$

$$= \text{Value of the } \left(\frac{30}{4}\right)\text{th observation}$$

$$= \text{Value of the 7.5th observation}$$

Referring to the column of cumulative frequency (cf), we see that the value of the 7.5th observation is 5.

$$\therefore Q_1 = 5 \text{ roses}$$

$$\begin{aligned}
Q_3 &= \text{value of the } 3\left(\frac{n}{4}\right)\text{th observation} \\
&= \text{value of the } 3(7.5)\text{ th observation} \\
&= \text{value of the } 22.5\text{ th observation}
\end{aligned}$$

Referring to the column of cumulative frequency (cf), we see that the value of the 22.5th observation lies in the class 8-12. Hence Q_3 class is 8-12.

$$\text{Now, } Q_3 = L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c$$

$$\text{Here, } L = 8, 3\left(\frac{n}{4}\right) = 22.5, cf = 18, f = 8, c = 4$$

$$\begin{aligned}
\therefore Q_3 &= 8 + \frac{22.5 - 18}{8} \times 4 \\
&= 8 + \frac{4.5}{2} \\
&= 8 + 2.25 \\
&= 10.25 \text{ roses}
\end{aligned}$$

$$\begin{aligned}
\therefore \text{Quartile Deviation } Q_d &= \frac{Q_3 - Q_1}{2} \\
&= \frac{10.25 - 5}{2} \\
&= \frac{5.25}{2} \\
&= 2.625
\end{aligned}$$

$$\therefore Q_d = 2.63 \text{ roses}$$

$$\begin{aligned}
\text{Coefficient of quartile deviation} &= \frac{Q_3 - Q_1}{Q_3 + Q_1} \\
&= \frac{5.25}{10.25 + 5} \\
&= \frac{5.25}{15.25} \\
&= 0.3443
\end{aligned}$$

$$\therefore \text{Coefficient of quartile deviation} = 0.34$$

$$\begin{aligned}
\text{Now, Mean Deviation } MD &= \frac{\sum f |x - \bar{x}|}{n} \\
&= \frac{87.6}{30} \\
&= 2.92
\end{aligned}$$

$$\therefore \text{Mean Deviation } MD = 2.92 \text{ roses}$$

$$\begin{aligned}
\text{Coefficient of mean deviation} &= \frac{MD}{\bar{x}} \\
&= \frac{2.92}{8.1} \\
&= 0.3605
\end{aligned}$$

$$\therefore \text{Coefficient of mean deviation} = 0.36$$

Some Useful Results

Suppose the range, quartile deviation, mean deviation and standard deviation for observations x_1, x_2, \dots, x_n are R_x, Q_{dx}, MD_x and s_x respectively. Now, if each observation x_i (where $i = 1, 2, \dots, n$) is multiplied by a non-zero constant “b” and a constant “a” is added to it, a new set of observations y_1, y_2, \dots, y_n is obtained. i.e. $y_i = bx_i + a$

Then the range, quartile deviation, mean deviation, standard deviation and variance of y can be obtained from the respective measures of x as follows :

Measures	For x	For y
Range	R_x	$R_y = b .R_x$
Quartile Deviation	Q_{dx}	$Q_{dy} = b .Q_{dx}$
Mean Deviation	MD_x	$MD_y = b .MD_x$
Standard Deviation	s_x	$s_y = b .s_x$
Variance	s_x^2	$s_y^2 = b^2.s_x^2$

Note : $|b| = b$ if $b \geq 0$

$|b| = -b$ if $b < 0$

Illustration 28 : The range, quartile deviation mean deviation and standard deviation for a variable x are 10, 2, 3 and 5 respectively. If $y = 5x + 3$ find the range, quartile deviation, mean deviation and standard deviation for the variable y .

Here, For variable x , Range $R_x = 10$, Quartile deviation $Q_{dx} = 2$, Mean deviation $MD_x = 3$, standard deviation $s_x = 5$.

Now, $y = 5x + 3$. Using the results discussed earlier, the measures of dispersion for the variable y are obtained as follows.

Range $R_y = |5| \cdot R_x = 5(10) = 50$

Quartile deviation $Q_{dy} = |5| \cdot Q_{dx} = 5(2) = 10$

Mean deviation $MD_y = |5| \cdot MD_x = 5(3) = 15$

standard deviation $s_y = |5| \cdot s_x = 5(5) = 25$

Illustration 29 : The demand function of a commodity is $d = 15 - 2p$, where $p =$ price (in ₹) per unit and $d =$ demand (units). From the closing price of each month of the last year, it is known that for the price, range is ₹ 5, mean deviation is ₹ 2 and variance is 9 (₹)². Find range, mean deviation and variance of the demand from it.

Here, for price, Range $R_p = 5$ ₹, Mean deviation $MD_p = 2$ ₹ and variance $s_p^2 = 9$ (₹)². Now, the demand function is $d = 15 - 2p$. Using the results discussed earlier, we get the following measures for the demand of the commodity :

Range $R_d = |-2| \cdot R_p = 2(5) = 10$ units

Mean deviation $MD_d = |-2| \cdot MD_p = 2(2) = 4$ units

variance $s_d^2 = (-2)^2 \cdot s_p^2 = 4(9) = 36$ (units)²

Illustration 30 : The range and standard deviation of marks obtained out of 100 in the first test by the students of a school are 80 marks and 20 marks respectively. These marks are divided by 4 for the calculation of the internal marks. Find the range and standard deviation of the marks so obtained.

Here, if marks obtained out of 100 are denoted by x then Range $R_x = 80$ marks and standard deviation $s_x = 20$ marks. Now for the calculation of the internal marks these marks are divided by 4. Let us denote the marks so obtained by y . Thus $y = \frac{x}{4} = \frac{1}{4} x$

So by using the results discussed earlier range and standard deviation of y are obtained as follows :

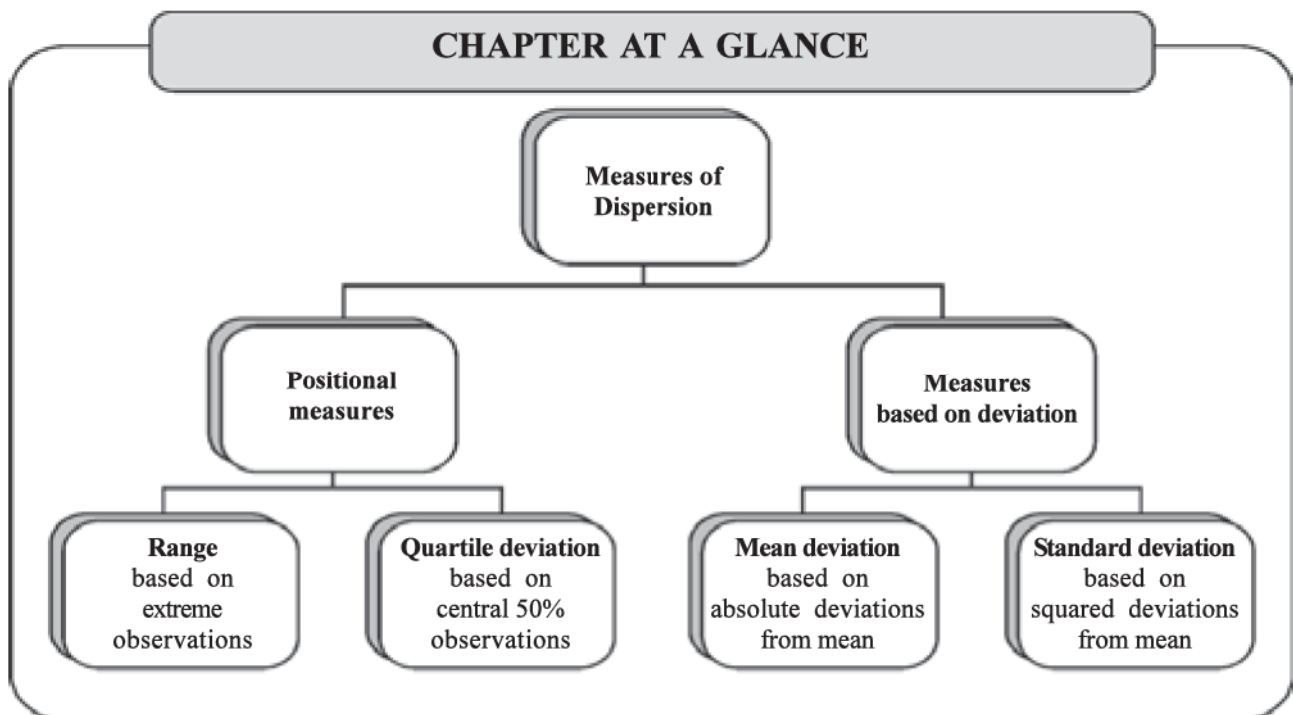
$$\text{Range } R_y = \left| \frac{1}{4} \right| R_x = \frac{1}{4} (80) = 20 \text{ Marks}$$

$$\text{Standard deviation } s_y = \left| \frac{1}{4} \right| s_x = \frac{1}{4} (20) = 5 \text{ Marks}$$

Summary

- **Dispersion or Variation :** It is a measure which shows scatter or spread the observations of a data.
- **Range :** It is a position measure of dispersion obtained by taking the difference between the largest and the smallest value of the data.
- **Quartile Deviation :** It is also a position measure of dispersion. It considers only middle 50% observation. It is also called semi-inter quartile range.
- **Mean Deviation :** It is the mean of the absolute deviations of the observation from its mean.
- **Variance :** The mean of the squares of deviation of all observation from its mean.
- **Standard Deviation :** It is the best measure of dispersion. The standard deviation can be obtained by taking positive square root of the variance s^2 .
- **Relative Measures of dispersion :** A measure free from the unit of the variable under study of dispersion is called relative measure. All relative measures of dispersion used to compare two or more groups in terms of their variability.
- **Coefficient of variation :** It is a percentage relative measure of dispersion based on the standard deviation. The lower value of the coefficient of variation suggests consistency of data.

CHAPTER AT A GLANCE



List of Formulae :

	Measure of Dispersion	Absolute Measure	Relative Measure
1.	Range	$R = x_H - x_L$	Coefficient of Range = $\frac{x_H - x_L}{x_H + x_L}$
2.	Quartile Deviation	$Q_d = \frac{Q_3 - Q_1}{2}$	Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$
3.	Mean Deviation	$MD = \frac{\sum x_i - \bar{x} }{n}$ (For Ungrouped Data) $MD = \frac{\sum f x - \bar{x} }{n}$ (For Grouped Data)	Coefficient of Mean Deviation = $\frac{MD}{\bar{x}}$
4.	Standard Deviation	For Ungrouped Data : $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$ OR $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$ Short-cut Method : $s = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$ For Grouped Data : $s = \sqrt{\frac{\sum f (x - \bar{x})^2}{n}}$ OR $\sqrt{\frac{\sum fx^2}{n} - \bar{x}^2}$ Short-cut Method : When, $d = x - A$ $s = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2}$ When, $d = \frac{x - A}{c}$ $s = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times c$	Coefficient of standard Deviation = $\frac{s}{\bar{x}}$ Coefficient of Variation = $\frac{s}{\bar{x}} \times 100$
5.	Combined Standard Deviation $s_c = \sqrt{\frac{n_1(s_1^2 - d_1^2) + n_2(s_2^2 - d_2^2)}{n_1 + n_2}}$		

EXERCISE 4

section A

For the following multiple choice questions choose the correct option :

- From the following, which is the formula for coefficient of range ?
(a) $x_H - x_L$ (b) $\frac{x_H - x_L}{x_H + x_L}$ (c) $\frac{x_L - x_H}{x_L + x_H}$ (d) $x_L - x_H$
- In which measure of dispersion, the absolute difference of the observation and its mean is considered ?
(a) Mean Deviation (b) Standard Deviation (c) Range (d) Quartile Deviation
- Which of the following measures is a unit free measure ?
(a) Mean Deviation (b) Quartile Deviation (c) Range (d) Coefficient of Variation
- Which measure of dispersion is least affected by the extreme values of the observations?
(a) Range (b) Standard Deviation (c) Quartile Deviation (d) Mean Deviation
- The coefficient of variation of group A is less than the coefficient of variation of group B. Which group is more consistent with respect to variability ?
(a) A (b) B (c) Both (d) None of these
- The weight (in kg.) for 10 students are 53, 47, 60, 55, 71, 65, 61, 68, 63, 70. What is the range of the data ?
(a) 17 kg. (b) 23 kg. (c) 24 kg. (d) 18 kg.
- If the first quartile and the third quartile of a data are 30 and 50 respectively then what is the value of coefficient of quartile deviation ?
(a) 0.25 (b) 50 (c) 4 (d) 20
- What is the value of any measure of dispersion for the observations 5, 5, 5, 5, 5 ?
(a) 1 (b) 5 (c) 0 (d) 25
- If mean of a variable is 10 and the coefficient of variation is 60%. What is the variance of the variable ?
(a) 6 (b) 36 (c) 60 (d) 50
- Suppose the standard deviation of the series $k_1, k_2, k_3, \dots, k_n$ is 5. What will be the standard deviation of the following series ?
(i) $k_1 + 2, k_2 + 2, k_3 + 2, \dots, k_n + 2$
(ii) $3k_1, 3k_2, 3k_3, \dots, 3k_n$
(a) (i) 7 (ii) 3 (b) (i) 5 (ii) 3 (c) (i) 5 (ii) 15 (d) (i) 7 (ii) 15
- The mean and standard deviation for a variable x are 5 and 2 respectively. Now, if $y = 3x + 4$ then what are the mean and the standard deviation of y ?
(a) 19 and 6 (b) 15 and 49 (c) 19 and 10 (d) 15 and 10
- The mean and standard deviation of a set of observations are 45 and 5 respectively. If a constant 5 is added to each observation, what is the coefficient of variation of the new set of the observations ?
(a) 10 % (b) 50 % (c) 11.11 % (d) 900 %

Section B

Give answer in one sentence for the following questions :

1. Define the range.
2. Define the quartile deviation.
3. Which types of measures of dispersion are used for comparing two or more groups in terms of their variability ?
4. Which is the best measure of dispersion ?
5. If the data of heights of 10 students are given in centimeter, what is the unit of its variance ?
6. For a company making pipes, the following information of diameter (in cm) of pipes is obtained. Find the range of the diameter of the pipes.

Diameter (cm)	20 - 40	40 - 60	60 - 80	80 - 100	100 - 120
No. of pipes	15	40	75	20	11

7. The 25th and 75th percentiles of a frequency distribution are 72.18 and 103.99 respectively. Find the quartile deviation.
8. Seven students of a group get 20, 20, 20, 20, 20, 20, 20 marks in a test of 25 marks. What is the standard deviation of their marks.
9. Find the mean deviation for the observations – 1, 0, 4.

Section C

Give answer for the following questions :

1. Define the following :
(i) Mean Deviation (ii) Standard Deviation (iii) Coefficient of Variation.
2. What is meant by the absolute and relative measures of dispersion ?
3. Write names of the absolute measures of dispersion.
4. Which measures of dispersion are based on the deviations of observation from their mean ?
5. Find the range and the coefficient of range for the observation 6, 11, – 3, 0, 8
6. Find the coefficient of quartile deviation for the observations :
8, 15, 2, 11, 20, 3, 5
7. Find the mean deviation for the observations :
3, 8, 1, 7, 6
8. If $\bar{x} = 25$ and the coefficient of variation is 20 %, find the variance.
9. Find the standard deviation for the observations 1, 2, 3, 4, 5.
10. Which of the following factories is more stable with respect to daily production ?

	Factory A	Factory B
Average Daily Production(units)	50	48
Standard deviation (units)	10	12

11. The 25th and the 75th percentiles of a data set are 20 and 36 respectively. Find the coefficient of quartile deviation of the data set.

Section D

Give answer for the following questions :

1. Explain the meaning of dispersion and state different measures of dispersion.
2. State the desirable characteristics of dispersion.
3. Write advantages and disadvantages of the range.
4. Write advantages and disadvantages of the quartile deviation.
5. Write advantages and disadvantages of the mean deviation.
6. Write advantages and disadvantages of the standard deviation.
7. What is standard deviation ? Why is it considered as the best measure of dispersion ?
8. Write a brief note on coefficient of variation.
9. The information of number of flowers on 100 plants of a nursery is given below. Find the quartile deviation of the number of flowers from it.

No. of flowers	11	13	15	17	19	21	23	25
No. of plants	5	8	13	20	22	18	10	4

10. The information of number of goals in 16 matches of hockey tournament is given. Find the mean deviation of number of goals for it.

No. of goals	1	2	3	4	5
No. of matches	1	4	6	4	1

11. In usual notations, $\Sigma d = 25$, $\Sigma d^2 = 272$, $n = 100$ and assumed mean = 4. Find the coefficient of variation.
12. Find the combined standard deviation using the following information :

	Data set A	Data set B
No. of observations	50	60
Mean	113	120
Standard deviation	6	7

13. The sum of 10 observations is 80 and the sum of their squares is 800. Find the coefficient of variation of the observations.

Section E

Solve the following :

1. In a language spelling test of 50 marks, the frequency distribution of marks secured by 30 students is given below. Find the mean deviation of the frequency distribution.

Marks	12 - 16	17 - 21	22 - 26	27 - 31	32 - 36
No. of students	2	3	14	8	3

2. Find the quartile deviation of advertisement expenditure using following frequency distribution of advertising expenditure of 50 companies.

Advertisement cost (thousand ₹)	0-5	5-15	15-30	30-40	40-60	60-100	Total
No. of companies	3	8	15	10	8	6	50

3. The information of runs scored by a batsman in his 100 matches is given below. Find the standard deviation of runs scored by him from it.

Runs	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of matches	10	15	25	25	10	10	5

4. The information of marks obtained by 220 students of a college is given below. Find the quartile deviation of the marks obtained by the students.

Marks	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 or more
No. of students	30	50	64	42	29	5

5. Goals scored by two teams in a Football session are as follows. Which team is more consistent in its game ?

No. of goals scored in a football match	No. of football matches played	
	Team A	Team B
0	15	20
1	10	10
2	7	5
3	5	4
4	3	2
5	2	1

6. For a sequence of 100 observations, the mean and standard deviation are 40 and 10 respectively. In calculating these measures, two observations were taken as 30 and 70 instead of 3 and 27 by mistake. Find the corrected mean and corrected standard deviation.
7. The total cost function for a factory is $y = 10 + 3x$ where $x =$ No. of units produced and $y =$ total cost of producing x units. The range, the quartile deviation, the mean deviation and standard deviation of daily production of the factory are 50, 5, 8 and 10 units respectively. Find the range quartile deviation mean deviation and standard deviation for total cost y from it.

Section F

Solve the following :

1. Find range, coefficient of range, quartile deviation, coefficient of quartile deviation, mean deviation and coefficient of mean deviation from the following data of number of emergency visits of 80 doctors to their patients in a town.

No. of visits	3	5	8	12	17	20	24	30	35
No. of doctors	1	3	7	15	20	13	10	7	4

2. Find the percentage of observations lying within the limits $\bar{x} \pm s$ using the following distribution of credit days taken by the merchants.

Credit days	12	13	14	15	16	17	18	19
No. of merchants	5	10	25	65	45	35	8	7

3. Find an appropriate measure of dispersion from the following data. Also find its relative measure :

Marks	Less than 10	10 - 20	20 - 30	30 - 40	Above 40
No. of students	2	4	10	3	1

4. The information of the salary of 200 employees of company is given below. Find the standard deviation of the salary of the employees.

Salary (thousand ₹)	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50	Less than 60	Less than 70
No. of persons	5	17	47	92	142	179	200

5. The following is a distribution of closing prices (in ₹) of shares of 100 different small scale industries on a certain day. Find the mean deviation of the closing prices of shares.

Price (₹)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
No. of industries	3	8	15	20	25	10	9	6	4

6. The information of daily wages (in ₹) of 230 workers of a factory is given below. Calculate the coefficient of variation from the following data for the daily wages of workers.

Daily wages (₹)	No. of workers
Less than 100	12
Less than 200	30
Less than 300	65
Less than 400	107
Less than 500	157
Less than 600	202
Less than 700	222
Less than 800	230

7. The marks obtained by two students A and B in 10 sets of examinations are given below :

Set	1	2	3	4	5	6	7	8	9	10
Marks of A	44	80	76	48	52	72	68	56	60	64
Marks of B	48	75	54	60	63	69	72	51	57	56

Which student is more consistent in his study ?

8. The following are the distributions of weights (in kg) for the students of two groups A and B. Find the coefficient of variation of each group. Which group has greater relative variation ?

Weight (kg)	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
Group A	7	10	20	18	7
Group B	5	9	21	15	6



Karl Pearson
(1857 - 1936)

Karl Pearson was a major contributor to the early development of statistics. His most famous contribution is the Pearson's chi-square test.

In 1911, he founded the world's first university statistics department at University College, London. He applied statistics to biological problems of heredity and evolution. These papers contain contributions to regression analysis, the correlation coefficient and include the chi-square test of statistical significance (1900). He coined the term 'standard deviation' in 1893. His work was influenced by the work of Edgeworth and in turn influenced the work of Yule. He was a co-founder of the statistical journal Biometrika.



5

Skewness of Frequency Distribution

Contents :

- 5.1 Meaning of Skewness**
- 5.2 Types of Skewness**
- 5.3 Concept of Absolute and Relative Measures of Skewness**
- 5.4 Methods of obtaining Measures of Skewness and Coefficients of Skewness**
 - 5.4.1 Karl Pearson's Method
 - 5.4.2 Bowley's Method
- 5.5 Comparison of two methods of Coefficient of Skewness**

5.1 Meaning of Skewness

Measures of central tendency and dispersion give important information about the population under study. These measures give the information about the values assumed by the units of the population. Moreover, a comparison of two or more populations and an analysis to obtain other information is possible using these measures. We have studied the nature of population observations by these measures.

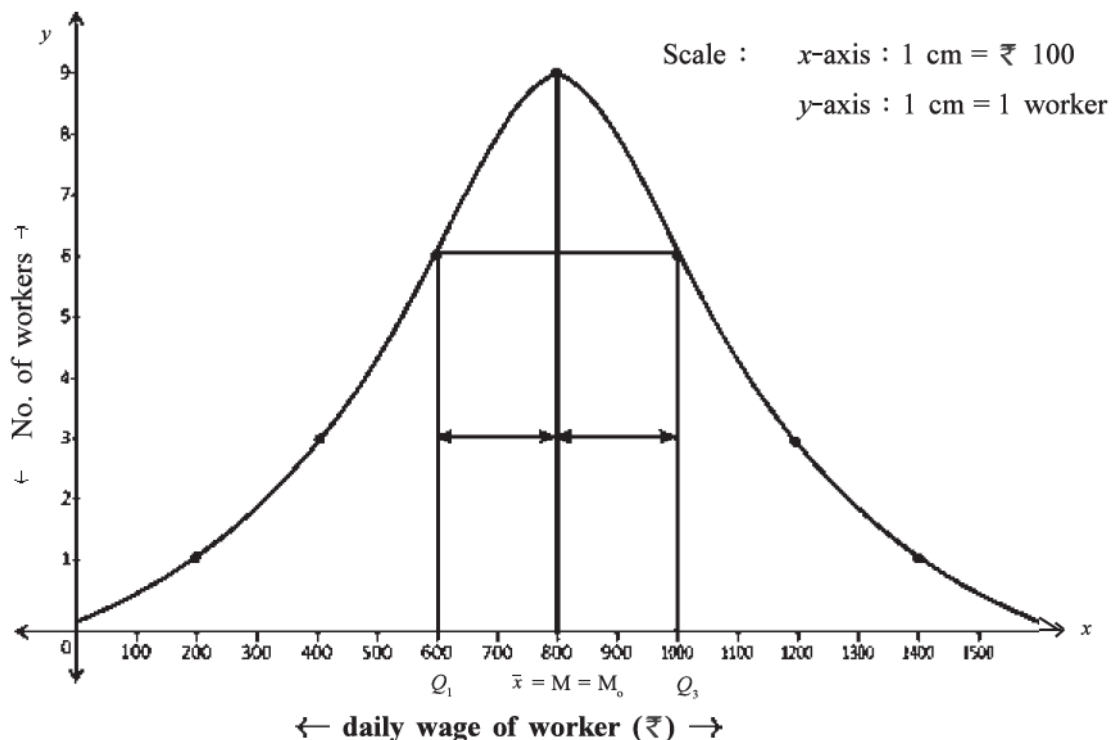
The measures of central tendency and dispersion give us partial information of the central tendency and dispersion of observations around the central tendency measure of raw data. But we cannot get its complete information. To get more information about this, the population is represented by frequency curve. Thus, the direction, shape and form of the frequency curve can be studied from the frequency curve of the frequency distribution. We will study the third important measure called skewness to obtain more information about the population. Let us understand the concept of symmetric frequency distribution before explaining the meaning of skewness.

Symmetric Frequency Distribution

A frequency distribution in which the observations of the population are evenly distributed on both the sides of the mode is called symmetric frequency distribution and its frequency curve is called symmetric frequency curve. Generally, the frequency curve of symmetric distribution is found to be bell-shaped. We shall study the concept of symmetry by the frequency curve and measures of central tendency using the illustration of frequency distribution of daily wages (in ₹) of workers.

Daily wage of worker (₹)	200	400	600	800	1000	1200	1400
No. of workers	1	3	6	9	6	3	1

To ascertain whether the above frequency distribution is symmetric or not, we will draw the frequency curve and find the measures of central tendency.



Thus, the frequency curve of the given distribution is bell-shaped. The observation ₹ 800 has maximum frequency 9. The observations at equal distance on both the sides of ₹ 800 have the same frequencies. Hence, the observations on both the sides of mode of the distribution are equally distributed.

Now we will compute mean, median, mode and quartiles of the frequency distribution and use them to understand the symmetric distribution.

Daily wage of workers (₹) x	200	400	600	800	1000	1200	1400	Total
No. of workers f	1	3	6	9	6	3	1	$n = 29$
fx	200	1200	3600	7200	6000	3600	1400	23,200
cf	1	4	10	19	25	28	29	

Mean $\bar{x} = \frac{\Sigma fx}{n}$
 $= \frac{23200}{29}$
 $= 800$
 $\therefore \bar{x} = ₹ 800$

Median $M =$ Value of the $\left(\frac{n+1}{2}\right)$ th observation
 $=$ Value of the $\left(\frac{29+1}{2}\right)$ th observation
 $=$ Value of the $\left(\frac{30}{2}\right)$ th observation
 $=$ Value of the 15th observation

Referring to *cf* column, we find that the value of the 15th observation is 800.

$\therefore M = 800$

$\therefore M = ₹ 800$

Mode $M_o =$ observation with highest frequency $9 = 800$

$\therefore M_o = ₹ 800$

First quartile $Q_1 =$ Value of the $\left(\frac{n+1}{4}\right)$ th observation
 $=$ Value of the $\left(\frac{29+1}{4}\right)$ th observation
 $=$ Value of the 7.5th observation

Referring to *cf* column, we find that the value of the 7.5th observation is 600

$\therefore Q_1 = 600$

$\therefore Q_1 = ₹ 600$

Third quartile $Q_3 =$ Value of the $3\left(\frac{n+1}{4}\right)$ th observation
 $=$ Value of the $3\left(\frac{29+1}{4}\right)$ th observation
 $=$ Value of the 3(7.5)th observation
 $=$ Value of the 22.5th observation

Referring to *cf* column, we find that the value of the 22.5th observation is 1000.

$\therefore Q_3 = 1000$

$\therefore Q_3 = ₹ 1000$

It is clear from the values of mean, median, mode and quartiles of the distribution as well as the frequency curve of the distribution that,

(1) We have $\bar{x} = M = M_o = ₹ 800$

(2) $(Q_3 - M) = (M - Q_1)$

$\therefore (1000 - 800) = (800 - 600)$

Thus, the quartiles are equidistant from median.

(3) The frequency curve of the frequency distribution is found to be bell-shaped.

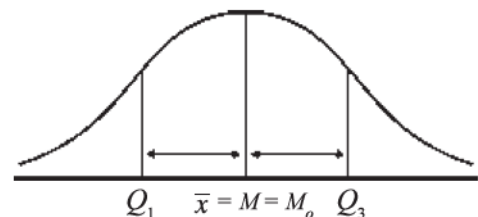
Thus, **the following characteristics are generally observed in a symmetric frequency distribution.**

(i) Mean, median and mode have same value.

That is, $\bar{x} = M = M_o$

(ii) First quartile Q_1 and third quartile Q_3 are at equal distance from median M .

That is, $(Q_3 - M) = (M - Q_1)$



(iii) The frequency curve of the frequency distribution is found to be bell-shaped.

(iv) The frequency of observation at equal distance on both the sides of mode is equally distributed.

If the above characteristics are absent in any given frequency distribution then it is said that the frequency distribution is not symmetric. **The frequency distribution with lack of symmetry is called skewed frequency distribution. Thus, lack of symmetry is called skewness.** The following situations indicate skewness in a frequency distribution :

(1) The values of mean, median and mode are not same.

(2) Quartiles Q_1 and Q_3 are not at equal distances from median $M (= Q_2)$.

$$\text{Thus } (Q_3 - M) \neq (M - Q_1)$$

(3) The right or left tail of the frequency curve is more elongated.

(4) The frequency of observation at equal distance on both sides of mode is not equally distributed.

The above conditions are called tests of skewness as they check whether the frequency distribution is skewed or not. Now we shall study the types of skewness.

5.2 Types of Skewness

There are two types of skewness for frequency distribution: (1) Positive Skewness and (2) Negative Skewness. We understand these two types with diagram and illustration.

(1) Positive Skewness :

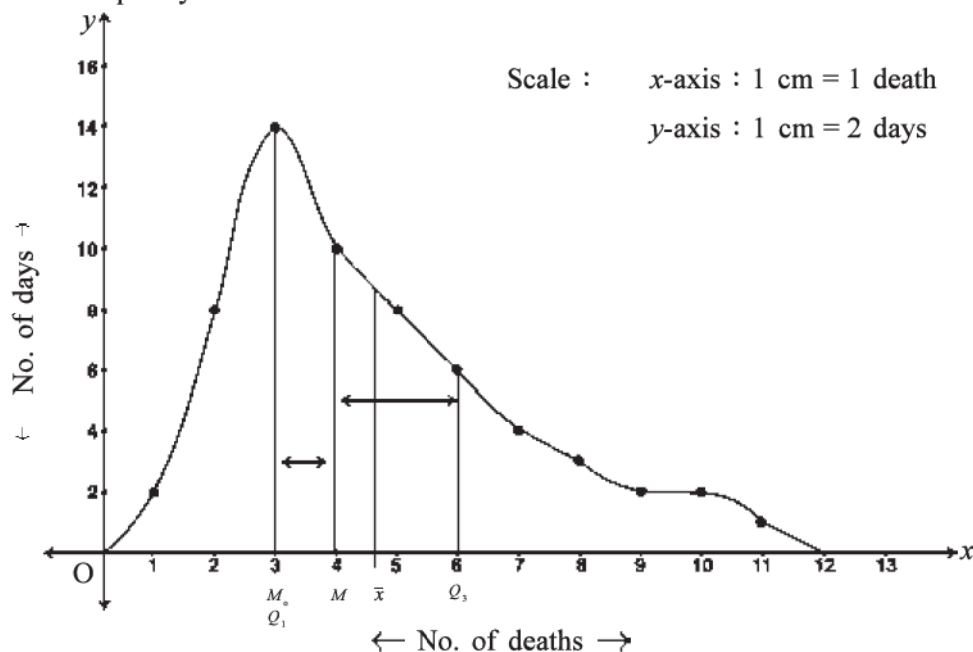
If the right tail of the frequency curve of a distribution is more elongated then it is called positively skewed distribution. The population is said to possess positive skewness due to this characteristic.

Now we take the following illustration to study positive skewness using frequency curve and measures of central tendency measures.

The frequency distribution of number of deaths in a hospital of a city during 60 days is as follows :

No. of deaths	1	2	3	4	5	6	7	8	9	10	11
No. of days	2	8	14	10	8	6	4	3	2	2	1

We will draw the frequency curve of the distribution.



The right tail of the frequency curve is elongated. Less observations are distributed to the left of the observation $x=3$ corresponding to the maximum frequency 14. More observations are distributed to its right side and their frequencies are decreasing gradually.

Now we find the mean, median, median and quartiles of the distribution.

No. of deaths x	No. of days f	fx	cf
1	2	2	2
2	8	16	10
3	14	42	24
4	10	40	34
5	8	40	42
6	6	36	48
7	4	28	52
8	3	24	55
9	2	18	57
10	2	20	59
11	1	11	60
Total	$n = 60$	277	

$$\begin{aligned} \text{Mean } \bar{x} &= \frac{\sum fx}{n} \\ &= \frac{277}{60} \\ &= 4.6166 \end{aligned}$$

$$\therefore \bar{x} = 4.62 \text{ deaths}$$

Mode M_o = Observation with highest frequency $14 = 3$

$$\therefore M_o = 3 \text{ deaths}$$

$$\begin{aligned} \text{Median } M &= \text{Value of the } \left(\frac{n+1}{2}\right)\text{th observation} \\ &= \text{Value of the } \left(\frac{60+1}{2}\right)\text{th observation} \\ &= \text{Value of the } \left(\frac{61}{2}\right)\text{th observation} \\ &= \text{Value of the } 30.5\text{th observation} \end{aligned}$$

Referring to cf column, we find that the value of the 30.5th observation is 4.

$$\therefore M = 4 \text{ deaths}$$

$$\begin{aligned} \text{First quartile } Q_1 &= \text{Value of the } \left(\frac{n+1}{4}\right)\text{th observation} \\ &= \text{Value of the } \left(\frac{60+1}{4}\right)\text{th observation} \\ &= \text{Value of the } \left(\frac{61}{4}\right)\text{th observation} \\ &= \text{Value of the } 15.25\text{th observation} \end{aligned}$$

Referring to cf column, we find that the value of the 15.25th observation is 3.

$$\therefore Q_1 = 3 \text{ deaths}$$

Third quartile Q_3 = Value of the $3\left(\frac{n+1}{4}\right)$ th observation
 = Value of the $3\left(\frac{60+1}{4}\right)$ th observation
 = Value of the 3(15.25)th observation
 = Value of the 45.75th observation

Referring to *cf* column, we find that the value of the 45.75th observation is 6.

$\therefore Q_3 = 6$ deaths

Thus, the following results are obtained for this frequency distribution :

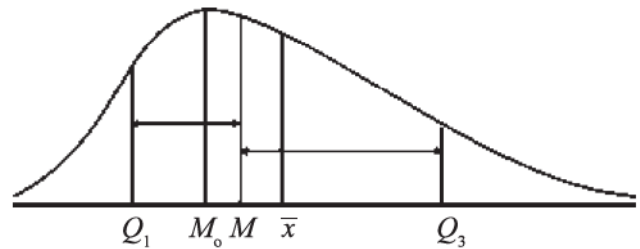
- (1) $\bar{x} = 4.62$, $M = 4$ and $M_o = 3$ Hence, $\bar{x} > M > M_o$
- (2) $Q_3 - M = 6 - 4 = 2$ and $M - Q_1 = 4 - 3 = 1$. Hence $Q_3 - M > M - Q_1$
- (3) The right tail of the frequency curve is more elongated.

Thus the following characteristics are generally observed in a positively skewed frequency distribution :

- (1) The values of median and mode are in decreasing order in this distribution. That is, $\bar{x} > M > M_o$

(2) The distance between third quartile Q_3 and median M is more than the distance between median and first quartile Q_1 . That is $(Q_3 - M) > (M - Q_1)$

(3) The right tail of the frequency curve of this distribution is more elongated.



Note : We find more variation in the observations to the right of the mode in a positively skewed distribution. e.g. positive skewness is found in the frequency distribution of number of deaths.

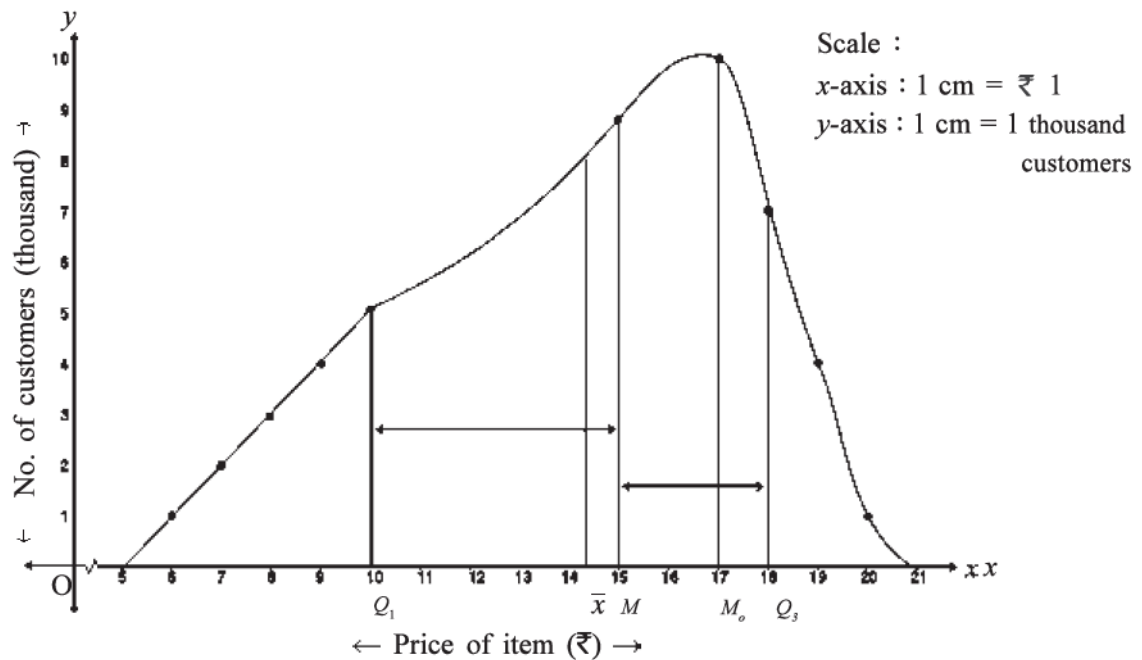
(2) Negative Skewness

If the left tail of the frequency curve of a distribution is more elongated then it is called negatively skewed distribution. Such a population is said to possess negative skewness.

We will study the negative skewness obtaining the frequency curve and central tendency measures of the following frequency distribution.

Price of an item (₹)	6	7	8	9	10	15	17	18	19	20	Total
No. of customers (thousand)	1	2	3	4	5	8	10	7	4	1	45

We will draw the frequency curve of this distribution.



The left tail of the frequency curve is elongated. Less observations are distributed to the right of observation $x = 17$ corresponding to maximum frequency 10 whereas more observations are distributed to the its left and their frequencies are decreasing gradually.

Now we find mean, median, mode and quartiles for the distribution.

Price of item (₹) x	No. of customers f	fx	cf
6	1	6	1
7	2	14	3
8	3	24	6
9	4	36	10
10	5	50	15
15	8	120	23
17	10	170	33
18	7	126	40
19	4	76	44
20	1	20	45
Total	$n = 45$	642	

$$\begin{aligned}
 \text{Mean } \bar{x} &= \frac{\Sigma fx}{n} \\
 &= \frac{642}{45} \\
 &= 14.27 \\
 \bar{x} &= ₹ 14.27
 \end{aligned}$$

$$\begin{aligned}
 \text{Median } M &= \text{value of the } \left(\frac{n+1}{2}\right)\text{th observation} \\
 &= \text{value of the } \left(\frac{45+1}{2}\right)\text{th observation} \\
 &= \text{value of the } \left(\frac{46}{2}\right)\text{th observation} \\
 &= \text{value of the 23rd observation}
 \end{aligned}$$

Referring to *cf* column, we find that the value of the 23rd observation is 15.

$$\therefore M = ₹ 15$$

$$\begin{aligned}
 \text{First quartile } Q_1 &= \text{value of the } \left(\frac{n+1}{4}\right)\text{th observation} \\
 &= \text{value of the } \left(\frac{45+1}{4}\right)\text{th observation} \\
 &= \text{value of the } \left(\frac{46}{4}\right)\text{th observation} \\
 &= \text{value of the 11.5th observation}
 \end{aligned}$$

Referring to *cf* column, we find that the value of the 11.5th observation is 10.

$$\therefore Q_1 = ₹ 10$$

$$\begin{aligned}
 \text{Third quartile } Q_3 &= \text{value of the } 3\left(\frac{n+1}{4}\right)\text{th observation} \\
 &= \text{value of the } 3\left(\frac{45+1}{4}\right)\text{th observation} \\
 &= \text{value of the } 3\left(\frac{46}{4}\right)\text{th observation} \\
 &= \text{value of the } 3(11.5)\text{th observation} \\
 &= \text{value of the 34.5th observation}
 \end{aligned}$$

Referring to *cf* column, we find that the value of the 34.5th observation is 18.

$$\therefore Q_3 = ₹ 18$$

Mode M_o = observation with highest frequency 10 = 17

$$\therefore M_o = ₹ 17$$

From the frequency curve showing mean \bar{x} , median M , mode M_o and the values of quartiles it is clear that

- (1) $\bar{x} = 14.27$, $M = 15$, $M_o = 17$ Hence $\bar{x} < M < M_o$
- (2) The left tail of the frequency curve is more elongated.
- (3) $Q_3 - M = 18 - 15 = 3$ and $M - Q_1 = 15 - 10 = 5$. Hence $Q_3 - M < M - Q_1$

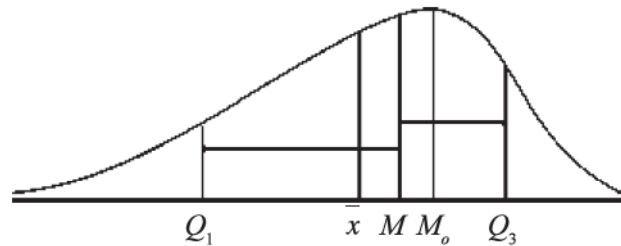
The following characteristics are generally seen in negatively skewed frequency distribution :

(1) Mean, median and mode are in increasing

order. That is $\bar{x} < M < M_o$

(2) The distance between third quartile Q_3 and median M is less than the distance between median and first quartile Q_1 . That is $(Q_3 - M) < (M - Q_1)$

(3) The left tail of the frequency curve of this distribution is more elongated.



Note: We find more variation in the observations to the left of the mode in a negatively skewed distribution. Many cases related to business and finance have negatively skewed frequency distribution. e.g. negatively skewed frequency curve is found in the frequency distributions of price of a certain item, number of investors, etc.

5.3 Concept of absolute and relative measure of skewness

We can find whether the skewness in the frequency distribution of population observations is positive or negative by drawing the frequency curve. But the degree of skewness in the distribution cannot be found using the graph.

Two types of measures are used to measure skewness: **(1) Absolute measure (2) Relative measure.** A measure of skewness which is expressed in the same units of population variable is called absolute measure of skewness. It is denoted by S_k .

The absolute measure of skewness is obtained by the difference of averages in Karl Pearson's method and by the difference of quartiles in Bowley's method. These measures cannot be used for comparing two populations having different units. Even when two populations have observations with same units, it is not advisable to use absolute measure as the measures of central tendency and dispersion may differ in the distributions of both the populations.

Thus, a relative measure is used for a comparative study of two or more populations which is called coefficient of skewness. The absolute measure of skewness of a population is divided by an appropriate measure of dispersion to obtain a measure of skewness which is free from the units. In short, **the relative measure of skewness is called coefficient of skewness.** It is denoted by j .

5.4 Methods for determining skewness and Coefficient of skewness

The following two methods are widely used to find skewness and coefficient of skewness of a frequency distribution: (1) Karl Pearson's method (2) Bowley's method.

5.4.1 Karl Pearson's method

The values of mean, median and mode are not same in a skewed distribution and the median is between mean and mode. Hence the difference between mean and mode is generally used to find the measure of skewness. Thus skewness $S_k = \bar{x} - M_o$. Coefficient of skewness j for a unimodal distribution is obtained by dividing skewness S_k by the standard deviation 's'. The coefficient of skewness is thus found by the

following formula : $j = \frac{\bar{x} - M_o}{s}$

When a distribution has multiple modes or when the value of mode is ill-defined, the mode is obtained by empirical formula $M_o = 3M - 2\bar{x}$ given by Karl Pearson. Thus, skewness is obtained by the following formula:

$$\text{Skewness } S_k = \text{mean} - \text{mode} = \bar{x} - M_o = \bar{x} - (3M - 2\bar{x}) = 3\bar{x} - 3M = 3(\bar{x} - M)$$

$$\text{and coefficient of skewness } j = \frac{3(\bar{x} - M)}{s}.$$

Note : There are two types of skewness: (1) Positive skewness (2) Negative skewness

(1) If the mode is well-defined in a distribution, skewness is found by the formula $S_k = \bar{x} - M_o$. For positively skewed distribution we have $\bar{x} > M_o$ and thus $S_k = \bar{x} - M_o > 0$. Hence when $S_k > 0$, the distribution is said to have positive skewness. Similarly in case of ill-defined mode, $\bar{x} > M$, $S_k = 3(\bar{x} - M) > 0$ is considered as positive skewness.

(2) If the mode is well-defined in a distribution, skewness is found by the formula $S_k = \bar{x} - M_o$. For negatively skewed distribution we have $\bar{x} < M_o$ and thus $S_k = (\bar{x} - M_o) < 0$. Hence when $S_k < 0$, the distribution is said to have negative skewness. Similarly in case of ill-defined mode, $\bar{x} < M$, $S_k = 3(\bar{x} - M) < 0$ is considered as negative skewness.

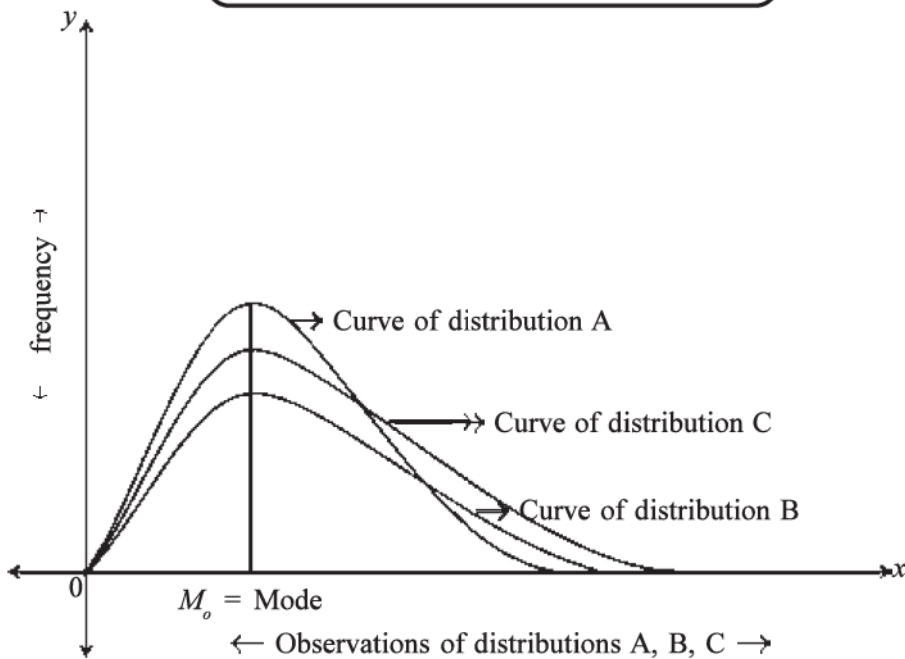
Note :

(1) In practice, coefficient of skewness lies between -1 and 1 for a skewed distribution based on sample data.

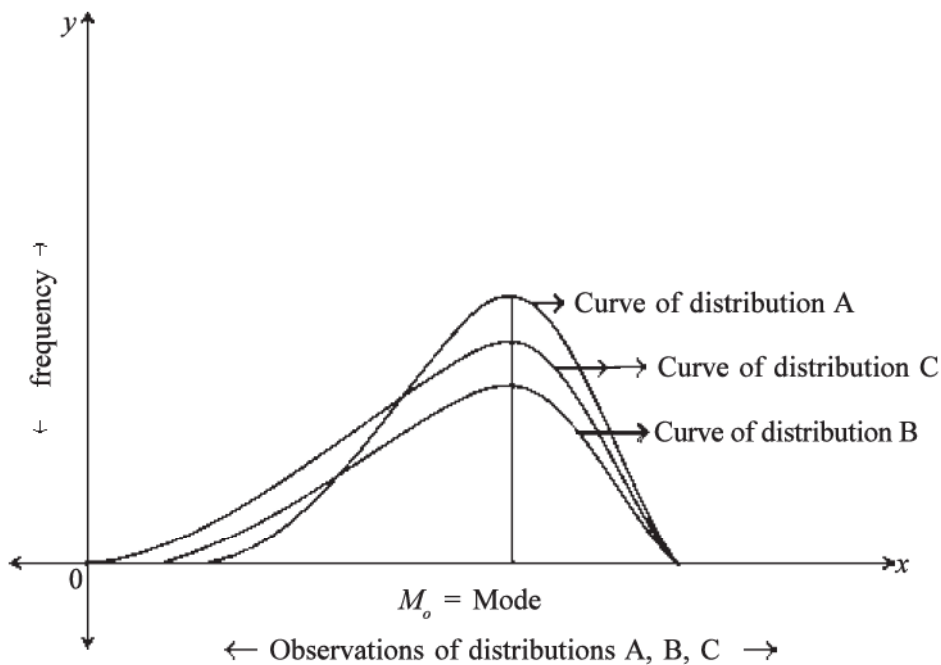
(2) If the frequency curve of a population has multiple modes, the coefficient of skewness $j = \frac{3(\bar{x} - M)}{s}$ lies between -3 and 3 .

(3) In the year 1951 statistician N. L. Johnson proved that the coefficient of skewness obtained as $j = \frac{\bar{x} - M_o}{s}$ theoretically lies between $-\sqrt{3}$ and $\sqrt{3}$ that is between -1.73 and 1.73 for a unimodal skewed distribution.

Additional Information for understanding



All the above frequency curves show positive skewness. The right tail for distribution C among them is the most elongated. So, it has maximum positive skewness. The curve of distribution B has less positive skewness than C. As the tail of distribution A is the least elongated as compared to B and C, it has the least positive skewness.



All the above frequency curves show negative skewness. The left tail for distribution C is the most elongated among them. So, it has the maximum negative skewness. The curve of distribution B has less negative skewness than C. As the tail of distribution A is the least elongated as compared to B and C, it has the least negative skewness.

Illustration 1 : The following data related to units transported by 50 trucks from railway yard to different factories on a day. Find the skewness and its coefficient using Karl Pearson's method from these data.

No. of units transported	120	130	140	150	160	170	180	190	200
No. of trucks	2	3	4	5	11	9	9	6	1

The given frequency distribution is unimodal. Hence, we will find mean, mode and standard deviation to find skewness by Karl Pearson's method.

No. of units transported x	No. of trucks f	$d = x - A$ $A = 160$	fd	fd^2
120	2	-40	-80	3200
130	3	-30	-90	2700
140	4	-20	-80	1600
150	5	-10	-50	500
160	11	0	0	0
170	9	10	90	900
180	9	20	180	3600
190	6	30	180	5400
200	1	40	40	1600
Total	$n = 50$		190	19500

$$\begin{aligned} \text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \\ &= 160 + \frac{190}{50} \\ &= 160 + 3.8 \\ &= 163.8 \end{aligned}$$

$$\therefore \bar{x} = 163.8 \text{ units}$$

Mode M_o = observation with the highest frequency = 160

$$\therefore M_o = 160 \text{ units}$$

$$\begin{aligned} \text{Standard deviation } s &= \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \\ &= \sqrt{\frac{19500}{50} - \left(\frac{190}{50}\right)^2} \\ &= \sqrt{390 - 14.44} \\ &= \sqrt{375.56} \\ &= 19.3794 \end{aligned}$$

$$\therefore s = 19.38 \text{ units}$$

$$\begin{aligned}
 \text{Skewness } S_k &= \bar{x} - M_o \\
 &= 163.8 - 160 \\
 &= 3.8 \\
 S_k &= 3.8 \text{ units}
 \end{aligned}$$

$$\begin{aligned}
 \text{Coefficient of skewness } j &= \frac{\bar{x} - M_o}{s} \\
 &= \frac{163.8 - 160}{19.38} \\
 &= \frac{3.8}{19.38} \\
 &= 0.1961 \\
 \therefore j &\approx 0.20
 \end{aligned}$$

This distribution has positive skewness. It should be noted that $j = 0.20$ is free from units as coefficient of skewness is free from units.

Illustration 2 : Find the skewness and coefficient of skewness by Karl Pearson's method from the following data about annual tax of 100 companies.

Annual tax (lakh ₹)	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100	100 - 120
No. of companies	5	20	40	25	6	4

As the given frequency distribution is unimodal, we will compute mean \bar{x} , mode M_o and standard deviation s .

Annual tax (lakh ₹)	No. of companies f	Mid-value x	$d = \frac{x - A}{c}$ $A = 50, c = 20$	fd	fd^2
0 - 20	5	10	- 2	- 10	20
20 - 40	20	30	- 1	- 20	20
40 - 60	40	50	0	0	0
60 - 80	25	70	1	25	25
80 - 100	6	90	2	12	24
100 - 120	4	110	3	12	36
Total	$n = 100$			19	125

$$\begin{aligned}
 \text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\
 &= 50 + \frac{19}{100} \times 20 \\
 &= 50 + \frac{380}{100} \\
 &= 50 + 3.8
 \end{aligned}$$

$$\begin{aligned}
 \therefore \bar{x} &= 53.8 \\
 \bar{x} &= ₹ 53.8 \text{ lakh}
 \end{aligned}$$

$$\begin{aligned}
 \text{Standard deviation } s &= \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times c \\
 &= \sqrt{\frac{125}{100} - \left(\frac{19}{100}\right)^2} \times 20 \\
 &= \sqrt{1.25 - 0.0361} \times 20 \\
 &= \sqrt{1.2139} \times 20 \\
 &= 22.0354 \\
 s &= ₹ 22.04 \text{ lakh}
 \end{aligned}$$

Mode M_o : The class interval with maximum frequency 40 is 40 - 60. The modal class in 40–60.

$$\text{Now, } M_o = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$$

Here, $L = 40$, $f_m = 40$, $f_1 = 20$, $f_2 = 25$, $c = 20$

$$\begin{aligned}
 M_o &= 40 + \frac{40 - 20}{2(40) - 20 - 25} \times 20 \\
 &= 40 + \frac{20 \times 20}{80 - 20 - 25} \\
 &= 40 + \frac{400}{35} \\
 &= 40 + 11.4285 \\
 &= 51.4285
 \end{aligned}$$

$$M_o = ₹ 51.43 \text{ lakh}$$

$$\begin{aligned}
 \text{Skewness } S_k &= \bar{x} - M_o \\
 &= 53.8 - 51.43 \\
 &= 2.37
 \end{aligned}$$

$$S_k = ₹ 2.37 \text{ lakh}$$

$$\begin{aligned}
 \text{Coefficient of skewness } j &= \frac{\bar{x} - M_o}{s} = \frac{53.8 - 51.43}{22.04} \\
 &= \frac{2.37}{22.04} \\
 &= 0.1075 \\
 j &= 0.11
 \end{aligned}$$

Hence, it can be said that the frequency distribution has positive skewness. Small value of coefficient of skewness indicates that it is almost symmetric.

Illustration 3 : A factory has 100 machines for production. The following data are obtained about rejected items during the production process. Find the skewness and its coefficient using Karl Pearson's method from these data.

No. of rejected items	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60	61 - 70	71-80
No. of machines	2	12	25	39	12	9	1

The given frequency distribution is unimodal. Hence, we will compute mean, mode and standard deviation to find coefficient of skewness.

No. of rejected items	No. of machines f	Mid-value x	$d = \frac{x - A}{c}$ $A = 45.5$ $c = 10$	fd	fd^2
11 - 20	2	15.5	- 3	- 6	18
21 - 30	12	25.5	- 2	- 24	48
31 - 40	25	35.5	- 1	- 25	25
41 - 50	39	45.5	0	0	0
51 - 60	12	55.5	1	12	12
61 - 70	9	65.5	2	18	36
71 - 80	1	75.5	3	3	9
Total	$n = 100$			- 22	148

$$\begin{aligned} \text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\ &= 45.5 + \frac{(-22)}{100} \times 10 \\ &= 45.5 - 2.2 \\ &= 43.3 \end{aligned}$$

$$\therefore \bar{x} = 43.3 \text{ units}$$

Mode M_o : The class interval with maximum frequency 39 is 41 - 50. Hence modal class of inclusive type is 41 - 50.

\therefore Taking class boundaries, the modal class is 40.5 - 50.5.

$$\text{Now, } M_o = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$$

$$\text{Here, } L = 40.5, f_m = 39, f_1 = 25, f_2 = 12, c = 10$$

$$\begin{aligned} M_o &= 40.5 + \frac{39 - 25}{2(39) - 25 - 12} \times 10 \\ &= 40.5 + \frac{14 \times 10}{78 - 37} \\ &= 40.5 + \frac{140}{41} \\ &= 40.5 + 3.4146 \\ &= 43.9146 \\ M_o &= 43.91 \text{ units} \end{aligned}$$

$$\begin{aligned}
 \text{Standard deviation } s &= \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times c \\
 &= \sqrt{\frac{148}{100} - \left(\frac{-22}{100}\right)^2} \times 10 \\
 &= \sqrt{1.48 - 0.0484} \times 10 \\
 &= \sqrt{1.4316} \times 10 \\
 &= 11.96 \text{ units}
 \end{aligned}$$

$$\begin{aligned}
 \text{Skewness } S_k &= \bar{x} - M_o \\
 &= 43.3 - 43.91 \\
 S_k &= -0.61 \text{ units}
 \end{aligned}$$

$$\begin{aligned}
 \text{Coefficient of skewness } j &= \frac{\bar{x} - M_o}{s} \\
 &= \frac{43.3 - 43.91}{11.96} \\
 &= \frac{-0.61}{11.96} \\
 &= -0.0509 \\
 \therefore j &\approx -0.05
 \end{aligned}$$

The given frequency distribution has negative skewness which is closer to symmetry.

Illustration 4 : The average monthly transportation cost of 200 families in a city in the year 2014 is as follows. Find skewness and coefficient of skewness using Karl Pearson's method.

Average monthly transportation cost (thousand ₹)	1 - 3	4 - 6	7 - 9	10 - 13	14 - 16	17 - 19
No. of families	5	40	120	20	10	5

As the given frequency distribution has classes of unequal lengths, we compute mean, median and standard deviation to find coefficient of skewness.

Average monthly transportation cost (thousand ₹)	No. of families f	Mid-value x	$d = x - A$ $A = 8$	fd	fd^2	cf
1-3	5	2	- 6	- 30	180	5
4-6	40	5	- 3	- 120	360	45
7-9	120	8	0	0	0	165
10-13	20	11.5	3.5	70	245	185
14-16	10	15	7	70	490	195
17-19	5	18	10	50	500	200
Total	$n = 200$			40	1775	

$$\begin{aligned}
 \text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \\
 &= 8 + \frac{40}{200} \\
 &= 8 + 0.2 \\
 &= 8.2
 \end{aligned}$$

$$\therefore \bar{x} = ₹ 8.2 \text{ thousand}$$

$$\begin{aligned}
 \text{Median } M &= \text{value of the } \left(\frac{n}{2}\right) \text{ th observation} \\
 &= \text{value of the } \left(\frac{200}{2}\right) \text{ th observation} \\
 &= \text{value of the 100 th observation}
 \end{aligned}$$

Referring to cf column, we find that the 100th observation lies in the interval 7 - 9. Hence, class of median is 7-9, which is inclusive type. \therefore The class boundaries of median class are 6.5 - 9.5

$$\text{Now, } M = L + \frac{\frac{n}{2} - cf}{f} \times c$$

$$\text{Here, } L = 6.5, \frac{n}{2} = 100, cf = 45, f = 120 \text{ and } c = 3$$

$$\begin{aligned}
 M &= 6.5 + \frac{100 - 45}{120} \times 3 \\
 &= 6.5 + \frac{55 \times 3}{120} \\
 &= 6.5 + 1.375 \\
 &= 7.875
 \end{aligned}$$

$$\therefore M = ₹ 7.88 \text{ thousand}$$

$$\begin{aligned}
 \text{Standard deviation } s &= \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \\
 &= \sqrt{\frac{1775}{200} - \left(\frac{40}{200}\right)^2} \\
 &= \sqrt{8.875 - 0.04} \\
 &= \sqrt{8.835} \\
 &= 2.9724 \\
 \therefore s &= ₹ 2.97 \text{ thousand}
 \end{aligned}$$

$$\begin{aligned}
 \text{Skewness } S_k &= 3 (\bar{x} - M) \\
 &= 3 (8.2 - 7.88) \\
 &= 3 (0.32) \\
 &= 0.96 \\
 S_k &= ₹ 0.96 \text{ thousand}
 \end{aligned}$$

$$\begin{aligned}
 \text{Coefficient of skewness } j &= \frac{3(\bar{x} - M)}{s} \\
 &= \frac{3(8.2 - 7.88)}{2.97} \\
 &= \frac{3(0.32)}{2.97} \\
 &= \frac{0.96}{2.97} \\
 &= 0.3232 \\
 &= 0.32 \\
 \therefore j &\approx 0.32
 \end{aligned}$$

Thus the given frequency distribution has positive skewness.

Illustration 5 : The following information is obtained for life (in complete hours) of 400 electric bulbs. Find the skewness and the coefficient of skewness by Karl Pearson's method.

Life of electric bulbs (completed hours)	4000-4199	4200-4399	4400-4599	4600-4799	4800-4999	5000-5199	5200-5399	5400-5599
No. of electric bulbs	14	46	58	76	70	76	40	20

Two classes of the above distribution have maximum frequency 76. Since this is a bimodal distribution, we will compute mean, median and standard deviation.

Life of electric bulbs (completed hours)	No. of electric bulbs f	Mid-value x	$d = \frac{x - A}{c}$ $A = 4699.5$ $c = 200$	fd	fd^2	cf
4000 - 4199	14	4099.5	- 3	- 42	126	14
4200 - 4399	46	4299.5	- 2	- 92	184	60
4400 - 4599	58	4499.5	- 1	- 58	58	118
4600 - 4799	76	4699.5	0	0	0	194
4800 - 4999	70	4899.5	1	70	70	264
5000 - 5199	76	5099.5	2	152	304	340
5200 - 5399	40	5299.5	3	120	360	380
5400 - 5599	20	5499.5	4	80	320	400
Total	$n = 400$			230	1422	

$$\begin{aligned}
 \text{Mean } \bar{x} &= A + \frac{\sum fd}{n} \times c \\
 &= 4699.5 + \frac{230}{400} \times 200 \\
 &= 4699.5 + 115 \\
 &= 4814.5 \\
 \therefore \bar{x} &= 4814.5 \text{ hours}
 \end{aligned}$$

Median $M =$ value of the $\left(\frac{n}{2}\right)$ th observation

$$= \text{value of the } \left(\frac{400}{2}\right) \text{th observation}$$

$$= \text{value of the 200th observation}$$

Referring to cf column, we find that the 200th observation lies in the interval 4800 - 4999.

Hence class of median is 4800 - 4999 which is inclusive type. \therefore The class boundaries of median class are 4799.5 - 4999.5.

$$\text{Now, } M = L + \frac{\frac{n}{2} - cf}{f} \times c$$

Here, $L = 4799.5$, $\frac{n}{2} = 200$, $cf = 194$, $f = 70$ and $c = 200$ \AA .

$$M = 4799.5 + \frac{200 - 194}{70} \times 200$$

$$= 4799.5 + \frac{6 \times 200}{70}$$

$$= 4799.5 + \frac{120}{7}$$

$$= 4799.5 + 17.1429$$

$$= 4816.6429$$

$\therefore M = 4816.64$ hours

$$\text{Standard deviation } s = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times c$$

$$= \sqrt{\frac{1422}{400} - \left(\frac{230}{400}\right)^2} \times 200$$

$$= \sqrt{3.555 - 0.3306} \times 200$$

$$= \sqrt{3.2244} \times 200$$

$$= 359.1323$$

$\therefore s = 359.13$ hours

$$\text{Skewness } S_k = 3(\bar{x} - M)$$

$$= 3(4814.5 - 4816.64)$$

$$= 3(-2.14)$$

$$= -6.42$$

$\therefore S_k = -6.42$ hours

$$\begin{aligned}
 \text{Coefficient skewness } j &= \frac{3(\bar{x}-M)}{s} \\
 &= \frac{3(4814.5-4816.64)}{359.13} \\
 &= \frac{-6.42}{359.13} \\
 &= -0.0178 \\
 j &\approx -0.02
 \end{aligned}$$

Thus, the given distribution has negative skewness.

Illustration 6 : The following information is about the annual advertisement cost (in lakh ₹) of 30 companies. Find skewness and coefficient of skewness using it.

Annual advertisement cost (lakh ₹)	0	3	5	8	10 - 20	20 - 30	30 - 40
No. of companies	3	4	5	10	5	2	1

The given distribution is mixed (discrete and continuous). Hence we find \bar{x} , M , s and coefficient of skewness.

Annual cost (lakh ₹)	No. of companies f	Mid-value x	$d = x - A$ $A = 15$	fd	fd^2	cf
0	3	0	- 15	- 45	675	3
3	4	3	- 12	- 48	576	7
5	5	5	- 10	- 50	500	12
8	10	8	- 7	- 70	490	22
10 - 20	5	15	0	0	0	27
20 - 30	2	25	10	20	200	29
30 - 40	1	35	20	20	400	30
Total	$n = 30$			-173	2841	

$$\begin{aligned}
 \text{Mean } \bar{x} &= A + \frac{\Sigma fd}{n} \\
 &= 15 + \frac{-173}{30} \\
 &= 15 - 5.7667 \\
 &= 9.2333
 \end{aligned}$$

$$\therefore \bar{x} = ₹ 9.23 \text{ lakh}$$

Median $M =$ Value of the $\left(\frac{n}{2}\right)$ th observation

$$= \text{Value of the } \left(\frac{30}{2}\right)\text{th observation}$$

$$= \text{Value of the 15th observation}$$

Referring to *cf* column, we find that the 15th observation is 8.

$$\therefore M = ₹ 8 \text{ lakh}$$

$$\text{Standard deviation } s = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2}$$

$$= \sqrt{\frac{2841}{30} - \left(\frac{-173}{30}\right)^2}$$

$$= \sqrt{94.7 - 33.2544}$$

$$= \sqrt{61.4456}$$

$$= 7.8387$$

$$\therefore s = ₹ 7.84 \text{ lakh}$$

$$\text{Skewness } S_k = 3(\bar{x} - M)$$

$$= 3(9.23 - 8)$$

$$= 3(1.23)$$

$$= 3.69$$

$$\therefore S_k = ₹ 3.69 \text{ lakh}$$

$$\text{Coefficient skewness } j = \frac{3(\bar{x} - M)}{s}$$

$$= \frac{3(9.23 - 8)}{7.84}$$

$$= \frac{3.69}{7.84}$$

$$= 0.4707$$

$$j = 0.47$$

Thus, the given frequency distribution has positive skewness.

Illustration 7 : The frequency distribution of hourly wages paid to 600 workers is given below. Find the Karl Pearson's coefficient of skewness from this distribution.

Hourly wage (lakh ₹)	40 - 60	60 - 80	80 - 100	100 - 120	120 - 140	140 - 160	160 - 180	180 - 200	200 - 220	220 - 240
No. of workers (<i>f</i>)	10	12	16	20	50	60	72	100	120	140

The maximum frequency in this distribution is 140 which is for the last class instead of the central class.

Hence, the mode is ill-defined. We will find the coefficient of skewness using the formula $j = \frac{3(\bar{x} - M)}{s}$

Hourly wages (₹)	No. of workers f	Mid-value x	$d = \frac{x - A}{c}$ $A = 130 \quad c = 20$	fd	fd^2	cf
40 - 60	10	50	- 4	- 40	160	10
60 - 80	12	70	- 3	- 36	108	22
80 - 100	16	90	- 2	- 32	64	38
100 - 120	20	110	- 1	- 20	20	58
120 - 140	50	130	0	0	0	108
140 - 160	60	150	1	60	60	168
160 - 180	72	170	2	144	288	240
180 - 200	100	190	3	300	900	340
200 - 220	120	210	4	480	1920	460
220 - 240	140	230	5	700	3500	600
Total	$n = 600$			1556	7020	

$$\text{Mean } \bar{x} = A + \frac{\sum fd}{n} \times c$$

$$= 130 + \frac{1556}{600} \times 20$$

$$= 130 + 51.8666$$

$$= 181.866$$

$$\therefore \bar{x} = ₹ 181.87$$

$$\text{Standard deviation } s = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times c$$

$$= \sqrt{\frac{7020}{600} - \left(\frac{1556}{600}\right)^2} \times 20$$

$$= \sqrt{11.7 - 6.7254} \times 20$$

$$= \sqrt{4.9746} \times 20$$

$$= 44.6076$$

$$\therefore s = ₹ 44.61$$

$$\begin{aligned} \text{Median } M &= \text{Value of the } \left(\frac{n}{2}\right) \text{ th observation} \\ &= \text{Value of the } \left(\frac{600}{2}\right) \text{ th observation} \\ &= \text{Value of the 300th observation} \end{aligned}$$

Referring to cf column, we find that the 300th observation lies in the interval 180-200. Hence, class of median is 180-200.

$$\text{Now, } M = L + \frac{\frac{n}{2} - cf}{f} \times c$$

$$\text{Here, } L = 180, \frac{n}{2} = 300, cf = 240, f = 100 \text{ and } c = 20.$$

$$\begin{aligned} M &= 180 + \frac{300 - 240}{100} \times 20 \\ &= 180 + \frac{60 \times 20}{100} \\ &= 180 + 12 \\ &= 192 \end{aligned}$$

$$\therefore M = ₹ 192$$

$$\begin{aligned} \text{Coefficient skewness } j &= \frac{3(\bar{x} - M)}{s} \\ &= \frac{3(181.87 - 192)}{44.61} \\ &= \frac{-30.39}{44.61} \\ &= -0.6812 \\ j &\approx -0.68 \end{aligned}$$

Thus, the given frequency distribution has negative skewness.

Illustration 8 : The following figures are given to describe the changes in the market prices of shares of a company before and after their general body meeting. Take them into consideration to comment whether the proceedings of this meeting have affected the market prices of shares by computing coefficient of skewness.

Details	Before meeting	After meeting
No. of share transactions	6000	5800
Mean of prices of share (₹)	440	460
Median of prices of share (₹)	500	480
Standard deviation of prices of share (₹)	60	52

We will be able to comment on the changes in the distribution of market prices of shares before and after the general body meeting by computing the coefficient of skewness. As we are given mean and median, we will find the coefficient of skewness using the following formula.

Coefficient of skewness Before meeting :

$$\begin{aligned}
 j &= \frac{3(\bar{x}-M)}{s} \\
 &= \frac{3(440-500)}{60} \\
 &= \frac{3(-60)}{60} \\
 j &= -3
 \end{aligned}$$

Coefficient of skewness After meeting :

$$\begin{aligned}
 j &= \frac{3(\bar{x}-M)}{s} \\
 &= \frac{3(460-480)}{52} \\
 &= \frac{3(-20)}{52} \\
 j &= -1.15
 \end{aligned}$$

(1) The distributions of share prices in both situations have negative skewness.

(2) The coefficient of skewness has decreased after the general body meeting. Hence, we can say that there is a partial effect of general body meeting on the share prices.

Illustration 9 : The data about the number of runs scored by two cricketers in 10 matches are as follows. Use this information to determine which cricketer's game is more skewed.

Details	Cricketer A	Cricketer B
Mean \bar{x}	50	35
Mode M_o	56	31
Standard deviation s	14.4	5.2

We will compute the coefficient of skewness for the scores of both cricketers from the given data :

Coefficient of skewness for cricketer A :

$$\begin{aligned}
 j &= \frac{\bar{x}-M_o}{s} \\
 &= \frac{50-56}{14.4} \\
 &= \frac{-6}{14.4} \\
 j &= -0.42
 \end{aligned}$$

Coefficient of skewness for cricketer B :

$$\begin{aligned}
 j &= \frac{\bar{x}-M_o}{s} \\
 &= \frac{35-31}{5.2} \\
 &= \frac{4}{5.2} \\
 j &= 0.77
 \end{aligned}$$

Ignoring the signs of coefficient of skewness for scores of cricketers, we see that the value of coefficient for cricketer B is more than that of A. Hence, we say that the game of cricketer B is more skewed.

Illustration 10 : From the following measures obtained for the frequency distributions of sales (in lakh Rs.) of potatoes by two merchants in a month, determine which distribution is more close to symmetry.

Measures for distribution of sales of potatoes by merchant A	Measures for distribution of sales of potatoes by merchant B
Mean $\bar{x} = ₹ 40$	Mean $\bar{x} = ₹ 45$
Median $M = ₹ 43$	Median $M = ₹ 40$
Standard deviation $s = ₹ 25$	Standard deviation $s = ₹ 16$

Coefficient of skewness for sale of A :

$$\begin{aligned}
 j &= \frac{3(\bar{x}-M)}{s} \\
 &= \frac{3(40-43)}{25} \\
 &= \frac{3(-3)}{25} \\
 &= -0.36 \\
 j &= -0.36
 \end{aligned}$$

Coefficient of skewness for sale of B :

$$\begin{aligned}
 j &= \frac{3(\bar{x}-M)}{s} \\
 &= \frac{3(45-40)}{16} \\
 &= \frac{3(5)}{16} \\
 &= 0.9375 \\
 j &= 0.94
 \end{aligned}$$

The distribution having less numerical value (ignoring the sign) of coefficient of skewness is said to be more close to symmetry. The coefficient of distribution A is numerically less than that of B. Hence, the distribution of A is more close to symmetry than distribution B.

Illustration 11 : The median of the distribution of marks in statistics obtained by students of a school is 72 and its mean is 75. Find the skewness for the marks obtained by the students from these data and state the type of skewness.

Here, median $M = 72$ and mean $\bar{x} = 75$. We will find the skewness using the following formula as the mode is not given.

$$\begin{aligned}
 \text{Skewness } S_k &= 3(\bar{x} - M) \\
 &= 3(75 - 72) \\
 &= 3(3) = 9 \text{ Marks} \\
 \therefore S_k &= 9 \text{ marks}
 \end{aligned}$$

As, $S_k > 0$ the frequency distribution of marks obtained by students has positive skewness.

Activity

Collect the data about the 'number of absent days in a month' for all the students of your class. Construct a frequency distribution for the number of absent days. Find the absolute and relative measures of skewness for it.

Find the absolute and relative measures of skewness for a similar distribution for the students of another class. Compare the frequency distributions of both the classes using frequency curve, coefficient of variation and coefficient of skewness.

EXERCISE 5.1

- The following distribution shows demand of 500 ml pouches of pasteurized toned milk by 59 consumers. Find Karl Pearson's coefficient of skewness from these data.

Demand of milk pouches (units)	1	2	3	4	5	6	7	8
No. of consumers	2	7	10	15	12	7	4	2

2. The following distribution shows purchase of T-shirts by 270 customers according to the shoulder lengths (in inches). Find Karl Pearson's coefficient of skewness from these data and interpret it.

Shoulder length of shirts (inches)	12.0	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0
No. of customers	5	20	30	47	56	56	37	16	3

3. The following information is obtained for the time taken (in completed minutes) by each worker to carry out a certain job. Find Karl Pearson's coefficient of skewness from these data and interpret it.

Time taken (completed min.)	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29
No. of workers	3	8	4	2	1

4. The students of standard 11 have collected the data about profits (in crore Rs.) of IT companies. Find Karl Pearson's coefficient of skewness from these data and interpret it.

Profit (crore ₹)	5 - 7	7 - 9	9 - 11	11 - 13	13 - 15	15 - 17
No. of companies	5	12	20	8	3	2

5. The following frequency distribution gives the amount of annual depreciation (in lakh ₹) of 38 companies. Using this information, find skewness and its coefficient by Karl Pearson's method. State the type of skewness.

Amount of annual depreciation (lakh ₹)	7	9	10	11 - 20	21 - 24	25 - 36
No. of companies	2	3	4	7	12	10

6. The monthly consumption of cotton (in thousand bales) of 35 cloth mills is as follows. Using this, find skewness and its coefficient by Karl Pearson's method. State the type of skewness.

Consumption of cotton (thousand bales)	0 - 2	2 - 6	6 - 12	12 - 20	20 - 22
No. of mills	3	10	7	12	3

7. The temperature (in Celsius) at a tourist place for 60 days in the year 2014 is as follows. Using this information, find skewness and its coefficient by Karl Pearson's method. State the type of skewness.

Temperature (Celsius)	-3° to -1°	-1° to 5°	5° to 11°	11° to 19°	19° to 23°	23° to 27°
No. of days	4	14	20	14	5	3

*

5.4.2 Bowley's Method

Prof. A. L. Bowley has given a measure of skewness based on quartiles. This measure is based on the position of quartiles. It considers the main assumption that 'both the quartiles Q_1 and Q_3 are not equidistant from the median $M (= Q_2)$ in a skewed frequency distribution. The absolute measure of skewness S_k is obtained using the quartile differences $Q_3 - M$ and $M - Q_1$.

Thus, skewness = $S_k = (Q_3 - M) - (M - Q_1)$

$$\therefore S_k = Q_3 + Q_1 - 2M$$

We have studied two types of skewness. We know that if the frequency distribution has positive skewness $Q_3 - M > M - Q_1$.

$$\therefore Q_3 + Q_1 > 2M \text{ and } S_k > 0.$$

For a negatively skewed distribution, we have $Q_3 - M < M - Q_1 \therefore Q_3 + Q_1 < 2M$ and $S_k < 0$.

If the frequency distribution is symmetric, we have $Q_3 - M = M - Q_1 \therefore Q_3 + Q_1 = 2M$ and $S_k = 0$.

The distances of Q_1 and Q_3 from the median M are $M - Q_1$ and $Q_3 - M$ respectively. When the difference between these two values is divided by their sum, we obtain the relative measure of skewness called coefficient of skewness. Thus, Bowley's formula for **coefficient of skewness j** is as follows:

$$j = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

The measure of skewness obtained by this formula is called Bowley's coefficient of skewness. We know that the values of $(Q_3 - M)$ and $(M - Q_1)$ are positive. The difference between any two real

numbers is less than or equal to their sum. Thus $\left| \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)} \right| \leq 1$.

\therefore |Bowley's coefficient of skewness j | ≤ 1

$\therefore -1 \leq j \leq 1$. Thus, the range of Bowley's coefficient of skewness is -1 to 1

Note :

(1) If the frequency distribution is open-ended, this is the only method to find absolute and relative measures of skewness.

(2) Bowley's method is appropriate when the skewness is to be found using the positional averages which are quartiles and median.

Illustration 12 : The data about rainfall (in cm) at a place in a month during monsoon are obtained as follows. Find the skewness and coefficient of skewness using Bowley's method.

Rainfall (cm)	6	7	13	5	15	20
No. of days	10	5	3	8	3	2

We will find quartiles Q_1 , $Q_2 (= M)$ and Q_3 after arranging the observations for rainfall in the ascending order and then compute skewness and its coefficient using Bowley's formula.

Rainfall (cm)	No. of days f	cf
5	8	8
6	10	18
7	5	23
13	3	26
15	3	29
20	2	31
Total	$n = 31$	

$$\begin{aligned}
 \text{First quartile } Q_1 &= \text{Value of the } \left(\frac{n+1}{4}\right) \text{ th observation} \\
 &= \text{Value of the } \left(\frac{31+1}{4}\right) \text{ th observation} \\
 &= \text{Value of the } \frac{32}{4} \text{ th observation} \\
 &= \text{Value of the 8th observation}
 \end{aligned}$$

Referring to *cf* column, we find that the value of the 8th observation is 5.

$$\therefore Q_1 = 5 \text{ cm}$$

$$\text{Median } M = \text{Value of the } \left(\frac{n+1}{2}\right) \text{ th observation}$$

$$M = \text{Value of the } \left(\frac{31+1}{2}\right) \text{ th observation}$$

$$= \text{Value of the } \frac{32}{2} \text{ th observation}$$

$$= \text{Value of the 16 th observation}$$

Referring to *cf* column, we find that the value of the 16th observation is 6.

$$\therefore M = 6 \text{ cm}$$

$$\text{Third quartile } Q_3 = \text{Value of the } 3 \left(\frac{n+1}{4}\right) \text{ th observation}$$

$$= \text{Value of the } 3 \left(\frac{31+1}{4}\right) \text{ th observation}$$

$$= \text{Value of the } 3 \left(\frac{32}{4}\right) \text{ th observation}$$

$$= \text{Value of the } 3(8) \text{ th observation}$$

$$= \text{Value of the 24th observation}$$

Referring to *cf* column, we find that the value of the 24th observation is 13.

$$\therefore Q_3 = 13 \text{ cm}$$

$$\text{Skewness } S_k = Q_3 + Q_1 - 2M$$

$$= 13 + 5 - 2(6)$$

$$= 18 - 12$$

$$= 6$$

$$S_k = 6 \text{ cm}$$

$$\text{Coefficient of skewness } j = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$= \frac{13 + 5 - 2(6)}{13 - 5}$$

$$= \frac{18 - 12}{8}$$

$$= \frac{6}{8}$$

$$j = 0.75$$

Illustration 13 : The frequency distribution of number of cheques received per day for clearing of 5 branches of a bank on 100 days in the year 2014 is as follows. Find the coefficient of skewness by Bowley's method using this distribution.

No. of cheques	0 - 199	200 - 399	400 - 599	600 - 799	800 - 999
No. of days	10	13	17	42	18

We will compute first quartile Q_1 , median $M (= Q_2)$ and third quartile Q_3 to find Bowley's coefficient.

No. of cheques	No. of days f	cf
0 - 199	10	10
200 - 399	13	23
400 - 599	17	40
600 - 799	42	82
800 - 999	18	100
Total	$n = 100$	

$$\begin{aligned} \text{First quartile } Q_1 &= \text{Value of the } \left(\frac{n}{4}\right)\text{th observation} \\ &= \text{Value of the } \left(\frac{100}{4}\right)\text{th observation} \\ &= \text{Value of the 25th observation} \end{aligned}$$

Referring to cf column, we find that the 25th observation lies in the interval 400 - 599. Hence class of Q_1 is 400 - 599 which is inclusive type. \therefore The class boundaries of class of Q_1 are 399.5 - 599.5

$$\text{Now, } Q_1 = L + \frac{\frac{n}{4} - cf}{f} \times c$$

$$\text{Here, } L = 399.5, \frac{n}{4} = 25, cf = 23, f = 17 \text{ and } c = 200$$

$$\begin{aligned} Q_1 &= 399.5 + \frac{25 - 23}{17} \times 200 \\ &= 399.5 + \frac{2 \times 200}{17} \\ &= 399.5 + \frac{400}{17} \\ &= 399.5 + 23.5294 \\ &= 423.0294 \end{aligned}$$

$$\therefore Q_1 = 423.03 \text{ cheques}$$

$$\begin{aligned} \text{Median } M &= \text{Value of the } \left(\frac{n}{2}\right)\text{th observation} \\ &= \text{Value of the } \left(\frac{100}{2}\right)\text{th observation} \\ &= \text{Value of the 50th observation} \end{aligned}$$

Referring to cf column, we find that the 50th observation lies in the interval 600 - 799. Hence, class of median is 600-799 which is inclusive type. \therefore The class boundaries of class of Q_1 are 599.5 - 799.5.

$$\text{Now, } M = L + \frac{\frac{n}{2} - cf}{f} \times c$$

Here, $L = 599.5$, $\frac{n}{2} = 50$, $cf = 40$, $f = 42$ and $c = 200$

$$\begin{aligned} M &= 599.5 + \frac{50-40}{42} \times 200 \\ &= 599.5 + \frac{10 \times 200}{42} \\ &= 599.5 + \frac{2000}{42} \\ &= 599.5 + 47.619 \\ &= 647.119 \end{aligned}$$

$$M \approx 647.12 \text{ cheque}$$

Third quartile $Q_3 =$ Value of the $3\left(\frac{n}{4}\right)$ th observation
 $=$ Value of the $3\left(\frac{100}{4}\right)$ th observation
 $=$ Value of the 3(25)th observation
 $=$ Value of the 75th observation

Referring to cf column, we find that the 75th observation lies in the interval 600 - 799. Hence, class of Q_3 is 600 - 799 which is inclusive type. \therefore The class boundaries of class of Q_3 are 599.5 - 799.5.

$$\text{Now, } Q_3 = L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c$$

Here, $L = 599.5$, $3\left(\frac{n}{4}\right) = 75$, $cf = 40$, $f = 42$ and $c = 200$

$$\begin{aligned} Q_3 &= 599.5 + \frac{75-40}{42} \times 200 \\ &= 599.5 + \frac{35 \times 200}{42} \\ &= 599.5 + 166.6667 \\ &= 766.1667 \end{aligned}$$

$$\therefore Q_3 \approx 766.17 \text{ cheque}$$

$$\begin{aligned} \text{Coefficient of skewness } j &= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \\ j &= \frac{766.17 + 423.03 - 2(647.12)}{766.17 - 423.03} \\ &= \frac{1189.20 - 1294.24}{343.14} \\ &= \frac{-105.04}{343.14} \\ &= -0.3061 \\ j &\approx -0.31 \end{aligned}$$

Thus, we say that the frequency distribution has negative skewness.

Illustration 14 : Find the coefficient of skewness for the frequency distribution of sales of 500 companies in the year 2014-2015 using an appropriate method.

Sales (thousand tonnes)	Less than 4	4 - 7	7 - 10	10 - 13	13 - 16	16 - 20	20 and above
No. of companies	26	119	198	86	39	20	12

As the distribution has open-ended classes, measure of skewness can be obtained only by Bowley's method. We will compute first quartile, median and third quartile.

Sales (thousand tonnes)	No. of companies f	cf
Less than 4	26	26
4-7	119	145
7-10	198	343
10-13	86	429
13-16	39	468
16-20	20	488
20 and above	12	500
Total	$n = 500$	

First quartile Q_1 = Value of the $\left(\frac{n}{4}\right)$ th observation

$$= \text{Value of the } \left(\frac{500}{4}\right) \text{ th observation}$$

$$= \text{Value of the 125th observation}$$

Referring to cf column, we find that the 125th observation lies in the interval 4 - 7. Hence, class of Q_1 is 4 - 7.

$$\text{Now, } Q_1 = L + \frac{\frac{n}{4} - cf}{f} \times c$$

$$\text{Here, } L = 4, \frac{n}{4} = 125, cf = 26, f = 119 \text{ and } c = 3$$

$$Q_1 = 4 + \frac{125 - 26}{119} \times 3$$

$$= 4 + \frac{297}{119}$$

$$= 4 + 2.4958$$

$$= 6.4958$$

$$\therefore Q_1 = 6.50 \text{ thousand tonnes}$$

Median M = Value of the $\left(\frac{n}{2}\right)$ th observation

$$= \text{Value of the } \left(\frac{500}{2}\right) \text{ th observation}$$

$$= \text{Value of the 250th observation}$$

Referring to cf column, we find that the 250th observation lies in the interval 7 - 10. Hence class of median is 7 - 10.

$$\text{Now, } M = L + \frac{\frac{n}{2} - cf}{f} \times c$$

$$\text{Here, } L = 7, \frac{n}{2} = 250, cf = 145, f = 198 \text{ and } c = 3$$

$$\begin{aligned} M &= 7 + \frac{250 - 145}{198} \times 3 \\ &= 7 + \frac{105 \times 3}{198} \\ &= 7 + \frac{315}{198} \\ &= 7 + 1.5909 \\ &= 8.5909 \end{aligned}$$

$\therefore M \approx 8.59$ thousand tonnes

$$\begin{aligned} \text{Third quartile } Q_3 &= \text{Value of the } 3\left(\frac{n}{4}\right) \text{ th observation} \\ &= \text{Value of the } 3\left(\frac{500}{4}\right) \text{ th observation} \\ &= \text{Value of the } 3(125) \text{ th observation} \\ &= \text{Value of the } 375 \text{ th observation} \end{aligned}$$

Referring to cf column, we find that the 375th observation lies in the interval 10 - 13.

Hence, class of Q_3 is 10 - 13.

$$\text{Now, } Q_3 = L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c$$

$$\text{Here, } L = 10, 3\left(\frac{n}{4}\right) = 375, cf = 343, f = 86 \text{ and } c = 3$$

$$\begin{aligned} Q_3 &= 10 + \frac{375 - 343}{86} \times 3 \\ &= 10 + \frac{32 \times 3}{86} \\ &= 10 + \frac{96}{86} \\ &= 10 + 1.1163 \\ &= 11.1163 \end{aligned}$$

$\therefore Q_3 \approx 11.12$ thousand tonnes

$$\begin{aligned} \text{Coefficient of skewness } j &= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \\ j &= \frac{11.12 + 6.5 - 2(8.59)}{11.12 - 6.5} \\ &= \frac{17.62 - 17.18}{4.62} \\ &= \frac{0.44}{4.62} \\ &= 0.0952 \\ \therefore j &\approx 0.10 \end{aligned}$$

Thus, we say that the frequency distribution has positive skewness. We can say that the distribution is close to symmetry as the coefficient is close to zero.

Illustration 15 : For a frequency distribution of monthly overtime (in hours) of employees of a company, the difference between the two extreme quartiles is 50 and their sum is 218. If the median is 112, find the coefficient of skewness.

Here, $Q_3 + Q_1 = 218$, $Q_3 - Q_1 = 50$, $M = 112$

$$\begin{aligned} \text{Coefficient of skewness } j &= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \\ &= \frac{218 - 2(112)}{50} \\ &= \frac{218 - 224}{50} \\ &= \frac{-6}{50} \\ j &= -0.12 \end{aligned}$$

Illustration 16 : Mode of a symmetric frequency distribution is 84. If the first quartile is 68, find the third quartile.

As the distribution is symmetric, we have $j = 0$ and $\bar{x} = M = M_o$

Thus, $M = M_o = 84$, $M = 84$ and $Q_1 = 68$

$$\begin{aligned} j &= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \\ \therefore 0 &= \frac{Q_3 + 68 - 2(84)}{Q_3 - 68} \\ \therefore 0 (Q_3 - 68) &= Q_3 + 68 - 168 \\ \therefore 0 &= Q_3 - 100 \\ \therefore Q_3 &= 100 \end{aligned}$$

Illustration 17 : The following information is available about marks in a subject obtained by the students of a school in the annual examination. 25% students have scored less than 28 marks whereas, 75% students have scored less than 47 marks. If the coefficient of skewness for marks is 0.4, find the median.

25 % observations have value less than 28 $\therefore Q_1 = 28$

75% observations have value less than 47 $\therefore Q_3 = 47, j = 0.4$

$$\text{Now, } j = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$\therefore 0.4 = \frac{47 + 28 - 2M}{47 - 28}$$

$$\therefore 0.4 = \frac{75 - 2M}{19}$$

$$\therefore 0.4 \times 19 = 75 - 2M$$

$$7.6 = 75 - 2M$$

$$\therefore 2M = 75 - 7.6$$

$$\therefore 2M = 67.4$$

$$\therefore M = 33.7 \text{ marks}$$

5.5 Comparison of two methods of Coefficient of Skewness

The values of coefficient of skewness computed by Karl Pearson's method and Bowley's method are generally not same as they are based on different averages. Mean, median and mode are used as averages along with standard deviation in Karl Pearson's method. On the other hand, positional averages namely quartiles are used in Bowley's method. Thus both the methods are based on different averages.

The calculation of coefficient of skewness by Bowley's method is easier than the calculation by Karl Pearson's method.

The value of coefficient of skewness obtained by Karl Pearson's method is more reliable than by Bowley's method as the coefficient by Bowley's method is based on quartiles. Quartiles are found using only central 50% observations of the data. Mean and standard deviation are used for the coefficient of skewness by Karl Pearson's method. These measures are based on all observations.

Skewness and its coefficient can only be found by Bowley's method for the frequency distribution with open-ended classes. Mean and standard deviation used in Karl Pearson's method cannot be found for a distribution with open-ended classes. In Karl Pearson's method, the absolute measure $\bar{x} - M_o$ for skewness is divided by standard deviation to obtain the coefficient of skewness whereas, the absolute measure of skewness $(Q_3 - M) - (M - Q_1)$ is divided by $(Q_3 - Q_1)$ to find its coefficient by Bowley's method. We should also note that it is inappropriate to compare the values obtained by Karl Pearson's formula to the one computed by Bowley's formula. If the coefficient of skewness $j = 0$, it indicates the absence of skewness which means that the distribution is symmetric. Although it is possible in some cases that the frequency curve of the distribution has positive or negative skewness even if $j = 0$.

Illustration 18 : The measures from the data about the monthly income of residents of two cities are obtained as follows. Find coefficient of skewness using Karl Pearson's and Bowley's method from these data.

Details	City A	City B
Mean \bar{x}	300	280
Median M	284	310
Standard deviation s	60	110
First quartile Q_1	124	160
Third quartile Q_3	390	520

City A :

Coefficient of skewness for Karl Pearson's method :

$$\begin{aligned}
 j &= \frac{3(\bar{x} - M)}{s} \\
 &= \frac{3(300 - 284)}{60} \\
 &= \frac{3(16)}{60} \\
 &= \frac{48}{60} \\
 j &= 0.8
 \end{aligned}$$

Bowley's method

$$\begin{aligned}
 j &= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \\
 &= \frac{390 + 124 - 2(284)}{390 - 124} \\
 &= \frac{514 - 568}{266} \\
 &= \frac{-54}{266} \\
 &= -0.203 \\
 j &= -0.2
 \end{aligned}$$

City B :

Coefficient of skewness for Karl Pearson's method :

$$\begin{aligned}
 j &= \frac{3(\bar{x} - M)}{s} \\
 &= \frac{3(280 - 310)}{110} \\
 &= \frac{3(-30)}{110} \\
 &= \frac{-9}{11} \\
 \therefore j &= -0.82
 \end{aligned}$$

Bowley's method

$$\begin{aligned}
 j &= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \\
 &= \frac{520 + 160 - 2 \times 310}{520 - 160} \\
 &= \frac{680 - 620}{360} \\
 &= \frac{60}{360} \\
 \therefore j &= 0.17
 \end{aligned}$$

The coefficient of skewness for city A is positive by Karl Pearson's method and negative by Bowley's method. On the other hand, it is negative by Karl Pearson's method and positive by Bowley's method for city B. Hence it is not appropriate to compare the coefficients obtained by two methods. It is advisable to compare using only one method.

EXERCISE 5.2

1. From the following frequency distribution of different youngsters exercising in a gymnasium, find coefficient of skewness by Bowley's method and state the type of skewness.

Age (years)	25	17	20	18	26	22	28	23
No. of youngsters	22	4	19	11	7	9	3	8

2. The frequency distribution of paid up share capital out of the issued share capital for 31 manufacturing companies is as follows. Find skewness and its coefficient by Bowley's method and state the type of skewness.

Paid up share capital (lakh ₹)	Less than 100	Less than 300	Less than 500	Less than 700	Less than 900	Less than 1100	Less than 1300
No. of companies	0	6	16	19	23	27	31

3. The distribution of sales (in thousand tonnes) of 400 companies during the year 2014 - 15 is as follows. Find skewness and its coefficient from these data and state the type of skewness.

Sales (thousand tonnes)	Less than 20	20-40	40-50	50-75	75-90	90-120	120 and above
No. of companies	30	70	125	100	40	20	15

4. The commission paid on insurance policy amount to agents in a branch of an insurance company during a month has the following frequency distribution. Find the coefficient of skewness by Bowley's method.

Commission paid (thousand Rs.)	10-12	12-14	14-16	16-18	18-20	20-22	22-24	24-26	26-28	28-30
No. of companies	4	10	16	29	52	80	32	23	17	1

*

Summary

- Generally there are two types of frequency distributions:
(1) Symmetric distribution (2) Skewed distribution
- Values of mean, median and mode are equal in a symmetric distribution and generally it has bell-shaped frequency curve.
- The right or left tail of the frequency curve of a skewed distribution is more elongated.
- The frequency curve with left tail more elongated is called negatively skewed curve and the one with right tail more elongated is called positively skewed curve.
- Two types of measures are used for measuring skewness:
(1) Absolute measure (skewness) (2) Relative measure (coefficient of skewness)
- Karl Pearson's method or Bowley's method is used for measuring skewness.
- Relative measure of skewness is called coefficient of skewness.
- The coefficient of skewness generally lies between - 1 and 1. In some specific cases, it lies between - 3 and 3.

List of formulae :

Formulae for Karl Pearson's method	Formulae for Bowley's method
<p style="text-align: center;">Absolute measure</p> <p>(1) For well-defined mode Skewness $S_k = \bar{x} - M_o$</p> <p>(2) For multiple modes or for ill-defined mode Coefficient of skewness $S_k = 3 (\bar{x} - M)$</p>	<p style="text-align: center;">Absolute measure</p> <p>Skewness $S_k = (Q_3 - M) - (M - Q_1)$ $= Q_3 + Q_1 - 2M$</p>
<p style="text-align: center;">Relative measure</p> <p>(1) For well-defined mode Coefficient of skewness $j = \frac{\bar{x} - M_o}{s}$</p> <p>(2) For multiple modes or for ill-defined mode Coefficient of skewness $j = \frac{3(\bar{x} - M)}{s}$</p>	<p style="text-align: center;">Relative measure</p> <p>Coefficient of skewness $j = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)}$ $j = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$</p>

EXERCISE 5

Section A

Find the correct option for the following multiple choice questions:

- Generally, what is the range of coefficient of skewness for data where the mode is ill-defined?
(a) 0 to 1 (b) -1 to +1 (c) -3 to +3 (d) 1 to 0
- For a frequency distribution having negative skewness, what will be the value of its mean?
(a) More than mode (b) Less than mode
(c) Equal to mode (d) Nothing can be said about mean
- The following measures are obtained for two distributions.
Distribution (i) Mean = 44, Median = 48 and Standard deviation = 20
Distribution (ii) Mean = 44, Median = 50 and Standard deviation = 24

Which of the following statements is true ?

- Distributions (i) and (ii) have same degree of skewness.
- Distribution (i) has more skewness than distribution (ii).
- Distribution (i) has less skewness than distribution (ii).
- Nothing can be said about skewness from the given data.

4. Two measures of central tendency are given for the following three frequency distributions. All of them are unimodal distributions. State the type of skewness for the three distributions.
- (i) Distribution A: mode = 100 and mean = 116
(ii) Distribution B: median = 142.8 and mean = 142.8
(iii) Distribution C: median = 208 and mean = 192
- (a) A is symmetric, B is negatively skewed and C is positively skewed.
(b) A is negatively skewed, B is positively skewed and C is symmetric.
(c) A is positively skewed, B is symmetric and C is negatively skewed.
(d) A is positively skewed, B is negatively skewed and C is symmetric.
5. Mode of a frequency distribution exceeds its mean by 2. What type of distribution is it ?
(a) Negatively skewed (b) symmetric (c) positively skewed (d) nothing can be said.
6. If $Q_3 + Q_1 = 60$ and $M = 30$ for a frequency distribution, which of the following statements about its skewness is true ?
(a) Distribution is highly skewed. (b) Distribution is less skewed.
(c) Distribution has lack of symmetry. (d) Distribution is symmetric.
7. In a moderately skewed distribution, (mean – mode) will be how many times (mean – median) ?
(a) 3 (b) –1 (c) $\frac{1}{3}$ (d) 0
8. Which of the following statements is false for a negatively skewed frequency distribution ?
(a) Value of mean is less than median and mode.
(b) The distance between the third quartile and median is less than the distance between median and the first quartile.
(c) The left tail of the frequency curve of the distribution is more elongated.
(d) The distance between the third quartile and median is more than the distance between median and the first quartile.
9. Which of the following statements is true for a symmetric distribution ?
(a) $Q_3 = 2M - Q_1$ (b) $Q_2 - Q_3 = Q_2 - Q_1$ (c) $Q_3 + Q_1 > 2M$ (d) $Q_3 + Q_1 < 2M$
10. If $(M - \bar{x}) = -\frac{1}{2}s$, find the value of j .
(a) $-\frac{1}{3}$ (b) $\frac{3}{2}$ (c) = –1.5 (d) 0.15
11. Which measure of central tendency has higher value in (i) negatively skewed distribution and (ii) positively skewed distribution?
(a) (i) mean (ii) mode (b) (i) median (ii) mode
(c) (i) mode (ii) mean (d) nothing can be said about mean, median and mode
12. If coefficients of skewness for distribution X is –0.99 and that of distribution Y is 0.90. Which of the following statements is true ?
(a) Distribution X is more skewed. (b) Distribution Y is more skewed.
(c) Nothing can be said about skewness of X and Y. (d) Distributions X and Y have same skewness.

13. Which of the following statements is false ?
- If $Q_3 + Q_1 > 2M$ the distribution is positively skewed.
 - Bowley's coefficient of skewness is found using positional averages.
 - The absolute measure is divided by standard deviation in Karl Pearson's method to eliminate the effect of unit of the variable whereas in Bowley's method, the absolute measure is divided by the difference of quartiles.
 - The frequencies of observations which are equidistant on both sides of the mode are equally distributed in a skewed distribution.
14. Which of the following statements is true ?
- The distribution in which observations on both sides of mode are equally distributed is called negatively skewed distribution.
 - The distances of median from the extreme quartiles are same in a symmetric distribution.
 - If $S_k > 0$, $\bar{x} > M$ and $\bar{x} < M_o$
 - If $S_k < 0$, $\bar{x} < M$ and $\bar{x} > M_o$

Section B

Answer the following questions in one sentence :

- What is meant by skewness ?
- When do we call a frequency distribution to be symmetric ?
- When do we say that a frequency distribution is skewed ?
- What can you say about the position of median in a skewed frequency distribution ?
- Explain how to determine the skewness using a frequency curve.
- What is coefficient of skewness ? State the range for its value.
- Which measures are used to obtain Karl Pearson's measure of skewness ?
- State the basic assumption of obtaining Bowley's measure of skewness.
- Which method gives more reliable value of coefficient of skewness ?
- Coefficient of skewness is absolute measure or relative measure. Give reasons.
- Which formula is used for finding coefficient of skewness when a distribution has open-ended classes ?
- What is the range for coefficient of skewness computed using Karl Pearson's method for a frequency distribution with unequal class intervals ?
- State the type of skewness for a frequency distribution whose three quartiles are 42, 36 and 40.
- State the type of skewness for a frequency distribution where $(Q_3 - Q_2) < (Q_2 - Q_1)$.
- The mean of a frequency distribution is less than its median by 2 units. State its type of skewness.
- If $Q_3 + Q_1 = 125$ and $M = 62.5$, for a set of data, what can be said about its skewness ?
- If $\bar{x} = M = M_o = 48$ in a frequency distribution, what can you say about its coefficient ?

Section C

Answer the following questions :

1. Explain the types of skewness.
2. Show the positions of averages and quartiles using a diagram for each types of skewness.
3. State any two characteristics of symmetric distribution.
4. State any two characteristics of skewed distribution.
5. For a frequency distribution, skewness $S_k = -2.8$. If its mode is 48.8, find mean.
6. Sum of two extreme quartiles is 138 in a symmetric frequency distribution. Find its median.
7. The coefficient of a skewed frequency distribution is 0.75. If its standard deviation is 20 and mean is 37.5, find median.
8. The mean of a skewed distribution exceeds its median by 3. If its coefficient of skewness is 0.75, find standard deviation.
9. If a frequency distribution has $Q_3 - Q_2 = 2(Q_2 - Q_1)$, find j .
10. If a frequency distribution has skewness $S_k = -6.6$ and quartiles deviation = 22, find j .
11. A skewed frequency distribution has mean = 40, mode = 46, $Q_3 + Q_1 = 76$ and $Q_3 - Q_1 = 20$. Find Bowley's coefficient of skewness.
12. 'The coefficient of skewness obtained by Bowley's method is not more reliable than the coefficient by Karl Pearson's method.' Explain this statement.
13. The three quartiles of a frequency distribution are 76, 98 and 40. Find j and state the type of skewness.
14. The coefficient of skewness of a distribution is 0.85. If its mean is 3.4 more than its mode, find its variance.
15. If a frequency distribution has $\bar{x} + M_o = 82$, $\bar{x} = 44$ and $s = 12$, find coefficient of skewness.

Section D

Answer the following questions :

1. Briefly explain Karl Pearson's method of finding the coefficient of skewness.
2. Write a short note on Bowley's method of finding the coefficient of skewness.
3. State the circumstances in which Karl Pearson's formula $j = 3 \left(\frac{\bar{x} - M}{s} \right)$ is used to find coefficient of skewness.
4. Differentiate between positive and negative skewness with details and using a diagram.
5. Which of the following populations is closer to symmetry ?
Population A : $\bar{x} = 56$, $M_o = 60$ and $s = 24$
Population B : $\bar{x} = 56$, $M = 60$ and $s = 30$
6. Find coefficient from the following data using an appropriate method and determine which population is more skewed among A and B.
Population A : $4Q_1 = 3Q_2 = 2Q_3 = 144$
Population B : $Q_1 = 34.8$, $Q_2 = 45.5$ and $Q_3 = 70$
7. Third quartile is at a distance of 12.8 from the median in a frequency distribution and its first quartile is at a distance of 11.2 from the median. Find skewness and its coefficient.

8. If coefficient of variation is 25 %, $\bar{x} = 32$ and $M_o = 32.2$ for a set of data, find its coefficient of skewness.
9. Find coefficient of skewness from the following data :
 $n = 20$, $\Sigma x = 640$, $\Sigma x^2 = 20,800$ and $M = 32.2$
10. Karl Pearson's coefficient of skewness for a data set is -0.6 . If mean = 60 and $s = 10$, find median and mode for the data.
11. Karl Pearson's skewness for a frequency distribution is 8 and coefficient of skewness is $= \frac{2}{3}$. If the mean is 64, find its median and coefficient of variation.
12. Find the coefficient of skewness for a frequency distribution with $Q_3 + Q_1 = 1.5 M$ and $3(Q_3 - Q_1) = 2M$
13. A frequency distribution has $4\bar{x} = 6M_o = 144$, $s = 64$ and $Q_3 + Q_1 = 6(Q_3 - Q_1) = 60$. Find the coefficient using Karl Pearson's and Bowley's method.

Section E

Answer the following questions :

1. Define skewed frequency distribution and state its characteristics.
2. Define symmetric frequency distribution and state its characteristics.
3. Differentiate: Karl Pearson's and Bowley's method for coefficient of skewness.
4. Explain skewness and coefficient of skewness.
5. What are the main objectives of studying skewness ?
6. Various measures for the frequency distributions of monthly salaries of two production firms are as follows. Compare the firms using coefficient of skewness by Bowley's and Karl Pearson's method.

Details	Mean	Median	First quartile	Third quartile	Standard deviation
Firm A	350	344	324	356	26
Firm B	360	340	330	370	38

7. The frequency distribution of sale of notebooks from a stationary shop in the month of June for the year 2014 is as follows. Find coefficient of skewness using Karl Pearson's method.

Sale of notebooks (dozens)	30	25	21	20	18	16	15	12
No. of days	2	2	7	3	4	7	2	3

8. The following information is available about defective staplers after testing 50 packets of 500 staplers each. Find coefficient of skewness using Karl Pearson's method.

No. of defective staplers	19	20	21	22	23	24	25	26
No. of packets	5	18	10	8	4	2	2	1

9. For the frequency distribution of a set of data, $n = 200$, $\Sigma f(x - 240) = 0$, $\Sigma f(x - 240)^2 = 11,250$ and median = 246, find coefficient of skewness and state the type of skewness.

Section F

Solve the following :

1. The frequency distribution of number of accidents due to driving for more than the prescribed time is given below. Find coefficient of skewness by Bowley's method.

No. of driving hours more than prescribed time	4	3.5	3	2.5	2	1.5	1	0
No. of accidents	5	4	3	2	1	2	2	1

2. The daily temperature of a city in the year 2014 is recorded as follows. The daily temperature has not been below -10° C. Find Karl Pearson's coefficient of skewness. State the type of skewness.

Mid-value (Celsius)	-5	5	12	18	25
No. of days	25	35	105	125	75

3. The distribution of marks obtained by 60 students in an examination is as follows. Find coefficient of skewness by Karl Pearson's method and state the type of skewness.

Marks of students	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	Total
No. of students	5	12	38	38	20	7	120

4. The frequency distribution of profits earned by 150 companies during the year 2015-2016 is as follows. Find the coefficient of skewness using an appropriate method and state the type of skewness.

Profit (lakh ₹)	Less than 10	10 - 20	20 - 30	30 - 40	40 and above
No. of companies	15	30	50	40	15

5. The frequency distribution of demand of a certain item is as follows. Find skewness and coefficient of skewness by Karl Pearson's method.

Demand (units)	1	2	3	4 - 8	8 - 12	12 - 16	16 - 20
No. of customers	10	8	12	10	5	15	20

6. A sample of 50 screws was taken from the lots of screws produced at a factory to measure the diameter (in mm) of head of each screw and its frequency distribution is as follows. Find Bowley's coefficient of skewness and interpret it.

Diameter of head of screw (mm)	4 - 4.1	4 - 4.2	4 - 4.3	4 - 4.4	4 - 4.5	4 - 4.6	4 - 4.7	4 - 4.8
No. of screws	6	13	23	33	41	46	48	50

7. From the following frequency distribution of daily sales of packets of bread from a departmental store, find coefficient of skewness by Karl Pearson's method and interpret it.

No. of bread packets	0 - 3	3 - 5	5 - 10	10 - 15	15 - 20	20 - 30	30 - 40	40 - 60
No. of customers	15	12	8	6	4	3	2	1

8. The frequency distribution of sales from a shop selling glass is as follows. Find Bowley's coefficient of skewness and interpret it.

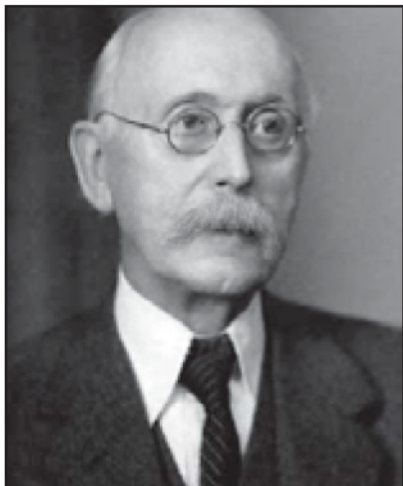
Size of glass (sq. m.)	1 - 1.9	2 - 2.9	3 - 3.9	4 - 4.9	5 - 5.9	6 - 6.9	7 - 7.9
No. of customers	10	40	20	50	30	30	20

9. A construction company builds houses with different areas. The frequency distribution of areas of houses is as follows. Find coefficient of skewness by Karl Pearson's method and interpret it.

Area of house (sq.m.)	100	140	180	220	260
No. of houses	10	25	50	25	10

10. The frequency distribution of units of power consumed in an hour for different machines during a production process of factory is as follows. Find coefficient of skewness by Bowley's method.

Units of power consumed	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40
No. of machines	5	10	15	20	25	30



Arthur Lyon Bowley
(1869 - 1957)

Sir Arthur Lyon Bowley was a British Statistician and Mathematical Economist. Among the distinguished posts he held were those of Professor of Mathematics, Economics and Statistics at University College, Reading and London and Director of University of Oxford Institute of Statistics. His major works were "Three studies on the National Income", published in 1938 and "Wages and Income in the UK since 1860", published in 1937. He occupied, at various times, high positions in the Royal Statistical Society, The Royal Economic Society (elected fellow in 1893), the International Statistical Institute, the Econometric Society and the British Association for the Advancement of Science.

“The nature of the infant is not just a new permutation - and - combination of elements contained in the natures of the parents. There is in the nature of the infant that which is utterly unknown in the natures of the parents.”

– David Herbert Lawrence

6

Permutations, Combinations and Binomial Expansion

Contents :

6.1 Permutation : Meaning

6.2 Combination : Meaning

6.3 Binomial Expansion : Meaning and Characteristics

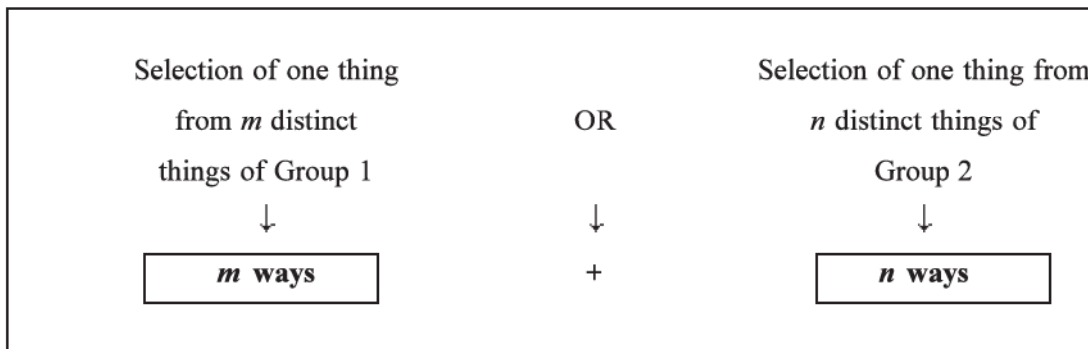
Permutation and Combination are useful for solving many problems in our day to day life. e.g. 4 students are to be arranged on a bench of a particular class then in how many ways can a teacher arrange these 4 students ? A person has 6 friends and he wishes to invite only 2 friends to his family function then how many options that person has. Normally, we solve these problems with common sense. But for the mathematical solution for different problems of this type, we shall study certain principles and methods in this chapter.

First of all we shall study two types of fundamental principles of counting namely addition and multiplication with reference to permutation and combination.

Fundamental principle of counting for addition :

We shall study some examples to understand this principle. A restaurant serves 4 types of pizzas (say P_1, P_2, P_3, P_4) and 3 types of burgers (say B_1, B_2, B_3). If a person wants to order a pizza or a burger then he has total $4 + 3 = 7$ $\{P_1, P_2, P_3, P_4, B_1, B_2, B_3\}$ options. In the same way,

a class teacher wishing to appoint a class representative from 40 boys and 20 girls studying in a class has $40 + 20 = 60$ options. Thus, if there are m distinct things in Group 1 and n distinct things in Group 2 then selection of one thing from total things of both groups can be done in $m + n$ ways. This rule is called fundamental principle of counting for addition.

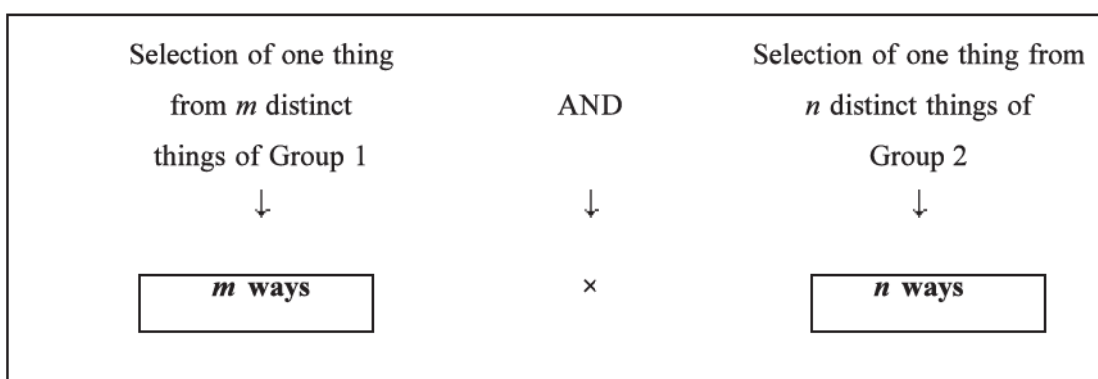


It means the word ‘OR’ indicates addition.

Note : Fundamental principle of counting for addition can also be applicable to more than two groups.

Fundamental Principle of Counting for Multiplication :

If the first operation can be done in m ways and second operation can be done in n ways, then two operations together can be done in $m \times n$ ways. This rule is called fundamental principle of counting for multiplication. We shall understand this rule using the previous example. If a person wants to order a pizza and a burger then he has total $4 \times 3 = 12$ $\{P_1B_1, P_1B_2, P_1B_3, P_2B_1, P_2B_2, P_2B_3, P_3B_1, P_3B_2, P_3B_3, P_4B_1, P_4B_2, P_4B_3\}$ options. In the same way, a class teacher wishing to appoint a boy and a girl as a class representative has $40 \times 20 = 800$ options.



It means the word ‘AND’ indicates multiplication.

Note : Fundamental principle of counting for multiplication can also be applicable to more than two groups.

6.1 Meaning of Permutation

Let us begin with some examples to understand permutation.

- How many different two digit numbers can be made using digits 2, 5 and 8 without repeating digits ?

If we make two digit numbers using given digits 2, 5, 8 then total 6 numbers 25, 28, 52, 58, 82, 85 can be made. It is necessary to note that the numbers 22, 55, 88 can not be formed as repetition is not allowed.

Now, we shall solve this problem using principle of counting. In two digit numbers, there are two places, unit and tens.

Arrange any one of the digits 2, 5 and 8

Tens Place

Now, excluding the digit placed in tens place, arrange any one digit here from the remaining two digits.

Unit place

Thus, there are three options 2, 5 and 8 for filling tens place. After filling tens place with any one digit there are only two options for filling unit place. That is total $3 \times 2 = 6$ numbers can be formed. Selecting digit here for unit place first and then selecting digit for tens place will not change the result.

Note : Permutations with repetition of items is not included in our syllabus.

- From 4 students A, B, C, D, one student is to be selected as captain and other student as vice captain of cricket team. How many alternative solutions are possible for this problem ?

If this problem is solved using common sense then following 12 different alternatives can be obtained :

Altern-ative	Captain	Vice captain
1	A	B
2	A	C
3	A	D
4	B	A
5	B	C
6	B	D
7	C	A
8	C	B
9	C	D
10	D	A
11	D	B
12	D	C

Now we shall solve this problem using principle of counting.

Any one student can be made captain from 4 students A, B, C, D

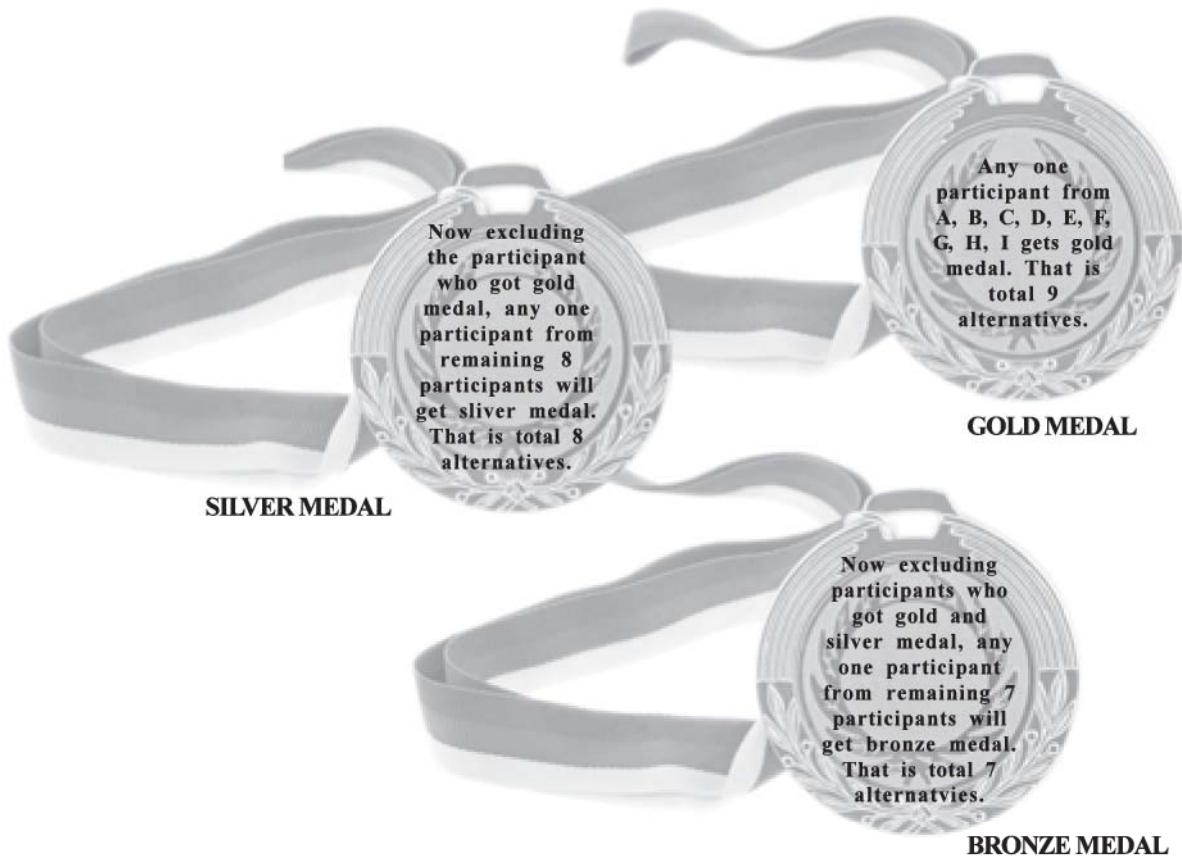
Captain's place

Now, excluding the student who is made captain, any one student can be made vice captain from remaining three students.

Vice-Captain's Place

Thus, there are four alternatives A, B, C and D for selection of a captain. There are three alternatives for selection of vice-captain after selecting captain. That is total $4 \times 3 = 12$ alternatives can be obtained.

- Nine participants A, B, C, D, E, F, G, H and I take part in a 100 meter running competition. Top three participants are given Gold, Silver and Bronze medals. Now, we have to find the number of alternatives for participants getting three medals.



Thus, according to fundamental principle of counting, the first action (Gold medal) can be in 9 ways, the second action (Silver medal) can be in 8 ways and the third action (Bronze medal) can be in 7 ways. Thus, total number of alternatives of all the three actions together are $9 \times 8 \times 7 = 504$. The solution of this problem can also be expressed in following way :

Medal	Gold	Silver	Bronze
Alternatives of participants getting medal	9	$9 - 1 = 8$	$9 - 2 = 7$

According to the fundamental principle of counting, total alternatives are $= 9 \times 8 \times 7 = 504$

(It is important to note here that there is no tie between the participants)

If r distinct things out of given n distinct things are to be arranged in r ($1 \leq r \leq n$) different places, then each such arrangement is called a permutation. The total number of such arrangements is denoted as ${}^n P_r$, ${}_n P_r$, $P(n, r)$, P_r^n . We will use the notation ${}^n P_r$.

Thus, total permutations of n distinct things in r places will be ${}^n P_r$

Understand the next table defining ${}^n P_r$.

Suppose r distinct things out of n different things are to be arranged in r places, i.e. we want to find ${}^n P_r$. This arrangement can be made as follows. (For easy understanding, compare the next table with the previous example of medals.)

<i>r</i> Places	1	2	3	4	...	<i>r</i>
Possible ways for each place	n $= n - (1-1)$	$n - 1$ $= n - (2-1)$	$n - 2$ $= n - (3-1)$	$n - 3$ $= n - (4-1)$...	$n - r + 1$ $= n - (r-1)$

It is clear from the above table that for filling the first place there are n ways available. Now, having filled the first place by any one thing, the second place can be filled by any one of the remaining $(n-1)$ things. Thus the total number of ways of filling the first two places is $n \times (n-1)$. Similarly the total number of ways of filling the first three places is $n \times (n-1) \times (n-2)$. Similarly, the total number of ways of filling all the r places is $n \times (n-1) \times (n-2) \times (n-3) \times \dots \times (n-r+1)$. This is called total permutations of arranging r things out of n different things in r places. So,

$${}^n P_r = n \times (n-1) \times (n-2) \times (n-3) \times \dots \times (n-r+1)$$

No. of ways of filling each place is one less than the no. of ways of filling previous place.

Factorial has a very important role in the study of permutations. First of all, let us understand the meaning of factorial. We have seen earlier that multiplication of consecutive numbers is involved while solving many problems. Factorial is used to express this multiplication in short. Factorial is denoted with symbol ‘!’ or ‘ $_$ ’. Factorial n is denoted by $n!$ or $_n$. Now let us understand the meaning of $n!$.

The product of natural numbers from 1 to n is $n!$, i.e.

$$n! = 1 \times 2 \times 3 \times \dots \times (n-1) \times n$$

OR

$$n! = n \times (n-1) \times \dots \times 3 \times 2 \times 1 \text{ (We will write this way for the sake of convenience)}$$

For example,

$$\begin{aligned} 7! &= 7 \times (7-1) \times \dots \times 3 \times 2 \times 1 \\ &= 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 \\ &= 5040 \end{aligned}$$

$$\begin{aligned} 10! &= 10 \times (10-1) \times \dots \times 3 \times 2 \times 1 \\ &= 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 \\ &= 36,28,800 \end{aligned}$$

Similarly,

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

$$3! = 3 \times 2 \times 1 = 6$$

$$1! = 1$$

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

$$= 6 \times 5 \times 4 \times 3 \times 2!$$

$$= 6 \times 5 \times 4 \times 3!$$

$$= 6 \times 5 \times 4!$$

$$= 6 \times 5!$$

See the utility...!

$$\frac{6!}{4!} = \frac{6 \times 5 \times 4!}{4!} = 30$$

Generalizing this, we get

$$\begin{aligned} n! &= n(n-1)(n-2) \times \dots \times 3 \times 2 \times 1 \\ &= n(n-1)(n-2) \times \dots \times 3 \times 2! \\ &= n(n-1)(n-2) \times \dots \times 3! \\ &= n(n-1)(n-2)! \\ &= n(n-1)! \end{aligned}$$

$n! = (n)(n-1)!$ Putting $n = 1$ $\therefore 1! = (1)(1-1)!$ $\therefore 1 = 1 \times 0!$ $\therefore 0! = 1$

Now, let us see some important results.

As we have seen earlier, ${}^n P_r = n(n-1)(n-2) \dots (n-r+1)$. This result can also be written as follows :

$$\begin{aligned} {}^n P_r &= n(n-1)(n-2) \dots (n-r+1) \times \frac{(n-r)(n-r-1) \times \dots \times 3 \times 2 \times 1}{(n-r)(n-r-1) \times \dots \times 3 \times 2 \times 1} \\ &= \frac{n(n-1)(n-2) \dots (n-r+1)(n-r)(n-r-1) \times \dots \times 3 \times 2 \times 1}{(n-r)!} \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

$\therefore {}^n P_r = \frac{n!}{(n-r)!}$ where, $n > 0$, $r \geq 0$, $n \geq r$, n is a positive integer and r is non-negative integer.

The following results can be derived using this formula.

Some important results			
${}^n P_0 = 1$	${}^n P_n = n!$	${}^n P_1 = n$	${}^n P_{n-1} = n!$
${}^n P_r = \frac{n!}{(n-r)!}$	${}^n P_r = \frac{n!}{(n-r)!}$	${}^n P_r = \frac{n!}{(n-r)!}$	${}^n P_r = \frac{n!}{(n-r)!}$
$\therefore {}^n P_0 = \frac{n!}{(n-0)!}$	$\therefore {}^n P_n = \frac{n!}{(n-n)!}$	$\therefore {}^n P_1 = \frac{n!}{(n-1)!}$	$\therefore {}^n P_{n-1} = \frac{n!}{[n-(n-1)]!}$
$= \frac{n!}{n!}$	$= \frac{n!}{0!}$	$= \frac{n(n-1)!}{(n-1)!}$	$= \frac{n!}{(n-n+1)!}$
$= 1$	$= n!$	$= n$	$= n!$
e.g. ${}^5 P_0 = 1$ ${}^{10} P_0 = 1$	e.g. ${}^5 P_5 = 5!$ ${}^{10} P_{10} = 10!$	e.g. ${}^5 P_1 = 5$ ${}^{10} P_1 = 10$	e.g. ${}^5 P_4 = 5!$ ${}^{10} P_9 = 10!$

Illustration 1 : Find the values of the following :

- (1) ${}^8 P_3$ (2) ${}^{60} P_2$ (3) ${}^7 P_6$ (4) ${}^5 P_5$

(1) ${}^n P_r = \frac{n!}{(n-r)!}$

$$\begin{aligned} {}^8 P_3 &= \frac{8!}{(8-3)!} \\ &= \frac{8!}{5!} \\ &= \frac{8 \times 7 \times 6 \times 5!}{5!} \\ &= 336 \end{aligned}$$

Alternative method :

According to the definition of ${}^n P_r$,
 ${}^n P_r = n(n-1)(n-2) \dots (n-r+1)$
 Put $n = 8$ and $r = 3$
 $\therefore {}^8 P_3 = 8(8-1)(8-2)$
 $= 8 \times 7 \times 6$
 $= 336$

(2) According to the definition of ${}^n P_r$

$${}^n P_2 = n(n-1)$$

$$\begin{aligned}\therefore {}^{60} P_2 &= 60(60-1) \\ &= 60 \times 59 \\ &= 3540\end{aligned}$$

(3) According to the definition of ${}^n P_r$

$${}^n P_6 = n(n-1)(n-2)(n-3)(n-4)(n-5)$$

$$\begin{aligned}\therefore {}^7 P_6 &= 7 \times 6 \times 5 \times 4 \times 3 \times 2 \\ &= 5040\end{aligned}$$

(4) According to the definition of ${}^n P_r$

$${}^n P_5 = n(n-1)(n-2)(n-3)(n-4)$$

$$\begin{aligned}\therefore {}^5 P_5 &= 5 \times 4 \times 3 \times 2 \times 1 \\ &= 120\end{aligned}$$

Illustration 2 : If ${}^n P_3 = 720$ then find the value of n .

According to the definition of ${}^n P_r$

$${}^n P_3 = n(n-1)(n-2)$$

$$\therefore n(n-1)(n-2) = 720$$

Now, instead of expansion we shall think about the three consecutive numbers whose multiplication is 720. Such numbers are 10, 9, 8. But we shall write it as follows :

$$\therefore n(n-1)(n-2) = 10(10-1)(10-2)$$

$$\therefore n = 10$$

Illustration 3 : If ${}^7 P_r = 42$ then find the value of r .

$${}^n P_r = \frac{n!}{(n-r)!}$$

$$\therefore {}^7 P_r = \frac{7!}{(7-r)!}$$

$$\therefore 42 = \frac{7!}{(7-r)!}$$

$$\therefore (7-r)! = \frac{7!}{42}$$

$$\therefore (7-r)! = \frac{7 \times 6 \times 5!}{42}$$

$$\therefore (7-r)! = 5!$$

$$\therefore 7-r = 5$$

$$\therefore r = 2$$

Alternative Method :

$${}^7 P_r = 42$$

$$\therefore {}^7 P_r = 7 \times 6$$

$$\therefore {}^7 P_r = {}^7 P_2$$

$$\therefore r = 2$$

Illustration 4 : If $56 \times n! = 8!$ then find the value of n .

$$56 \times n! = 8 \times 7 \times 6!$$

$$\therefore n! = 6!$$

$$\therefore n = 6$$

Illustration 5 : If $\frac{n!}{2} = 60$ then find value the of n .

$$\frac{n!}{2} = 60$$

$$\therefore n! = 120 (= 1 \times 2 \times 3 \times 4 \times 5)$$

$$\therefore n! = 5!$$

$$\therefore n = 5$$

Illustration 6 : If ${}^{(n+2)}P_3 : {}^{(n+1)}P_3 = 10 : 7$ then find n .

$$\frac{{}^{(n+2)}P_3}{{}^{(n+1)}P_3} = \frac{10}{7}$$

$$\therefore \frac{(n+2)(n+1)(n)}{(n+1)(n)(n-1)} = \frac{10}{7}$$

$$\therefore \frac{n+2}{n-1} = \frac{10}{7}$$

$$\therefore 7(n+2) = 10(n-1)$$

$$\therefore 7n + 14 = 10n - 10$$

$$\therefore 3n = 24$$

$$\therefore n = 8$$

Illustration 7 : In how many ways 3 members of a family; husband, wife and a daughter can be arranged in a row for a group photograph ?

There are three members in this family. For a group photograph, they can be arranged in a row in 3P_3 ways.

$$\begin{aligned} \therefore \text{Total permutations} &= {}^3P_3 \\ &= 3! \\ &= 6 \end{aligned}$$

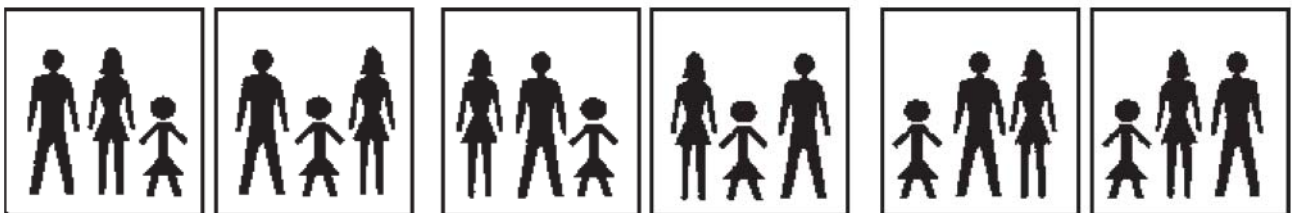


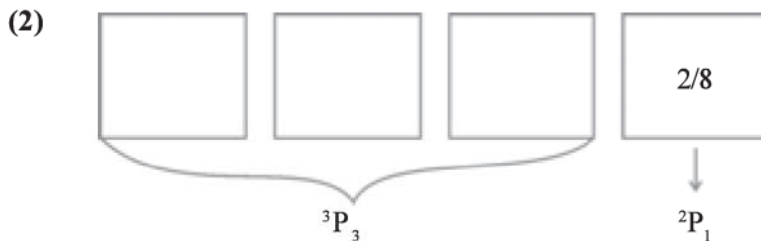
Illustration 8 : Using all the digits 2, 5, 7, 8, four digit numbers are to be formed.

- (1) How many such numbers can be made ?
 - (2) How many of them will be even numbers ?
 - (3) How many of them will be divisible by 5 ?
 - (4) How many of them will be greater than 5000 ?
- (1) Using all the digits 2, 5, 7, 8 i.e. 4 digits, total number of 4 digit numbers that can be formed is 4P_4 .

$$\begin{aligned} \therefore \text{Total permutations} &= {}^4P_4 \\ &= 4! \\ &= 24 \end{aligned}$$

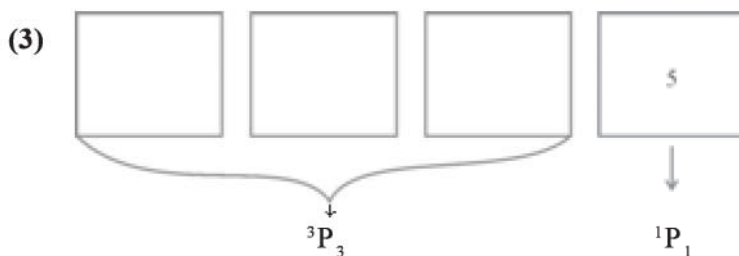
2578	5278	7258	8257
2587	5287	7285	8275
2758	5728	7528	8527
2785	5782	7582	8572
2857	5827	7825	8725
2875	5872	7852	8752

Verify total permutations in each question with the arrangement in this table.



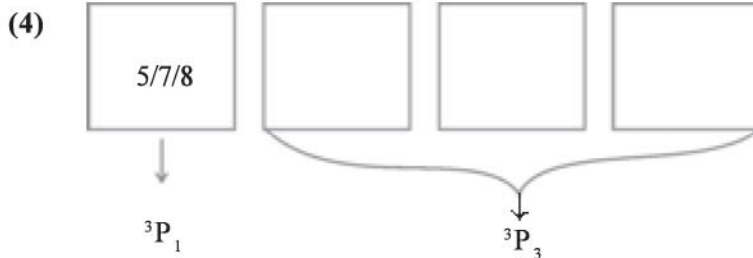
If even numbers are to be formed using all the digit 2, 5, 7, 8 then digit 2 or 8 should be in unit place. This arrangement can be done in 2P_1 ways. After arranging 2 or 8 in unit place, remaining 3 digits can be arranged in remaining 3 places in 3P_3 ways.

$$\begin{aligned} \therefore \text{Total Permutations} &= {}^2P_1 \times {}^3P_3 \\ &= 2 \times 3! \\ &= 2 \times 6 \\ &= 12 \end{aligned}$$



If numbers divisible by 5 are to be formed using all the digits 2, 5, 7, 8 then digit 5 should be in unit place. This arrangement can be done in 1P_1 ways. After arranging 5 in unit place, remaining 3 digits can be arranged in remaining 3 places in 3P_3 ways.

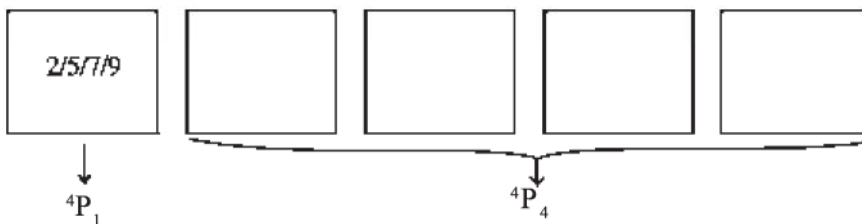
$$\begin{aligned}
 \therefore \text{Total Permutations} &= {}^1P_1 \times {}^3P_3 \\
 &= 1! \times 3! \\
 &= 1 \times 6 \\
 &= 6
 \end{aligned}$$



If numbers greater than 5000 are to be formed using all digits 2, 5, 7, 8 then digit 5, 7 or 8 should be in thousands place (first place). This arrangement can be done in 3P_1 ways. After arranging 5, 7 or 8 in thousands place, remaining 3 digits can be arranged in remaining 3 places in 3P_3 ways :

$$\begin{aligned}
 \therefore \text{Total Permutations} &= {}^3P_1 \times {}^3P_3 \\
 &= 3 \times 3! \\
 &= 3 \times 6 \\
 &= 18
 \end{aligned}$$

Illustration 9 : How many 5 digit numbers can be formed using all the digits 2, 5, 0, 7 and 9 ?



If 5 digit numbers are to be formed using all the digits 2, 5, 0, 7, 9 then digit 0 should not be in the first place. Therefore, excluding digit 0, one of the four digits can be placed in the first place in 4P_1 ways. Now, after arranging one digit from 2, 5, 7 or 9 in the first place, remaining 4 digits (including 0) can be arranged in remaining 4 places in 4P_4 ways.

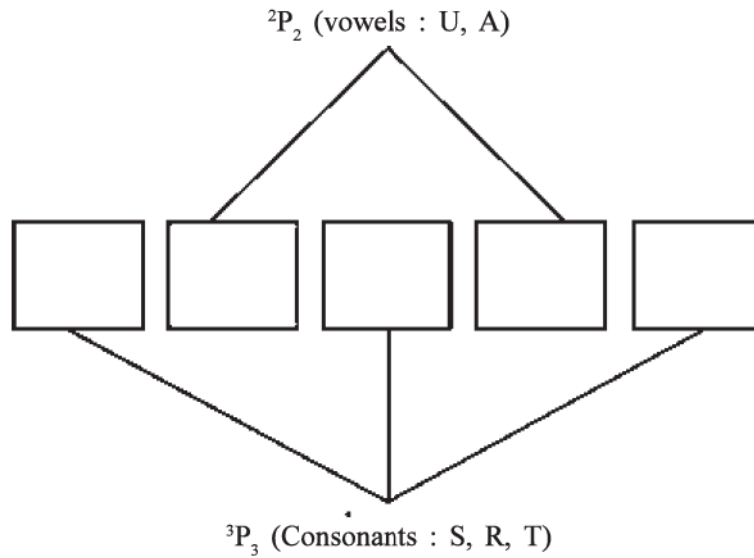
$$\begin{aligned}
 \therefore \text{Total Permutations} &= {}^4P_1 \times {}^4P_4 \\
 &= 4 \times 4! \\
 &= 4 \times 24 \\
 &= 96
 \end{aligned}$$

If zero is put in the first place, the number will be only 4 digit number.
For example,

$$02579 = 2579$$

$$09752 = 9752$$

Illustration 10 : In how many ways can all the letters of the word SURAT be arranged such that vowels are at even places only ?

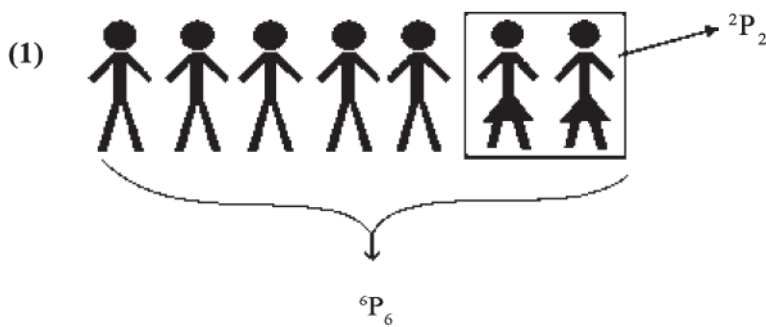


There are two vowels U and A in word SURAT. Now, they can be arranged in even places i.e the 2nd and the 4th places in 2P_2 ways. Remaining 3 letters (consonants) can be arranged in remaining 3 places (odd places) in 3P_3 ways.

$$\begin{aligned} \therefore \text{Total Permutations} &= {}^2P_2 \times {}^3P_3 \\ &= 2! \times 3! \\ &= 2 \times 6 \\ &= 12 \end{aligned}$$

Illustration 11 : In how many ways can 5 boys and 2 girls be arranged in a row such that,

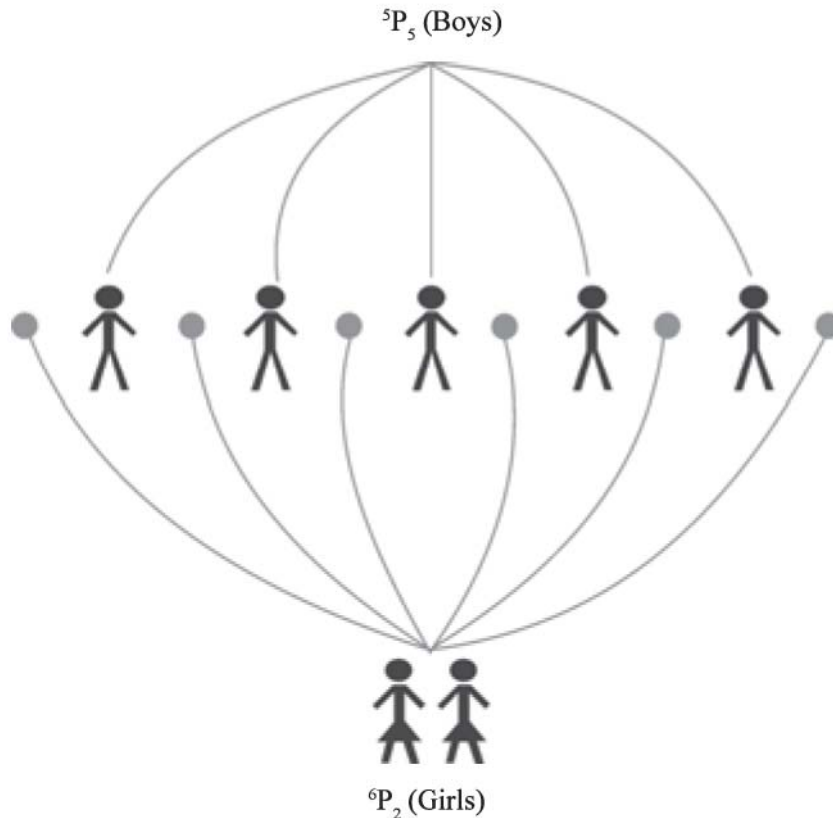
- (1) both the girls are together ?
- (2) both the girls are not together ?



Two girls are to be arranged together, so considering them as one person, total 6 persons can be arranged in 6P_6 ways and in each of these arrangements, two girls can be arranged among themselves in 2P_2 ways.

$$\begin{aligned} \therefore \text{Total Permutations} &= {}^6P_6 \times {}^2P_2 \\ &= 6! \times 2! \\ &= 720 \times 2 \\ &= 1440 \end{aligned}$$

(2)



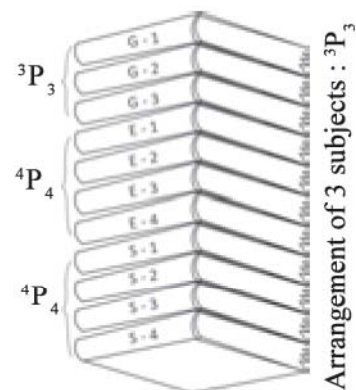
As the two girls are not to be arranged together, they can be arranged between 5 boys and on either sides. So, two girls can be placed in total 6 places in 6P_2 ways. Moreover 5 boys can be arranged in 5P_5 ways.

$$\begin{aligned}\therefore \text{Total Permutations} &= {}^6P_2 \times {}^5P_5 \\ &= 6 \times 5 \times 5! \\ &= 30 \times 120 \\ &= 3600\end{aligned}$$

Illustration 12 : There are 3 different books of Gujarati, 4 different books of English and 4 different books of Sanskrit on a table. In how many ways these books can be arranged such that the books of each subject are together ?

3 different books of Gujarati can be arranged in 3P_3 ways, 4 different books of English can be arranged in 4P_4 ways and 4 different books of Sanskrit can be arranged in 4P_4 ways. Moreover, arrangement of 3 subjects can be done in 3P_3 ways.

$$\begin{aligned}\therefore \text{Total Permutations} &= {}^3P_3 \times {}^4P_4 \times {}^4P_4 \times {}^3P_3 \\ &= 3! \times 4! \times 4! \times 3! \\ &= 6 \times 24 \times 24 \times 6 \\ &= 20,736\end{aligned}$$



Activity

Take your note books of Statistics, Accountancy and Economics. Also take your friend's notebooks of Statistics, Accountancy having his name on it. Now using your common sense, arrange all these notebooks one upon one such that notebooks of each subject remain together. Find out how many such different arrangements can be made? Now obtain solution of the same problem using fundamental principle of counting and compare both the answers.

Illustration 13 : In how many ways can 3 boys and 3 girls be arranged in a row such that boys and girls are alternately arranged ?

Boys (B) and girls (G) are to be arranged alternately, so they can be arranged as shown alongside :

B	G	B	G	B	G
		OR			
G	B	G	B	G	B

If we start arrangement with boys then that arrangement can be done in 3P_3 ways **OR** if we start arrangement with girls then that arrangement can be done in 3P_3 ways.

$$\begin{aligned}
 \therefore \text{Total Permutations} &= ({}^3P_3 \times {}^3P_3) + ({}^3P_3 \times {}^3P_3) \\
 &= (3! \times 3!) + (3! \times 3!) \\
 &= (6 \times 6) + (6 \times 6) \\
 &= 36 + 36 \\
 &= 72
 \end{aligned}$$

Illustration 14 : Arrangements are made using all the letters of the word YOUNG. If all these arrangements are arranged in the order of dictionary, what will be the rank of the word YOUNG ?

There are 5 letters Y, O, U, N, G in the word YOUNG which can be arranged in ${}^5P_5 = 5! = 120$ ways. Now, we have to obtain the order of the word YOUNG from all 120 arrangements as per dictionary order.

Alphabetical order of all letters of the word YOUNG will be G, N, O, U, Y

Arrangements with G in the first place will be ${}^1P_1 \times {}^4P_4 = 24$.

Arrangements with N in the first place will be ${}^1P_1 \times {}^4P_4 = 24$.

Arrangements with O in the first place will be ${}^1P_1 \times {}^4P_4 = 24$.

Arrangements with U in the first place will be ${}^1P_1 \times {}^4P_4 = 24$.

Arrangements with Y in the first place and G in the second place will be ${}^1P_1 \times {}^1P_1 \times {}^3P_3 = 6$

Arrangements with Y in the first place and N in the second place will be ${}^1P_1 \times {}^1P_1 \times {}^3P_3 = 6$

Arrangements with Y in the first place, O in the second place and G in the third place will be ${}^1P_1 \times {}^1P_1 \times {}^1P_1 \times {}^2P_2 = 2$

Arrangements with Y in first place, O in second place and N in the third place will be ${}^1P_1 \times {}^1P_1 \times {}^1P_1 \times {}^2P_2 = 2$.

Arrangements with Y in first place, O in the second place, U in the third place and G in the fourth place will be ${}^1P_1 \times {}^1P_1 \times {}^1P_1 \times {}^1P_1 \times {}^1P_1 = 1$.

Thereafter, the word YOUNG comes which itself takes 1 position.

$$\therefore \text{Dictionary order of the word YOUNG} = 24 + 24 + 24 + 24 + 6 + 6 + 2 + 2 + 1 + 1 = 114$$

To understand the above illustration, study the information in the table below :

If all the letters of the word TAB are used, ${}^3P_3 = 3! = 6$ arrangements can be made. These arrangement are TAB, TBA, ATB, ABT, BTA, BAT. These 6 arrangements arranged according to dictionary order are ABT, ATB, BAT, BTA, TAB, TBA. Thus, the order of the word TAB is 5.

Activity

Using common sense obtain all arrangements using all the letters of the word ZERO. Arrange all these arrangements according to dictionary order and find the rank of the word ZERO. Now obtain solution of this problem as per calculations of above illustration and compare both the answers.

Permutation of identical things

When out of total N things, A things are identical and remaining things are different then total arrangements will be $\frac{N!}{A!}$. There are three letters ($N = 3$) in the word BEE of which E is repeated two times ($A = 2$). So, total arrangement of the word BEE will be $\frac{N!}{A!} = \frac{3!}{2!} = \frac{6}{2} = 3$. We can understand that only three arrangements BEE, EBE, EEB are possible. In broader sense, out of N total things, A things are identical of the first type, B things are identical of the second type, C things are identical of third type and remaining things are different then total number of permutations of N things is $\frac{N!}{A! B! C!}$

Illustration 15 : How many arrangements can be made using all the letters of the word CINCINNATI ?

There are total 10 letters in the word CINCINNATI of which C is repeated 2 times, I is repeated 3 times and N is repeated 3 times.

$$\begin{aligned}\therefore \text{Total arrangement} &= \frac{10!}{2! 3! 3!} \\ &= \frac{3628800}{2 \times 6 \times 6} \\ &= 50,400\end{aligned}$$

EXERCISE 6.1

- Obtain the values of the following :
(1) ${}^{10}P_3$ (2) ${}^{50}P_2$ (3) 8P_7 (4) 9P_9
- If ${}^nP_3 = 990$ then find the value of n .
- If ${}^rP_r = 3024$, find the value of r .
- If $3 \cdot ({}^{n+3}P_4) = 5 \cdot ({}^{n+2}P_4)$ then find the value of n .
- In how many ways can 4 persons be arranged in a row ?
- How many six digit numbers can be formed using all the digits 1, 2, 3, 0, 7, 9 ?
- In how many ways can 5 boys and 3 girls be arranged in a row such that all the boys are together ?

8. There are 7 cages for 7 lions in a zoo. 3 cages out of 7 cages are so small that 3 out of 7 lions cannot fit in it. In how many ways can 7 lions be caged in 7 cages ?
9. Using all the digits 2, 3, 5, 8, 9, how many numbers greater than 50,000 can be formed ?
10. A person has 5 chocolates of different sizes. These chocolates are to be distributed among 5 children of different ages. If the biggest chocolate is to be given to the youngest child then in how many ways, 5 chocolates can be distributed among 5 children ?
11. How many total arrangements can be made using all letters of the following words ?
(1) STATISTICS (2) BOOKKEEPER (3) APPEARING
12. What is the ratio of number of arrangements of all letters of the word ASHOK and GEETA ?
13. There are 5 seats in a car including the driver's seat. If 3 out of 10 members in a family know driving then in how many ways, 5 persons out of 10 members can be arranged in the car ?
14. If all the arrangements formed using all the letters of the following words are arranged in the order of dictionary then what will be the rank of that word ?
(1) PINTU (2) NURI (3) NIRAL (4) SUMAN
15. How many arrangements of the letters of the word SHLOKA can be made such that all vowels are together ?
16. 7 speakers A, B, C, D, E, F are invited to deliver a speech in a program. Speakers have to deliver speech one after the other. In how many ways, speeches of 7 speakers can be arranged if the speaker B has to deliver his speech immediately after the speaker A ?

*

6.2 Meaning of Combination

We have studied permutation which gives number of arrangements of distinct things. Now, we shall think about the number of ways to select some of the things from given distinct things. For example, Tanya has three friends Rutva, Kathan and Kirti. Now if these three friends are to be arranged in two places then it can be done in ${}^3P_2 = 3 \times 2 = 6$ ways as under :

Types of Arrangement	1	2	3	4	5	6
First Place	Rutva	Rutva	Kathan	Kathan	Kirti	Kirti
Second Place	Kathan	Kirti	Rutva	Kirti	Rutva	Kathan

But now we have to think about types of selection. Tanya wants to invite only two friends out of the above three friends for a function at her home. How many options does Tanya have to invite two friends out of three ? Will Tanya invite Rutva and Kirti ? or Rutva and Kathan ? or Kathan and Kirti ? Thus, only these three different ways can arise which are shown in the following table :

Selection ways	1	2	3
Selected friends	Rutva and Kathan	Rutva and Kirti	Kathan and Kirti

It should be noted here that the order is important in permutation so ‘Rutva and Kathan’ and ‘Kathan and Rutva’ are different with reference to permutation, whereas the order is not important in selection. That is, whether ‘Rutva and Kathan’ or ‘Kathan and Rutva’ come to Tanya’s home, it means the same. Thus, selection of 2 friends out of 3 friends can be done in 3 ways. Each of these options of selection is called combination. Total number of such combinations is denoted by 3C_2 . Thus, it can be said that ${}^3C_2 = 3$. In general, the total number of combinations of selecting r ($\leq n$) things out of n different things will be nC_r . It is also denoted by ${}_nC_r$, $C(n, r)$, C_r^n . We will use the notation nC_r .

Thus, total combinations of selecting r things out of n different things is nC_r .

Now, let us see how the value of nC_r is obtained. Total combinations of selecting r things out of n distinct things is nC_r . In each type of selection, r things are involved. r things among themselves can be arranged in ${}^rP_r = r!$ ways. Thus, we get $r!$ permutations corresponding to each combination. Hence, nC_r combinations give ${}^nC_r \times r!$ permutations. But as we have seen earlier, total number of permutations of r things out of n distinct things is nP_r .

Hence, ${}^nP_r = {}^nC_r \times r!$

$$\therefore {}^nC_r = \frac{{}^nP_r}{r!}$$

Substituting the formula of nP_r ,

$$\begin{aligned} \therefore {}^nC_r &= \frac{n!}{(n-r)! r!} \\ &= \frac{n!}{r! (n-r)!} \end{aligned}$$

$$\therefore {}^nC_r = \frac{n!}{r!(n-r)!} \text{ where, } n > 0, r \geq 0, n \geq r, n \text{ is positive integer and } r \text{ is non-negative integer.}$$

Using this formula, following results can be derived :

Some important results				
${}^n C_0 = 1$	${}^n C_n = 1$	${}^n C_1 = n$	${}^n C_{n-1} = n$	${}^n C_r = {}^n C_{n-r}$
${}^n C_r = \frac{n!}{r!(n-r)!}$	${}^n C_r = \frac{n!}{r!(n-r)!}$	${}^n C_r = \frac{n!}{r!(n-r)!}$	${}^n C_r = \frac{n!}{r!(n-r)!}$	${}^n C_r = \frac{n!}{r!(n-r)!}$
$\therefore {}^n C_0 = \frac{n!}{0!(n-0)!}$	$\therefore {}^n C_n = \frac{n!}{n!(n-n)!}$	$\therefore {}^n C_1 = \frac{n!}{1!(n-1)!}$	$\therefore {}^n C_{n-1} = \frac{n!}{(n-1)!(n-(n-1))!}$	$\therefore {}^n C_{n-r} = \frac{n!}{(n-r)!(n-(n-r))!}$
$= \frac{n!}{n!}$	$= \frac{n!}{n!}$	$= \frac{n(n-1)!}{1 \times (n-1)!}$	$= \frac{n!}{(n-1)!1!}$	$= \frac{n!}{(n-r)! r!}$
$= 1$	$= 1$	$= n$	$= n$	$= {}^n C_r$
e.g. ${}^5 C_0 = 1$ ${}^{10} C_0 = 1$	e.g. ${}^5 C_5 = 1$ ${}^{10} C_{10} = 1$	e.g. ${}^5 C_1 = 5$ ${}^{10} C_1 = 10$	e.g. ${}^5 C_4 = 5$ ${}^{10} C_9 = 10$	e.g. ${}^5 C_4 = {}^5 C_1$ ${}^{10} C_7 = {}^{10} C_3$
If ${}^n C_x = {}^n C_y$ then, $x + y = n$ or $x = y$				

Illustration 16 : Find the values of the following :

- (1) ${}^8 C_3$ (2) ${}^{20} C_3$ (3) ${}^5 C_4$ (4) ${}^6 C_6$

(1) ${}^n C_r = \frac{n!}{r!(n-r)!}$

$$\begin{aligned} \therefore {}^8 C_3 &= \frac{8!}{3!(8-3)!} \\ &= \frac{8!}{3! \times 5!} \\ &= \frac{8 \times 7 \times 6 \times 5!}{3 \times 2 \times 1 \times 5!} \\ &= 56 \end{aligned}$$

Alternative method :

According to definition of ${}^n C_r$

$$\begin{aligned} {}^n C_r &= \frac{{}^n P_r}{r!} \\ \therefore {}^8 C_3 &= \frac{{}^8 P_3}{3!} \\ &= \frac{8 \times 7 \times 6}{3 \times 2 \times 1} \\ &= 56 \end{aligned}$$

- (2) According to the definition of ${}^n C_r$

$$\begin{aligned} {}^n C_r &= \frac{{}^n P_r}{r!} \\ \therefore {}^{20} C_3 &= \frac{{}^{20} P_3}{3!} \\ &= \frac{20 \times 19 \times 18}{3 \times 2 \times 1} \\ &= 1140 \end{aligned}$$

(3) According to the definition of ${}^n C_r$,

$${}^n C_r = \frac{{}^n P_r}{r!}$$

$$\therefore {}^5 C_4 = \frac{{}^5 P_4}{4!}$$

$$= \frac{5 \times 4 \times 3 \times 2}{4 \times 3 \times 2 \times 1}$$

$$= 5$$

(4) According to the definition of ${}^n C_r$,

$${}^n C_r = \frac{{}^n P_r}{r!}$$

$$\therefore {}^6 C_6 = \frac{{}^6 P_6}{6!}$$

$$= \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{6 \times 5 \times 4 \times 3 \times 2 \times 1}$$

$$= 1$$

Illustration 17 : If ${}^n C_2 = 45$, find the value of n .

$${}^n C_r = \frac{n!}{r!(n-r)!}$$

$$\therefore {}^n C_2 = \frac{n!}{2!(n-2)!}$$

$$\therefore 45 = \frac{n(n-1)(n-2)!}{2 \times 1 \times (n-2)!}$$

$$\therefore 90 = n(n-1)$$

$$\therefore n(n-1) = 10(10-1)$$

$$\therefore n = 10$$

Alternative method :

$${}^n C_2 = 45$$

$$\therefore \frac{n(n-1)}{2 \times 1} = 45$$

$$\therefore n(n-1) = 90$$

$$\therefore n(n-1) = 10(10-1)$$

$$\therefore n = 10$$

Illustration 18 : If $3 \cdot {}^{2n} C_3 = 44 \cdot {}^n C_2$ then find the value of n .

$$3 \cdot {}^{2n} C_3 = 44 \cdot {}^n C_2$$

$$\therefore \frac{3 \times 2n(2n-1)(2n-2)}{3 \times 2 \times 1} = \frac{44 \times n(n-1)}{2 \times 1}$$

$$\therefore \frac{2n(2n-1)2(n-1)}{2} = \frac{44n(n-1)}{2}$$

$$\therefore 4(2n-1) = 44$$

$$\therefore 2n-1 = 11$$

$$\therefore 2n = 12$$

$$\therefore n = 6$$

Illustration 19 : If ${}^n C_{n-3} = 56$, find the value of n .

$$\begin{aligned} {}^n C_{n-3} &= {}^n C_3 & [\because {}^n C_r &= {}^n C_{n-r}] \\ \therefore {}^n C_3 &= 56 \\ \therefore \frac{n(n-1)(n-2)}{3 \times 2 \times 1} &= 56 \\ \therefore n(n-1)(n-2) &= 336 \\ \therefore n(n-1)(n-2) &= 8(8-1)(8-2) \\ \therefore n &= 8 \end{aligned}$$

Illustration 20 : If ${}^n C_4 = {}^n C_6$ then find the value of n .

we know that if ${}^n C_x = {}^n C_y$ then,

Option 1 :

$$\begin{aligned} x + y &= n \\ \therefore 4 + 6 &= n \\ \therefore n &= 10 \end{aligned}$$

Thus, the value of n is 10.

Option 2 :

$$\begin{aligned} x &= y \\ \therefore 4 &= 6 \\ \text{Which is impossible.} \end{aligned}$$

Illustration 21 : If ${}^{50} C_{r+2} = {}^{50} C_{2r-3}$ then find the value of r .

We know that if ${}^n C_x = {}^n C_y$ then,

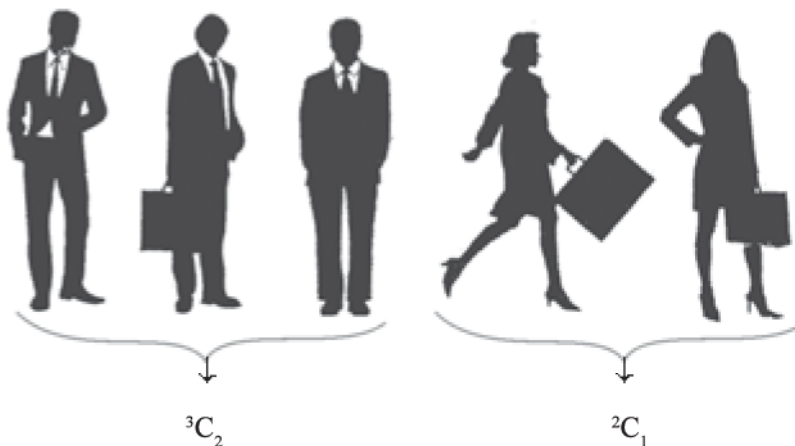
Option 1 :

$$\begin{aligned} x + y &= n \\ \therefore (r+2) + (2r-3) &= 50 \\ \therefore r+2+2r-3 &= 50 \\ \therefore 3r &= 51 \\ \therefore r &= 17 \end{aligned}$$

Option 2 :

$$\begin{aligned} x &= y \\ \therefore r+2 &= 2r-3 \\ \therefore r &= 5 \end{aligned}$$

Illustration 22 : In a company 2 male managers and 1 female manager are to be selected from 3 male managers and 2 female managers for training. How many ways this selection can be done ?



2 male managers out of 3 male managers can be selected in ${}^3 C_2$ ways and 1 female manager out of 2 female managers can be selected in ${}^2 C_1$ ways.

$$\begin{aligned} \therefore \text{Total Combinations} &= {}^3 C_2 \times {}^2 C_1 \\ &= 3 \times 2 \\ &= 6 \end{aligned}$$

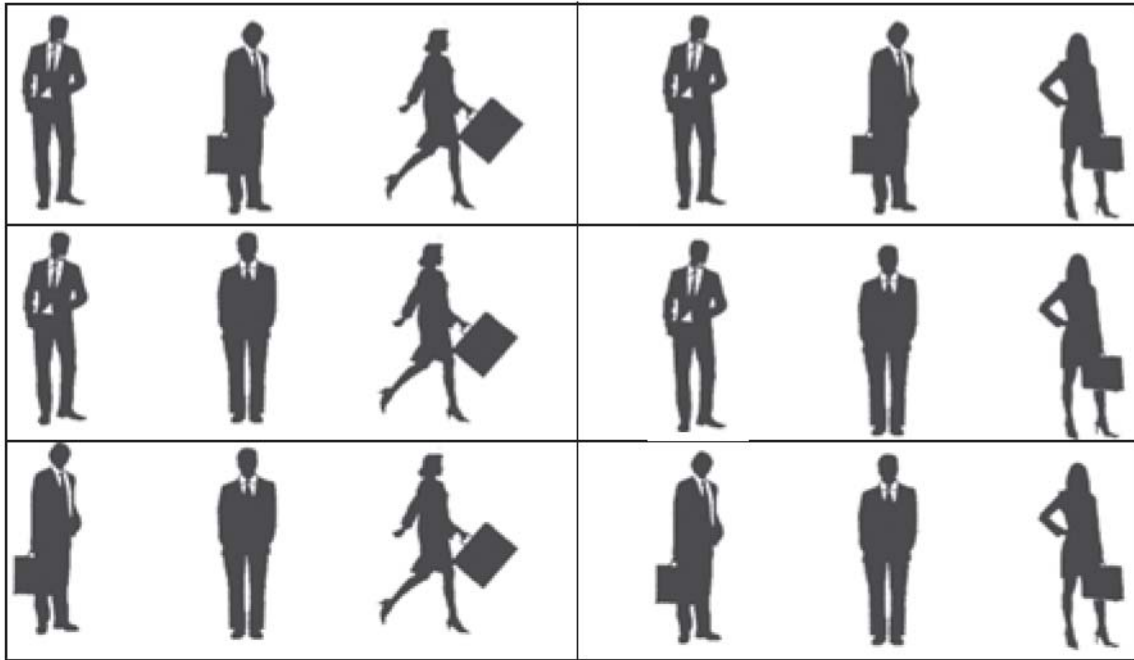


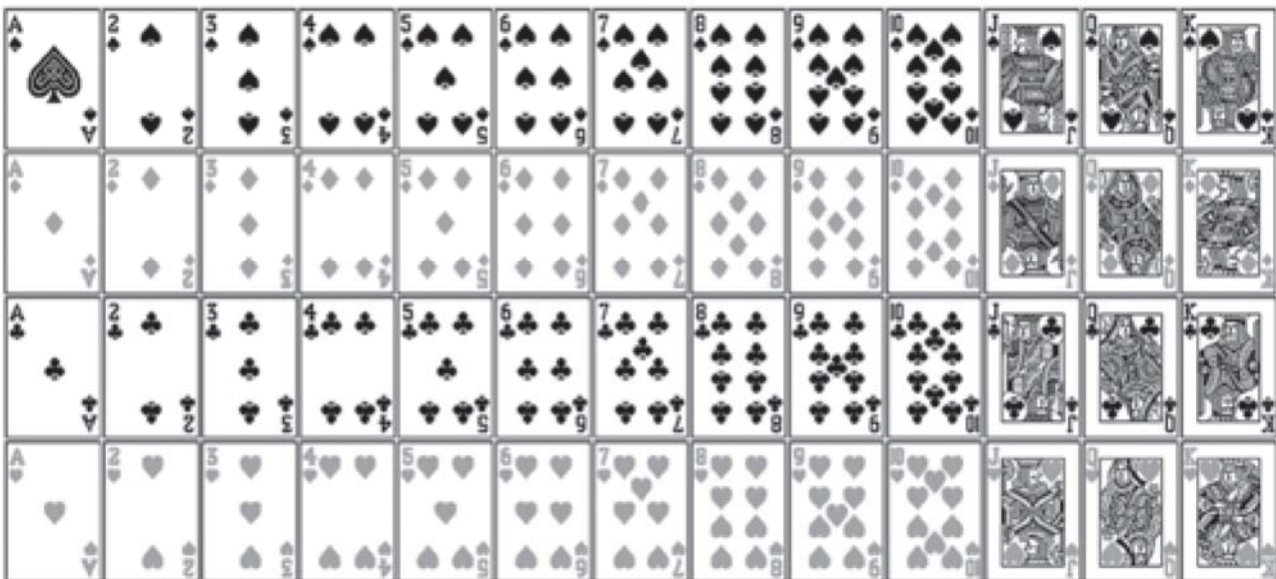
Illustration 23 : There are 5 yellow, 4 white and 3 pink flowers in a basket. In how many ways 3 yellow, 2 white and 1 pink flower can be selected from it ?

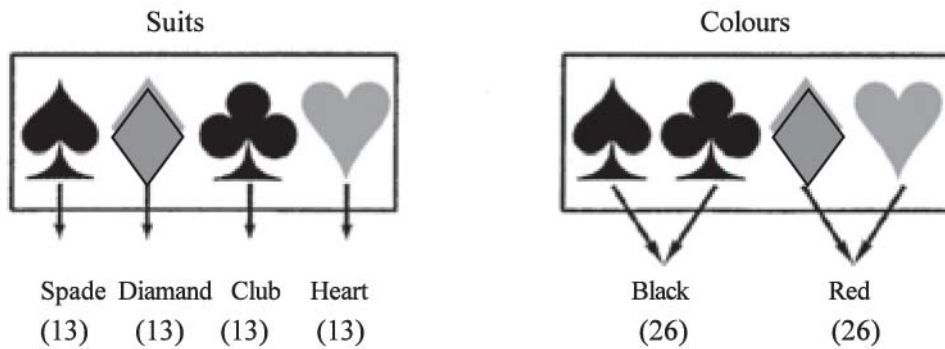
3 yellow flowers from 5 yellow flowers can be selected in 5C_3 ways, 2 white flower from 4 white flowers can be selected in 4C_2 ways and 1 pink flower from 3 pink flowers can be selected in 3C_1 ways.

$$\begin{aligned} \therefore \text{Total combinations} &= {}^5C_3 \times {}^4C_2 \times {}^3C_1 \\ &= 10 \times 6 \times 3 \\ &= 180 \end{aligned}$$

Illustration 24 : 2 cards are selected from a pack of 52 cards. In how many ways, this selection can be done such that

- (1) One is a face card and the other is a number card ?
- (2) both are of different colours ?
- (3) both are of same suit ?

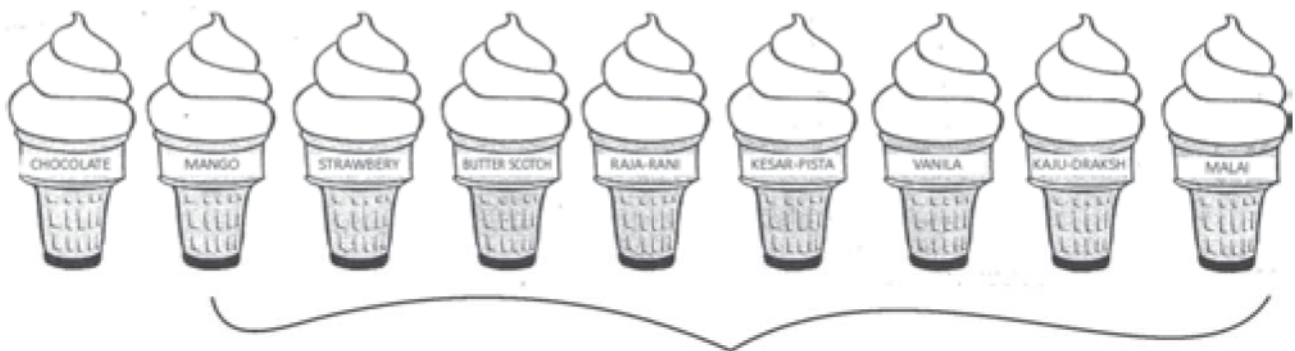




- (1) There are 12 face cards (king, Queen, Jack) and 40 number cards in a pack of 52 cards. One face card can be selected in $^{12}C_1$ ways and one number card can be selected in $^{40}C_1$ ways.
 \therefore **Total Combinations** = $^{12}C_1 \times ^{40}C_1$
 = 12×40
 = 480
- (2) There are 26 black and 26 red cards in a pack of 52 cards. One black card can be selected in $^{26}C_1$ ways and one red card can be selected in $^{26}C_1$ ways.
 \therefore **Total Combinations** = $^{26}C_1 \times ^{26}C_1$
 = 26×26
 = 676
- (3) There are 4 suits, spade, diamond, club and heart, in a pack of 52 cards and each suit has 13 cards. 2 cards are of same suit i.e. 2 of spade or 2 of diamond or 2 of club or 2 of heart.
 \therefore **Total Combinations** = $^{13}C_2 + ^{13}C_2 + ^{13}C_2 + ^{13}C_2$
 = $78 + 78 + 78 + 78$
 = 312

Illustration 25 : A boy named Kathan wants to select 5 different flavours of ice-cream cones out of 9 different flavours of ice-cream cones. If he wants one of the selected ice-cream cone to be of chocolate flavour then find total number of selections ?

Kathan wants to select 5 different flavours of ice-cream cones out of 9 different flavours of ice-cream cones. Chocolate ice-cream cone has to be selected which can be selected in 1C_1 ways. Now out of remaining 8 ice-cream cones, Kathan can select remaining 4 ice-cream cones in 8C_4 ways.



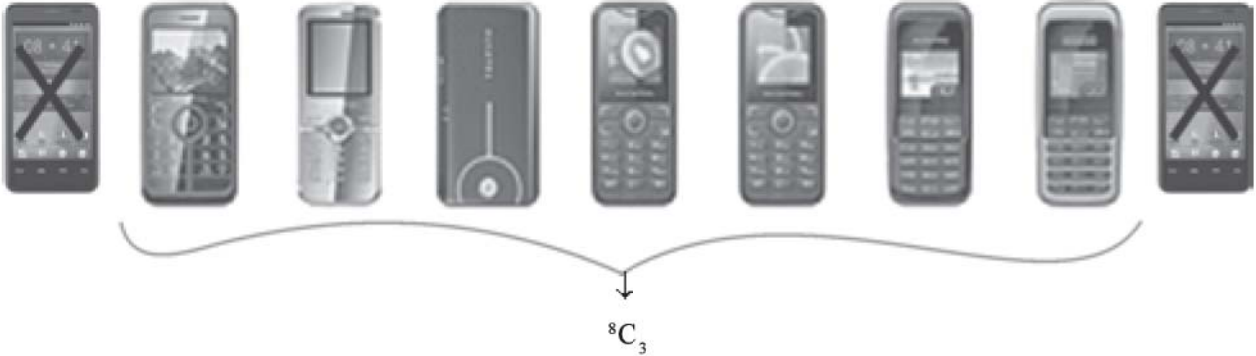
1C_1

8C_4

$$\begin{aligned} \therefore \text{Total Combinations} &= ^1C_1 \times ^8C_4 \\ &= 1 \times \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} \\ &= 70 \end{aligned}$$

Illustration 26 : A person wishes to purchase 3 mobile handsets from 10 different handsets of a company. But 2 mobile handsets do not fit into his budget. In how many ways can a person purchase 3 different handsets ?

Here there are 10 different handsets. But 2 mobile hand sets do not fit into his budget. So, selection of 3 mobile handsets out of remaining 8 mobile handsets can be done in 8C_3 ways.



$$\begin{aligned} \therefore \text{Total Combinations} &= {}^8C_3 \\ &= \frac{8 \times 7 \times 6}{3 \times 2 \times 1} \\ &= 56 \end{aligned}$$

Illustration 27 : There are 6 engineers and 4 managers in a company. In how many ways can a committee of 5 members be made such that

- (1) there are at least 2 managers ?
- (2) there are at the most 2 engineers ?
- (3) engineers are in majority ?

(1) In a committee of 5 members, selection of at least 2 managers can be done in following ways :

Manager (4)		Engineer (6)
2	and	3
	OR	
3	and	2
	OR	
4	and	1

$$\begin{aligned} \therefore \text{Total Combinations} &= ({}^4C_2 \times {}^6C_3) + ({}^4C_3 \times {}^6C_2) + ({}^4C_4 \times {}^6C_1) \\ &= (6 \times 20) + (4 \times 15) + (1 \times 6) \\ &= 120 + 60 + 6 \\ &= 186 \end{aligned}$$

(2) In a committee of 5 persons, selection of at the most 2 engineers can be done in following ways:

Engineer (6)		Manager (4)
2	and	3
	OR	
1	and	4

$$\begin{aligned}
 \therefore \text{Total Combinations} &= ({}^6C_2 \times {}^4C_3) + ({}^6C_1 \times {}^4C_4) \\
 &= (15 \times 4) + (6 \times 1) \\
 &= 60 + 6 \\
 &= 66
 \end{aligned}$$

(3) The selection of committee of 5 members such that engineers are in majority can be done in following ways :

Engineer (6)		Manager (4)
5	and	0
	OR	
4	and	1
	OR	
3	and	2

$$\begin{aligned}
 \therefore \text{Total Combinations} &= ({}^6C_5 \times {}^4C_0) + ({}^6C_4 \times {}^4C_1) + ({}^6C_3 \times {}^4C_2) \\
 &= (6 \times 1) + (15 \times 4) + (20 \times 6) \\
 &= 6 + 60 + 120 \\
 &= 186
 \end{aligned}$$

Illustration 28 : In an interview, 6 questions are asked to a person. In how many ways can a person give (1) at least 4 correct answers ? (2) at the most 3 correct answers ?

(1) If a person has given at least 4 correct answers of 6 questions asked then it means he has given 4 or 5 or 6 correct answers.

$$\begin{aligned}
 \therefore \text{Total Combinations} &= {}^6C_4 + {}^6C_5 + {}^6C_6 \\
 &= 15 + 6 + 1 \\
 &= 22
 \end{aligned}$$

(2) If a person has given at the most 3 correct answers of 6 questions asked then it means he has given 3 or 2 or 1 or 0 correct answers.

$$\begin{aligned}
 \therefore \text{Total Combinations} &= {}^6C_3 + {}^6C_2 + {}^6C_1 + {}^6C_0 \\
 &= 20 + 15 + 6 + 1 \\
 &= 42
 \end{aligned}$$

Activity

Make a group of 10 friends including yourself. Each one shakes hand with remaining in the group. Count total number of handshakes. Now obtain the solution of this problem with the formula of combination and compare both the answers.

EXERCISE 6.2

- Obtain the values of the following :
(1) ${}^{11}C_4$ (2) 9C_0 (3) ${}^{25}C_{23}$ (4) 8C_8
- Find the unknown value :
(1) ${}^nC_2 = 28$ (2) ${}^{27}C_{r+4} = {}^{27}C_{2r-1}$ (3) ${}^nC_{n-2} = 15$ (4) $4 \cdot {}^nC_4 = 7 \cdot {}^nC_3$
- 8 candidates applied for 2 posts of peon in a school. In how many ways can 2 peons be selected from 8 candidates ?
- 5 countries participate in a cricket tournament. In the first round, every country plays a match with the other country. How many matches will be played in this round ?
- There are 200 items in a box and 5% of them are defective. In how many ways can 3 items can be selected from the box so that all the items selected are defective ?
- In how many ways can 3 clerks and 1 peon be selected from 14 clerks and 6 peons working in a bank ?
- There are 3 white and 5 pink flowers in a box. In how many ways can
(1) three flowers of same colour be selected ?
(2) 2 flowers of different colours be selected ?
- Two cards are randomly selected from a pack of 52 cards. In how many ways can 2 cards be selected such that,
(1) both are of heart ?
(2) one is a king and the other is a queen ?
- There are 9 employees in a bank of which 6 are clerks, 2 are peons and 1 is a manager. In how many ways can a committee of 4 members be formed such that
(1) the manager must be selected ?
(2) two peons are not to be selected and the manager is to be selected ?
- In an office, there are 8 employees of which 3 are females and remaining are males. 3 employees are to be selected from the office for training. In how many ways can the selection be done so that at least one male is selected ?
- A person has 6 friends. In how many ways can he invite atleast one friend to his house ?
- In how many ways can 5 books be selected from 8 different books so that,
(1) a particular book is always selected ?
(2) a particluar book is never selected ?
- A student in 12th standard commerce stream has to appear for exam in 7 subjects. It is necessary to pass in all the subjects to pass an exam. Certain minimum marks must be obtained to pass in a subject. In how many ways can a student appearing for the exam fail ?
- In how many ways can a hotel owner subscribe 3 newspapers and 2 magazines from 8 different newspapers and 5 different magazines available in the city ? If a particular newspaper is to be selected and a particular magazine is not to be selected then in how many ways can this selection be done ?
- If ${}^nP_2 + {}^nC_2 = 84$ then find the value of n .
- If ${}^nP_r \div {}^nC_r = 24$ then find the value of r .

*

6.3 Meaning of Binomial Expansion

Any expression consisting of two terms connected by + or – sign is called a binomial expression. For example, $a + b$, $x - y$, $4a + 3b$, $x + 2a$ etc. are called binomial expressions. We have already studied the expansions of binomial expressions up to third power, which are as follows :

- $(x + a)^1 = x + a$
- $(x + a)^2 = x^2 + 2xa + a^2$
- $(x - a)^2 = x^2 - 2xa + a^2$
- $(x + a)^3 = x^3 + 3x^2a + 3xa^2 + a^3$
- $(x - a)^3 = x^3 - 3x^2a + 3xa^2 - a^3$

Now, let us think about an easy way to obtain the expansion of binomial expression with positive integer power greater than 3.

This expansion can be obtained using binomial theorem. We can write the coefficients of different terms of the above binomial expansions using combinations as follows :

- $(x + a)^1 = x + a$
 $= {}^1C_0 x + {}^1C_1 a$
- $(x + a)^2 = x^2 + 2xa + a^2$
 $= {}^2C_0 x^2 + {}^2C_1 xa + {}^2C_2 a^2$
- $(x + a)^3 = x^3 + 3x^2a + 3xa^2 + a^3$
 $= {}^3C_0 x^3 + {}^3C_1 x^2a + {}^3C_2 xa^2 + {}^3C_3 a^3$

Similarly, expansion of $(x + a)^4$ and $(x + a)^5$ can be written as follows :

$$(x + a)^4 = {}^4C_0 x^4 + {}^4C_1 x^3a + {}^4C_2 x^2a^2 + {}^4C_3 xa^3 + {}^4C_4 a^4$$

$$= x^4 + 4x^3a + 6x^2a^2 + 4xa^3 + a^4$$

$$(x + a)^5 = {}^5C_0 x^5 + {}^5C_1 x^4a + {}^5C_2 x^3a^2 + {}^5C_3 x^2a^3 + {}^5C_4 xa^4 + {}^5C_5 a^5$$

$$= x^5 + 5x^4a + 10x^3a^2 + 10x^2a^3 + 5xa^4 + a^5$$

If we observe the above expansion in detail, we come to know that the power of x decreases by one in each successive term and power of ‘ a ’ increases by one in each successive terms. From this, the expansion of binomial expression with positive integer power n can be written as follows :

$$(x + a)^n = {}^nC_0 x^n a^0 + {}^nC_1 x^{n-1} a^1 + {}^nC_2 x^{n-2} a^2 + \dots + {}^nC_n x^0 a^n$$

Expansion of $(x + a)^n$ is called binomial expansion. The general term of this expansion is ${}^nC_r x^{n-r} a^r$. If we put $r = 0, 1, 2, \dots, n$ in this term we get all the terms of binomial expansion.

r	Order in the expansion	${}^nC_r x^{n-r} a^r$
0	First term	${}^nC_0 x^{n-0} a^0 = x^n$
1	Second term	${}^nC_1 x^{n-1} a^1$
2	Third term	${}^nC_2 x^{n-2} a^2$
•	•	•
•	•	•
•	•	•
n	$(n + 1)$ th term	${}^nC_n x^{n-n} a^n = a^n$

Thus, ${}^n C_r x^{n-r} a^r$ is the $(r + 1)$ th term of the binomial expansion of $(x + a)^n$ which is called as general term of $(x + a)^n$.

i.e. $T_{r+1} = {}^n C_r x^{n-r} a^r$

We can see the following characteristics in the above expansion :

- (1) Total number of terms in the expansion of $(x + a)^n$ is $n + 1$ i.e. the number of terms is 1 more than the power of the binomial expression $(x + a)$.
- (2) The coefficients of the terms of expansion are ${}^n C_0, {}^n C_1, {}^n C_2, \dots, {}^n C_n$ respectively.
- (3) The first term of the expansion is x^n . In the successive terms of the expansion, the power of 'x' goes on decreasing by 1 and the power of 'a' goes on increasing by 1. The last term of the expansion is a^n .
- (4) The sum of powers of x and a is equal to n in any term of the expansion.
- (5) The coefficients of terms equidistant from the middle term or terms are equal.

The coefficients of binomial expansion are given below in a triangular form. This triangular arrangement was constructed by the French mathematician **Blaise Pascal**.

Pascal's triangle

power <i>n</i>	Coefficients										Sum of coefficients = 2^n
1											$2^1 = 2$
2											$2^2 = 4$
3											$2^3 = 8$
4											$2^4 = 16$
5											$2^5 = 32$
6											$2^6 = 64$
7											$2^7 = 128$
8											$2^8 = 156$
9											$2^9 = 512$

Illustration 29 : Expand $(x + y)^6$.

$$\begin{aligned}
 &(x + y)^6 \\
 &= {}^6 C_0 (x)^6 (y)^0 + {}^6 C_1 (x)^5 (y)^1 + {}^6 C_2 (x)^4 (y)^2 + {}^6 C_3 (x)^3 (y)^3 \\
 &\qquad\qquad\qquad + {}^6 C_4 (x)^2 (y)^4 + {}^6 C_5 (x)^1 (y)^5 + {}^6 C_6 (x)^0 (y)^6 \\
 &= x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6
 \end{aligned}$$

Illustration 30 : Expand $(1 + x)^4$.

$$\begin{aligned}
 &(1 + x)^4 \\
 &= {}^4 C_0 (1)^4 (x)^0 + {}^4 C_1 (1)^3 (x)^1 + {}^4 C_2 (1)^2 (x)^2 + {}^4 C_3 (1)^1 (x)^3 + {}^4 C_4 (1)^0 (x)^4 \\
 &= 1 + 4x + 6x^2 + 4x^3 + x^4
 \end{aligned}$$

Illustration 31 : Expand $(3a + 2y)^3$.

$$\begin{aligned}(3a + 2y)^3 &= {}^3C_0 (3a)^3 (2y)^0 + {}^3C_1 (3a)^2 (2y)^1 + {}^3C_2 (3a)^1 (2y)^2 + {}^3C_3 (3a)^0 (2y)^3 \\ &= 27a^3 + 3 (9a^2) (2y) + 3 (3a) (4y^2) + 8y^3 \\ &= 27a^3 + 54a^2y + 36ay^2 + 8y^3\end{aligned}$$

Illustration 32 : Expand $(3x - y)^4$.

$$\begin{aligned}(3x - y)^4 &= [3x + (-y)]^4 \\ &= {}^4C_0 (3x)^4 (-y)^0 + {}^4C_1 (3x)^3 (-y)^1 + {}^4C_2 (3x)^2 (-y)^2 + {}^4C_3 (3x)^1 (-y)^3 + {}^4C_4 (3x)^0 (-y)^4 \\ &= 81x^4 + 4 (27x^3) (-y) + 6 (9x^2) (y^2) + 4 (3x) (-y^3) + y^4 \\ &= 81x^4 - 108x^3y + 54x^2y^2 - 12xy^3 + y^4\end{aligned}$$

Note : If there is a negative sign between two terms of binomial expression, then its binomial expansion has the sign of the first term +, sign of the second term -, sign of the third term + and so on.

Illustration 33 : Expand $(2x - a)^5$.

$$\begin{aligned}(2x - a)^5 &= {}^5C_0 (2x)^5 (a)^0 - {}^5C_1 (2x)^4 (a)^1 + {}^5C_2 (2x)^3 (a)^2 - {}^5C_3 (2x)^2 (a)^3 \\ &\quad + {}^5C_4 (2x)^1 (a)^4 - {}^5C_5 (2x)^0 (a)^5 \\ &= 32x^5 - 5(16x^4) (a) + 10 (8x^3) (a^2) - 10 (4x^2) (a^3) + 5(2x) (a^4) - a^5 \\ &= 32x^5 - 80x^4a + 80x^3a^2 - 40x^2a^3 + 10xa^4 - a^5\end{aligned}$$

Illustration 34 : Expand $\left(x - \frac{2}{x}\right)^5$.

$$\begin{aligned}\left(x - \frac{2}{x}\right)^5 &= {}^5C_0 (x)^5 \left(\frac{2}{x}\right)^0 - {}^5C_1 (x)^4 \left(\frac{2}{x}\right)^1 + {}^5C_2 (x)^3 \left(\frac{2}{x}\right)^2 - {}^5C_3 (x)^2 \left(\frac{2}{x}\right)^3 + {}^5C_4 (x)^1 \left(\frac{2}{x}\right)^4 - {}^5C_5 (x)^0 \left(\frac{2}{x}\right)^5 \\ &= x^5 - 5 (x)^4 \left(\frac{2}{x}\right) + 10 (x)^3 \left(\frac{4}{x^2}\right) - 10 (x^2) \left(\frac{8}{x^3}\right) + 5 (x) \left(\frac{16}{x^4}\right) - \frac{32}{x^5} \\ &= x^5 - 10x^3 + 40x - \frac{80}{x} + \frac{80}{x^3} - \frac{32}{x^5}\end{aligned}$$

Activity

To obtain the value of $(11)^4$ using binomial expansion, first of all convert 11 to a binomial expression $(10 + 1)$. Now obtain the value of $(11)^4 = (10 + 1)^4$ using binomial expansion and check whether the value obtained is correct or not using calculator.

Illustration 35 : Expand $\left(\frac{a}{b} - \frac{b}{a}\right)^4$.

$$\begin{aligned} & \left(\frac{a}{b} - \frac{b}{a}\right)^4 \\ &= {}^4C_0 \left(\frac{a}{b}\right)^4 \left(\frac{b}{a}\right)^0 - {}^4C_1 \left(\frac{a}{b}\right)^3 \left(\frac{b}{a}\right)^1 + {}^4C_2 \left(\frac{a}{b}\right)^2 \left(\frac{b}{a}\right)^2 - {}^4C_3 \left(\frac{a}{b}\right)^1 \left(\frac{b}{a}\right)^3 + {}^4C_4 \left(\frac{a}{b}\right)^0 \left(\frac{b}{a}\right)^4 \\ &= \frac{a^4}{b^4} - 4 \left(\frac{a^3}{b^3}\right) \left(\frac{b}{a}\right) + 6 \left(\frac{a^2}{b^2}\right) \left(\frac{b^2}{a^2}\right) - 4 \left(\frac{a}{b}\right) \left(\frac{b^3}{a^3}\right) + \frac{b^4}{a^4} \\ &= \frac{a^4}{b^4} - \frac{4a^2}{b^2} + 6 - \frac{4b^2}{a^2} + \frac{b^4}{a^4} \end{aligned}$$

Illustration 36 : Expand $\left(\sqrt{a} + \frac{1}{\sqrt{a}}\right)^4$.

$$\begin{aligned} & \left(\sqrt{a} + \frac{1}{\sqrt{a}}\right)^4 \\ &= {}^4C_0 (\sqrt{a})^4 \left(\frac{1}{\sqrt{a}}\right)^0 + {}^4C_1 (\sqrt{a})^3 \left(\frac{1}{\sqrt{a}}\right)^1 + {}^4C_2 (\sqrt{a})^2 \left(\frac{1}{\sqrt{a}}\right)^2 \\ & \quad + {}^4C_3 (\sqrt{a})^1 \left(\frac{1}{\sqrt{a}}\right)^3 + {}^4C_4 (\sqrt{a})^0 \left(\frac{1}{\sqrt{a}}\right)^4 \\ &= a^2 + 4 (\sqrt{a})^2 + 6 + 4 \frac{1}{(\sqrt{a})^2} + \frac{1}{a^2} \\ &= a^2 + 4a + 6 + \frac{4}{a} + \frac{1}{a^2} \end{aligned}$$

Illustration 37 : Expand $(1 + x)^6$ and verify it by putting $x = 1$ on both the sides.

$$\begin{aligned} & (1 + x)^6 \\ &= {}^6C_0 (1)^6 (x)^0 + {}^6C_1 (1)^5 (x)^1 + {}^6C_2 (1)^4 (x)^2 + {}^6C_3 (1)^3 (x)^3 \\ & \quad + {}^6C_4 (1)^2 (x)^4 + {}^6C_5 (1)^1 (x)^5 + {}^6C_6 (1)^0 (x)^6 \\ &= 1 + 6x + 15x^2 + 20x^3 + 15x^4 + 6x^5 + x^6 \end{aligned}$$

Thus,

$$(1 + x)^6 = 1 + 6x + 15x^2 + 20x^3 + 15x^4 + 6x^5 + x^6$$

Putting $x = 1$,

$$\text{L.H.S.} = (1 + x)^6 = (1 + 1)^6 = (2)^6 = 64$$

$$\begin{aligned} \text{R.H.S.} &= 1 + 6x + 15x^2 + 20x^3 + 15x^4 + 6x^5 + x^6 \\ &= 1 + 6(1) + 15(1)^2 + 20(1)^3 + 15(1)^4 + 6(1)^5 + (1)^6 \\ &= 1 + 6 + 15 + 20 + 15 + 6 + 1 \\ &= 64 \end{aligned}$$

Thus, L.H.S. = R.H.S.

Illustration 38 : Obtain the value of $(\sqrt{3}+2)^5 - (\sqrt{3}-2)^5$ using binomial expansion method.

$$\begin{aligned}
 & (\sqrt{3}+2)^5 - (\sqrt{3}-2)^5 \\
 &= \left[\begin{array}{l} {}^5C_0(\sqrt{3})^5(2)^0 \\ + {}^5C_1(\sqrt{3})^4(2)^1 \\ + {}^5C_2(\sqrt{3})^3(2)^2 \\ + {}^5C_3(\sqrt{3})^2(2)^3 \\ + {}^5C_4(\sqrt{3})^1(2)^4 \\ + {}^5C_5(\sqrt{3})^0(2)^5 \end{array} \right] - \left[\begin{array}{l} {}^5C_0(\sqrt{3})^5(2)^0 \\ - {}^5C_1(\sqrt{3})^4(2)^1 \\ + {}^5C_2(\sqrt{3})^3(2)^2 \\ - {}^5C_3(\sqrt{3})^2(2)^3 \\ + {}^5C_4(\sqrt{3})^1(2)^4 \\ - {}^5C_5(\sqrt{3})^0(2)^5 \end{array} \right]
 \end{aligned}$$

(As there is a negative sign between two binomial expressions, the signs of terms of the second expression will change. As a result, the first, third and the fifth terms in the expansion of both expressions will get cancelled.)

$$\begin{aligned}
 &= 2 [{}^5C_1 (\sqrt{3})^4 (2)^1 + {}^5C_3 (\sqrt{3})^2 (2)^3 + {}^5C_5 (\sqrt{3})^0 (2)^5] \\
 &= 2 [5 (9) (2) + 10 (3) (8) + 32] \\
 &= 2 [90 + 240 + 32] \\
 &= 2 [362] \\
 &= 724
 \end{aligned}$$

Illustration 39 : Obtain the value of $(\sqrt{3}+\sqrt{2})^4 + (\sqrt{3}-\sqrt{2})^4$ using binomial expansion method.

$$\begin{aligned}
 & (\sqrt{3} + \sqrt{2})^4 + (\sqrt{3} - \sqrt{2})^4 \\
 &= \left[\begin{array}{l} {}^4C_0(\sqrt{3})^4(\sqrt{2})^0 \\ + {}^4C_1(\sqrt{3})^3(\sqrt{2})^1 \\ + {}^4C_2(\sqrt{3})^2(\sqrt{2})^2 \\ + {}^4C_3(\sqrt{3})^1(\sqrt{2})^3 \\ + {}^4C_4(\sqrt{3})^0(\sqrt{2})^4 \end{array} \right] + \left[\begin{array}{l} {}^4C_0(\sqrt{3})^4(\sqrt{2})^0 \\ - {}^4C_1(\sqrt{3})^3(\sqrt{2})^1 \\ + {}^4C_2(\sqrt{3})^2(\sqrt{2})^2 \\ - {}^4C_3(\sqrt{3})^1(\sqrt{2})^3 \\ + {}^4C_4(\sqrt{3})^0(\sqrt{2})^4 \end{array} \right]
 \end{aligned}$$

(As there is a positive sign between two binomial expressions, the signs of terms of the second expression will not change. As a result, the second, fourth terms in the expansion of both the expressions will be cancelled.)

$$\begin{aligned}
 &= 2 [{}^4C_0 (\sqrt{3})^4 (2)^0 + {}^4C_2 (\sqrt{3})^2 (\sqrt{2})^2 + {}^4C_4 (\sqrt{3})^0 (\sqrt{2})^4] \\
 &= 2 [9 + 6 (3) (2) + 4] \\
 &= 2 [9 + 36 + 4] \\
 &= 2 [49] \\
 &= 98
 \end{aligned}$$

EXERCISE 6.3

1. Obtain the expansion of following binomial expressions :

$$(1) (3a + 4b)^3 \quad (2) (1 + x)^7 \quad (3) \left(\frac{3}{x} - \frac{4x}{3}\right)^4 \quad (4) \left(\frac{\sqrt{x}}{3} + \frac{3}{\sqrt{x}}\right)^6 \quad (5) \left(\frac{a}{2} - \frac{b}{3}\right)^5$$

2. Obtain the values using binomial expansion :

$$(1) (\sqrt{5} + 1)^5 - (\sqrt{5} - 1)^5$$

$$(2) (\sqrt{2} + 1)^6 + (\sqrt{2} - 1)^6$$

$$(3) (\sqrt{5} + \sqrt{3})^4 + (\sqrt{5} - \sqrt{3})^4$$

3. Expand $(1 + x)^5$ and verify by putting $x = 1$ on both sides.

4. Expand $(1 + a)^6$ and verify by putting $a = 2$ on both sides.

Summary

- If there are m distinct things in Group 1 and n distinct things in Group 2 then selection of one thing from combined groups can be done in $m + n$ ways.
- If the first action can be done in m ways and second action can be done in n ways then two actions together can be done in $m \times n$ ways.
- $n! = n(n - 1)(n - 2) \times \dots \times 3 \times 2 \times 1$
- Total permutations of n different things in r places will be ${}^n P_r$.
- The fundamental difference between permutation and combination is that the order is important in permutation and the order is not important in combination. i.e. in permutation ab and ba are different while in combination ab and ba are same.
- The expansion of $(x + a)^n$ has $(n + 1)$ terms with coefficients ${}^n C_0, {}^n C_1, {}^n C_2, \dots, {}^n C_n$ respectively.

List of formulae

- ${}^n P_r = \frac{n!}{(n-r)!}$
- ${}^n P_0 = 1, {}^n P_n = n!, {}^n P_1 = n, {}^n P_{n-1} = n!$
- Out of N total things, A things are identical of the first type, B things are identical of the second type, C things are identical of the third type and remaining things are different then total number of permutations of N things is $\frac{N!}{A! B! C!}$.
- ${}^n C_r = \frac{n!}{r!(n-r)!}$
- The mathematical relation between ${}^n P_r$ and ${}^n C_r$ with reference to permutations and combinations :

$${}^n C_r = \frac{{}^n P_r}{r!}$$
- Expansion of expression $(x + a)^n$:

$${}^n C_0 (x)^n (a)^0 + {}^n C_1 (x)^{n-1} (a)^1 + {}^n C_2 (x)^{n-2} (a)^2 + \dots + {}^n C_n (x)^0 (a)^n$$

EXERCISE 6

Section A

For the following multiple choice questions select the correct option :

- In one group there are m distinct things and in the other group there are n distinct things. In how many ways can one thing be selected from both the groups ?
(a) mn (b) $\frac{m}{n}$ (c) $m - n$ (d) $m + n$
- If the first action can be done in m ways and second action can be done in n ways then in how many ways can both the actions be done together ?
(a) mn (b) $\frac{m}{n}$ (c) $m - n$ (d) $m + n$
- What is $n!$?
(a) Addition of 1 to n natural numbers
(b) Multiplication of 1 to n natural numbers
(c) Multiplication of 1 to $n - r$ natural numbers
(d) Multiplication of 0 to n natural numbers
- In usual notations, which of the following relation between permutation and combination is true ?
(a) ${}^n C_r = {}^n P_r \times r!$ (b) ${}^n P_r = {}^n C_r + r!$ (c) ${}^n P_r = \frac{{}^n C_r}{r!}$ (d) ${}^n C_r = \frac{{}^n P_r}{r!}$
- Which of the following is equivalent to ${}^n C_r$?
(a) $\frac{n!}{(n-r)!}$ (b) ${}^n C_{n-r}$ (c) ${}^n C_{r-1}$ (d) $\frac{{}^n C_{r+1}}{r}$
- Find the value of ${}^n C_0 + {}^n C_n$.
(a) 0 (b) 1 (c) 2 (d) $2n$
- If $(n + 1)! = 120$ then find the value of n .
(a) 3 (b) 4 (c) 5 (d) 6
- How many terms are there in the expansion of $(x + a)^{n-1}$?
(a) n (b) $n - 2$ (c) $n + 1$ (d) $n + 2$
- If $10 \times n! = 240$ then find value of n .
(a) 6 (b) 3 (c) 5 (d) 4
- State the last term in the expansion of $(x + a)^n$.
(a) a^n (b) a^{n-1} (c) x^0 (d) x^{n-1}
- A ride in a fun fair has 8 seats. In how many ways can 3 persons be arranged in this ride ?
(a) ${}^8 C_3$ (b) ${}^3 P_8$ (c) ${}^3 C_8$ (d) ${}^8 P_3$

Section B

Answer the following questions in one sentence :

- What is the main difference between permutation and combination ?
- Write the fundamental principle of counting for addition.
- Write the fundamental principle of counting for multiplication.

4. Write the mathematical relationship between permutation and combination in usual notations.
5. Write the coefficients of the terms in the expansion of $(x + a)^n$ for $n = 6$.
6. Write the general term of the expansion of $(x + a)^n$.
7. There are 5 empty seats in the coach of a train. In how many ways will 3 persons be seated ?
8. If ${}^n C_2 = 15$ then find the value of n .
9. If ${}^n P_3 = 210$, find the value of n .
10. How many new arrangements can be made using all the letters of the word TUESDAY ?
11. How many arrangements can be made using all the letters of the word VIAAN ?
12. In how many ways can 5 different letters be placed in 5 covers ?
13. What is ${}^n P_r$?
14. Write the coefficients of $(n + 1)$ terms in the expansion of $(x + a)^n$.
15. If ${}^n C_x = {}^n C_y$ then write the two possible relationships between x and y .

Section C

Answer the following questions :

1. Write the characteristics of binomial expansion.
2. 10 schools participate in a science fair. In how many ways can the first, second and the third prizes be distributed among these schools ?
3. In how many ways can 4 boys and 3 girls be arranged in a row such that no two boys and no two girls are together ?
4. There are 6 different books of Statistics, 5 different books of Accounts and 4 different books of English on a table. In how many ways can these books be arranged in a row such that the books of same subject remain together ?
5. How many 5 digit numbers can be formed using all the digits 3, 8, 0, 7, 6 ?
6. In how many ways can all the letters of the word TANI be arranged so that vowels remain together ?
7. In how many ways can all the letters of the word MANGO be arranged so that vowels are not together ?
8. How many numbers can be formed using all the digits of the number 1234321 such that odd digits occupy odd places only ?
9. What will be the ratio of number of arrangements obtained using all the letters of the word ROLLS and DOLLS ?
10. There are 2 defective screws in a box having 6 screws. In how many ways can 2 non-defective screws be selected from the box ?
11. In how many ways can 2 cards of queen or king can be selected from a well shuffled pack of 52 cards ?
12. In how many ways can 3 cards of same colour be selected from a well shuffled pack of 52 cards ?
13. Expand $(2x + 3y)^3$.
14. Expand $\left(x - \frac{1}{x}\right)^3$.
15. Expand $(y + k)^5$.

Section D

Answer the following :

1. Using all the first five natural numbers
 - (1) how many numbers can be formed ?
 - (2) how many numbers greater than 30,000 can be formed ?
 - (3) how many numbers divisible by 5 can be formed ?
2. In how many ways can 4 boys and 4 girls be arranged in a row such that no two boys and no two girls appear together ?
3. In how many ways can 3 boys and 2 girls be arranged in a row such that
 - (1) both the girls remain together ?
 - (2) boys and girls are alternately arranged ?
 - (3) all the three boys remain together ?
4. If all the arrangements formed using all the letters of the word WAKEFUL are arranged in the order of dictionary then what will be the rank of the word WAKEFUL ?
5. 4 couples (husband-wife) attend a party. In how many ways can 2 persons be selected from these 8 persons such that
 - (1) two persons selected are husband and wife ?
 - (2) one is a male and the other is a female ?
 - (3) one is a male and the other is a female but they are not husband and wife ?
6. There are 4 different books of Statistics and 3 different books of Economics on a table. In how many ways can 2 books be selected such that
 - (1) both the books are of the same subject ?
 - (2) both the books are of different subject ?
 - (3) no book of Economics is selected ?
7. 3 dolls, 4 kitchen sets and 3 cars are displayed in a toy shop. In how many ways can 3 toys be selected such that,
 - (1) all are dolls ?
 - (2) all are different toys ?
 - (3) two are dolls and one is a kitchen set ?
8. A committee of 3 members is to be formed from 4 chartered accountants and 5 doctors associated with a social organization. In how many ways can the committee be formed such that,
 - (1) the chartered accountants are in majority ?
 - (2) the doctors are in majority ?
9. Obtain the value of $(\sqrt{7} + 1)^3 - (\sqrt{7} - 1)^3$ using binomial expansion method.
10. Obtain the value of $(\sqrt{3} + 1)^6 + (\sqrt{3} - 1)^6$ using binomial expansion method.



Blaise Pascal
(1623 - 1662)

French mathematician – Pascal’s inventions and discoveries have been instrumental to development in the fields of geometry, physics and computer science. His exploration of binomial coefficients influenced Sir Isaac Newton, leading him to uncover his “general binomial theorem for fractional and negative powers.”

In the 1970s, the Pascal (Pa) unit was named after Blaise Pascal in honour of his contributions. Pascal is also credited with building the foundation of probability theory.



“Sampling theory deals with inductive inference which is the process by which we draw a conclusion about some measure of population based on a sample value.”

– W. A. Spur and C. P. Bonini

7

Sampling Methods

Contents :

- 7.1 Population and Sample: Meaning**
- 7.2 Population inquiry and sample inquiry**
- 7.3 Need for Sampling**
- 7.4 Characteristics of an Ideal sample**
- 7.5 Points to be considered while determining the sample size**
- 7.6 Sampling Methods :**
 - 7.6.1 Simple Random Sampling
 - 7.6.1.1 Meaning
 - 7.6.1.2 Method of lottery
 - 7.6.1.3 Method of random number table
 - 7.6.1.4 Advantages and disadvantages
 - 7.6.2 Stratified Random Sampling
 - 7.6.2.1 Meaning
 - 7.6.2.2 Advantages and disadvantages
 - 7.6.3 Systematic Sampling
 - 7.6.3.1 Meaning
 - 7.6.3.2 Advantages and disadvantages

7.1 Population and Sample

Meaning :

A set or group of all units or items under study is called a **Population**. A set or group of units selected from a population on the basis of some definite criteria is called a **sample**.

Suppose we want to study the IQ level of students in XII standard of Gujarat Board, then all XII standard students of Gujarat Board is the **Population**. If we select 1000 students from all XII standard students of Gujarat board based on some criteria, then set of these 1000 students becomes our **Sample**. Selection of sample is often used in day to day life. Before purchasing vegetables or fruits, we select and inspect a few from the whole lot of vegetables and fruits that shopkeeper has. All the vegetables or fruits this shopkeeper has, is the population and a few selected vegetables or fruits is the sample.

A sample can be selected with or without replacement from the population. A sample in which each unit is selected from the population after replacing the unit selected earlier in the population is called a **sample with replacement**. A sample in which each unit is selected from the population without replacing the unit selected earlier in the population is called a **sample without replacement**. Similarly, if units are selected simultaneously from the population then it is also a sample without replacement.

7.2 Population Inquiry and Sample Inquiry

There are two methods of collecting information or data about the population under study. The inquiry in which information is collected from each and every unit of the population is called **population inquiry or census survey**.

The census survey conducted every ten years in India is an example of population inquiry. Suppose, the study is about marks of XI and XII standard students of a particular school and we collect marks of each and every XI and XII standard student of that school then it is a population inquiry. Inspection of all votes cast during a particular election is another example of population inquiry.

The inquiry in which information is collected from a few units selected from the population is called **sample inquiry or sample survey**. Suppose, the study is to analyze spending habit of students of a particular college. If we collect information of spending habit of few students (selected with definite criteria) of that college then it is an example of sample inquiry. Testing a drop of blood drawn from the body to determine the blood group of a person is also an example of sample survey.

The procedure of selecting a sample from a population is called **sampling**.

7.3 Need for sampling

We often use sampling in real life. The quality control department of a factory randomly selects a few units from the total production to check the quality of the units produced. The total production of that factory is considered as population and a few units selected for determining the quality becomes the sample. The procedure of selecting some milk for determining the fat from the milk brought by milkman to the dairy, checking the blood sample of few patients suffering from a particular disease to detect the cause of the disease, etc. are other examples of sampling.

Sampling is inevitable in following situations:

- The number of units in a population is very large.
- Units of the population are spread over wide geographical area.
- The units under inquiry are to be destroyed i.e. when inspection of units is destructive in nature.
e.g. Inquiry of life of electric bulb.
- Limited availability of resources like time, money and experts for conducting an inquiry.

Generally, population inquiry is undertaken according to the constitutional and legal provision and for administrative reasons as well. Even in situations where population inquiry is feasible, preference is given to sample inquiry because a population inquiry involves more time, money and man power. Moreover, a large number of errors creep into the population inquiry because the task of organizing population inquiry is extensive and complicated.

The main objective of sampling is to draw inferences about the characteristics of the population on the basis of a sample inquiry. The measures such as mean, standard deviation, etc. calculated from the numerical data obtained from sample units are called **sample** statistics. All these measures for population data are called **parameters**.

7.4 Characteristics of an Ideal sample

Sample inquiry is used to draw conclusions of the population on the basis of the sample selected from the population. Thus, a sample selected from a population plays a vital role in obtaining information of the population. Hence, it is absolutely essential that a sample is appropriately selected from the population. A sample possessing the following characteristics is called an **ideal sample**.

- (1) It should represent the population i.e. all the characteristics of the population should be included in the sample.
- (2) It should be randomly selected i.e. no units of the population should be favoured or prejudiced.
- (3) The sample units should be selected in the same period of time and under identical conditions i.e. there should not be any major fluctuations affecting the study during the period of sampling.
- (4) Selection of sample units should be independent i.e. the selection of a unit should not be dependent on selection of any other unit of the sample.
- (5) The size of the sample should be adequate and determined in an appropriate manner.

7.5 Points to be considered while determining the sample size

The sample size means the number of units in the sample selected from the population. The following points should be considered while determining the sample size :

- (1) The size of the population and scope of the study
- (2) The variability in the units of the population or heterogeneity of population

- (3) Availability of resources like time, money and technical expertise
- (4) Expected level of accuracy of sample results

Additional information for understanding

Considering the availability of resources, large sample should be selected if population is large, heterogeneity of population is more and expected level of accuracy is high.

7.6 Sampling Methods

Method of selecting sample from the population is called **sampling method**. There are various methods of sampling that may be used singly or along with the other. The choice will be determined by the purpose for which sampling is sought and the nature of the population.

There are various methods of sampling but we shall study following methods:

- (1) Simple random sampling
- (2) Stratified random sampling
- (3) Systematic sampling

7.6.1 Simple Random Sampling

7.6.1.1 Meaning

Under this method, the whole population is taken as a single composite unit for the purpose of sampling. A sampling in which units are selected independent of each other in such a way that each unit belonging to the population has an equal chance of being a part of the sample is called a **simple random sampling**. Simple random sampling gives reliable results if the population is fairly homogenous, i.e. population observations possess almost same characteristics.

If the units are selected one by one and each unit is selected from the population after replacing the unit selected earlier in the population then the sampling procedure is called **simple random sampling with replacement**. If the units are selected one by one and each unit is selected from the population without replacing the unit selected earlier in the population then the sampling procedure is called **simple random sampling without replacement**.

Generally, in simple random sampling, the following two methods are used to select a random sample:

- (a) Method of Lottery
- (b) Method of random number table

7.6.1.2 Method of Lottery

This is the simplest and the most popular method of drawing a random sample from a population. In this method, units of the population are assigned numbers 1, 2, 3,... to identify them. Small and identical slips of paper bearing each number are prepared. (Identical slips means similar in colour, size and shape so that there is no bias in selection.) The slips are folded and mixed together in a vessel or a box. A blindfold selection of slips is made one by one until a sample of required size is obtained. The units bearing the numbers on the slips selected constitute a random sample of required size. In modern times, a mechanical device is also used to draw a random sample.

7.6.1.3 Method of Random Number Table

The method of lottery for selecting a random sample becomes time consuming and cumbersome when the population is very large. In such a situation, drawing a random sample becomes easy using readymade tables of random numbers. Random number tables are scientifically generated. The random number tables in common use are listed below :

(i) L.H.C. Tippett's random number tables (ii) Fisher and Yates' random number tables (iii) Random number tables of Rand Corporation of America

When a random sample is to be drawn, a page of the booklet of random number tables is opened at random and any row or column of that page is randomly selected. The random numbers appearing in order in the selected row or column are taken as the units according to the size of the population.

Tippett's random number table is commonly used to draw a random sample. We shall use it for illustrations of drawing a random sample. A part of the table is reproduced below for the purpose of illustration :

	1	2	3	4
(1)	053	274	323	599
(2)	667	484	786	833
(3)	992	347	253	338
(4)	428	982	564	785
(5)	278	154	490	076
(6)	819	314	589	889
(7)	195	222	428	924
(8)	390	379	699	786
(9)	420	598	443	692
(10)	664	430	343	118
(11)	171	035	189	236
(12)	289	505	667	484
(13)	535	300	112	089
(14)	784	280	257	154
(15)	640	143	364	326

Illustration 1 : Draw a random sample of 10 students without replacement from a population of 70 students of X standard of a particular school for testing the IQ level.

We shall first assign numbers 1 to 70 to the students of X standard of the school.

The size of this population (N) = 70, a two - digit number. Hence, we consider only first two digits of random numbers. We can select any row or column of any page of random number table booklet but we shall use the part of the table given above for clarity of understanding.

Suppose, we begin with the random numbers from the third row and then the consecutive rows for selecting the random sample. The random numbers with first two digits of each number are listed below :

(3)	99	34	25	33
(4)	42	98	56	78
(5)	27	15	49	07
(6)	81	31	58	88
(7)	19	22	42	92
(8)	39	37	69	78
(9)	42	59	44	69
(10)	66	43	34	11
(11)	17	03	18	23
(12)	28	50	66	48
(13)	53	30	11	08
(14)	78	28	25	15
(15)	64	14	36	32

Since the size of population is 70, we shall ignore random numbers which are greater than 70. As the sample has to be selected without replacement, we shall also ignore the random numbers which are repeated more than once. Considering this, we are left with the following numbers :

(3)	—	34	25	33
(4)	42	—	56	—
(5)	27	15	49	07
(6)	—	31	58	—
(7)	19	22	—	—
(8)	39	37	69	—
(9)	—	59	44	—
(10)	66	43	—	11
(11)	17	03	18	23
(12)	28	50	—	48
(13)	53	30	—	08
(14)	—	—	—	—
(15)	64	14	36	32

As we require a random sample of size 10, we shall select the first 10 random numbers. The selected random numbers are: 34, 25, 33, 42, 56, 27, 15, 49, 7, 31.

Thus, students with above numbers are selected. So, we have 10 students randomly selected from the school for testing their IQ level.

NOTE : As the selection of random numbers from random number tables can be from any row or column, random numbers obtained can be different for different persons.

Illustration 2 : Draw a random sample of 7 Income tax files without replacement from a population of 5000 Income tax files for scrutiny.

We shall first assign numbers 1 to 5000 to 5000 Income tax files.

The size of this population is $N = 5000$, a four - digit number. Hence, we shall consider three digits of a column and first digit of the next column. We can select any row or column of any page of random number table booklet, but we shall use the part of the table given above for clarity of understanding.

Suppose, we begin with the random numbers from the first row and then the consecutive rows for selecting the random sample.

	1	2	3	4
(1)	053	274	323	599
(2)	667	484	786	833
(3)	992	347	253	338
(4)	428	982	564	785
(5)	278	154	490	076
(6)	819	314	589	889
(7)	195	222	428	924
(8)	390	379	699	786
(9)	420	598	443	692
(10)	664	430	343	118
(11)	171	035	189	236
(12)	289	505	667	484
(13)	535	300	112	089
(14)	784	280	257	154
(15)	640	143	364	326

Since the size of the population is 5000, we shall ignore a random numbers which are greater than 5000. As the sample has to be selected without replacement, we shall also ignore the random numbers which are repeated more than once. The population size (N) = 5000, which has a four digits but the random number given in the above table is a three digit number. So, as shown in the above table, we shall join the three digits of the first column with the first digit of the second column to make random numbers of 4 digits. As we require random sample of size 7, we shall select first 7 random numbers. The selected random numbers are:

0532, 4289, 2781, 1952, 3903, 4205, 1710.

Thus, Income tax files with above numbers are selected. So, we have 7 randomly selected Income tax files for scrutiny.

Illustration 3 : Draw a random sample of size 5 (i) with replacement (ii) without replacement from a population of 50 units using the following 15 two digit random numbers: 62, 25, 6, 60, 95, 55, 98, 11, 71, 25, 20, 45, 89, 27, 40

We shall first assign numbers 1 to 50 to the 50 units of the the population.

Since the size of the population is 50, we shall ignore random numbers which are greater than 50. So we are left with random numbers : 25, 6, 11, 25, 20, 45, 27 and 40.

(i) As the sample has to be selected **with replacement**, we shall not ignore the random numbers which are repeated more than once. As we require random sample of size 5, we shall select first 5 random numbers. The selected random numbers are: 25, 6, 11, 25, and 20.

(ii) As the sample has to be selected **without replacement**, we shall ignore the random numbers which are repeated more than once. As we require random sample of size 5, we shall select first 5 random numbers. The selected random numbers are: 25, 6, 11, 20, and 45. (as 25 cannot be selected again)

7.6.1.4 Advantages and Disadvantages of Simple Random Sampling

Advantages :

- (1) There is no scope of bias or prejudice in the selection of a sample since each unit of the population has equal chance of selection.
- (2) The chance that a random sample represents the population is high. As the size of the sample increases, it becomes increasingly representative of the population.
- (3) Reliable information about the characteristics of the population can be obtained with less cost and time.

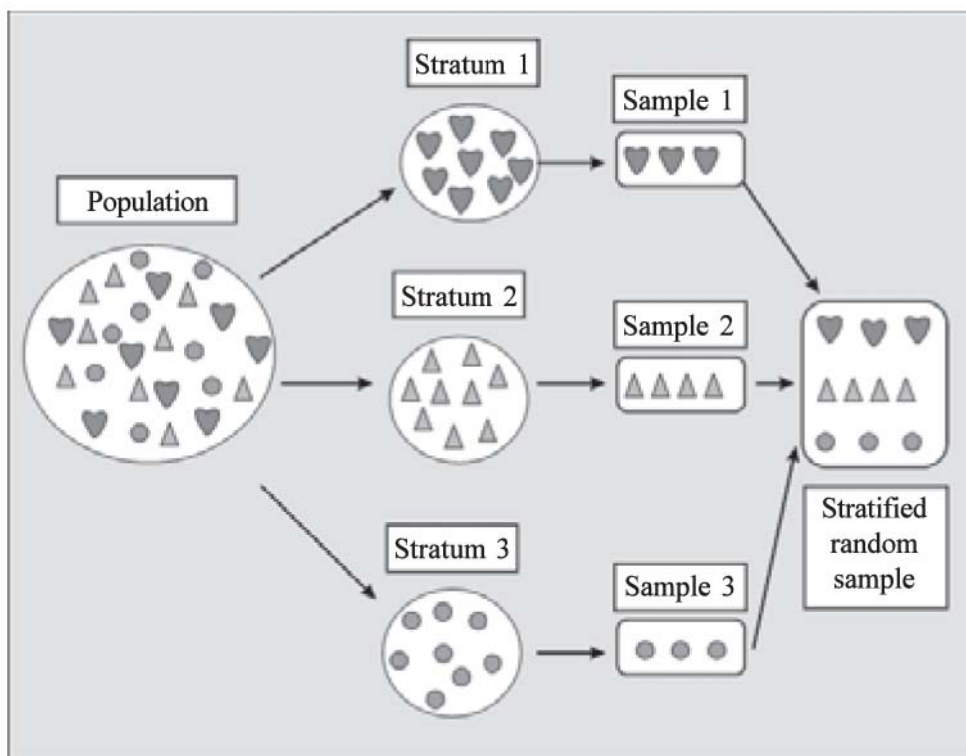
Disadvantages :

- (1) It requires complete list of units of the population. Absence of complete list restricts the use of this sampling design.
- (2) When the size of population is large, the preparation of slips or assigning numbers to the population units becomes tedious and time consuming.
- (3) Sample will not be a representative sample if the sample size is small and population is heterogeneous.

7.6.2 Stratified Random Sampling

7.6.2.1 Meaning

When the population is heterogeneous i.e. when there is considerable amount of variation among the units of population, the use of stratified random sampling is more appropriate than simple random sampling. In this form of sampling, the population is first divided into two or more mutually exclusive groups based on some characteristics of variables under study. A process of dividing heterogeneous population into non-overlapping fairly homogeneous groups is called stratification and these groups are called strata. These strata differ from one another and the units of each stratum have almost same characteristics in terms of variation. A random sample is independently drawn from each stratum and random samples so obtained from all strata are combined to get a sample which is called a **stratified random sample** and the method of selecting such a sample is called **stratified random sampling method**. Strategies like proportional allocation, optimal allocation, cost minimization etc. can be used for deciding number of units taken from each stratum.



In order to study the spending habits of commerce college students, all students of the college are divided into three income groups (strata) as high, middle and low income group of students. Certain fixed number of students is randomly selected from each of the three groups (strata). When these three samples are combined into a single sample, we get a stratified random sample.

Similarly, when we want to check the quality of the units produced in a factory, total production of the factory is divided into two groups (strata) as units produced in morning shift and units produced in night shift. Certain fixed number of units is randomly selected from each of two groups (strata). When these two samples are combined into a single sample, we get stratified random sample.

Illustration 4 : Draw a random sample of 1 percent students without replacement from 1500 students of a particular college for giving their opinion on college facilities. There are 600 students in First year (FY), 500 students in Second year (SY) and 400 students of Third year (TY) class in the college. Use the following 40 three – digit random numbers :

158, 092, 411, 745, 009, 724, 674, 550, 716, 359, 419, 969, 200, 458,
 384, 019, 676, 631, 390, 557, 299, 786, 706, 206, 729, 344, 543, 309,
 227, 483, 741, 766, 027, 070, 648, 956, 238, 912, 480, 558.

(Use the first 14 random numbers for FY, next 14 random numbers for SY and remaining random numbers for TY)

We shall first divide 1500 students of the college into three groups (strata) First Year, Second Year and Third Year students. Now we will select a random sample of 1% from each group. 1% of 600 of First year class i.e. 6, 1% of 500 of Second year class i.e. 5 and 1% of 400 of Third year class i.e. 4 will be selected using simple random sampling method.

Selecting sample of size 6 from 600 FY students :

Random numbers for FY : 158, 092, 411, 745, 009, 724, 674, 550, 716, 359, 419, 969, 200, 458.

Numbers 1 to 600 will be assigned to 600 students of FY. Since there are 600 students in FY, we shall ignore random numbers which are greater than 600. As the sample has to be selected without replacement, we shall also ignore the random numbers which are repeated more than once. We require a random sample of size 6, so we shall select first 6 random numbers. Thus, the selected random numbers are: 158, 092, 411, 009, 550, 359. Students with these numbers are selected from FY in our sample.

Selecting sample of size 5 from 500 SY students :

Random numbers for SY: 384, 019, 676, 631, 390, 557, 299, 786, 706, 206, 729, 344, 543, 309

Numbers 1 to 500 will be assigned to 500 students of SY. Since there are 500 students in SY, we shall ignore random numbers which are greater than 500. As the sample has to be selected without replacement, we shall also ignore the random numbers which are repeated more than once. We require a random sample of size 5, so we shall select first 5 random numbers. Thus the selected random numbers are: 384,019,390,299,206. Students with these numbers are selected from SY in our sample.

Selecting sample of size 4 from 400 TY students :

Random numbers for TY : 227, 483, 741, 766, 027, 070, 648, 956, 238, 912, 480, 558

Numbers 1 to 400 will be assigned to 400 students of TY. Since there are 400 students in TY, we shall ignore random numbers which are greater than 400. As the sample has to be selected without replacement, we shall also ignore the random numbers which are repeated more than once. We require a random sample of size 4, so we shall select first 4 random numbers. Thus the selected random numbers are: 227, 027, 070, 238. Students with these numbers are selected from TY in our sample.

Stratified random sample of size 15 is the combination of 6 students selected from FY, 5 students selected from SY and 4 students selected from TY.

Illustration 5 : For studying the usage of mobile phones, randomly select 7 persons without replacement from 30 boys and 20 girls from a locality of a city. There should be 3 boys and 4 girls in the randomly selected 7 persons.

Random numbers for boys : 82, 95, 18, 96, 20, 84, 56, 11, 52, 03

Random numbers for girls : 04, 40, 34, 13, 72, 11, 50, 55, 08, 11, 76, 18

We shall first divide 50 persons into two groups (strata) Boys and Girls. Now we will select a random sample of 3 boys from 30 boys and 4 girls from 20 girls using simple random sampling method.

Selecting sample of size 3 from 30 boys :

Numbers 1 to 30 will be assigned to 30 boys. Since there are 30 boys, we shall ignore random numbers which are greater than 30. As the sample has to be selected without replacement, we shall also ignore the random numbers which are repeated more than once. We require a random sample of size 3, so we shall select first 3 random numbers. Thus, the selected random numbers are: 18, 20 and 11. Boys with these numbers are selected from all boys of the locality.

Selecting sample of size 4 from 20 girls :

Numbers 1 to 20 will be assigned to 20 girls. Since there are 20 girls, we shall ignore random numbers which are greater than 20. As the sample has to be selected without replacement, we shall also ignore the random numbers which are repeated more than once. We require a random sample of size 4, so we shall select first 4 random numbers. Thus, the selected random numbers are: 04, 13, 11 and 8. Girls with these numbers are selected from all girls of the locality.

Stratified random sample of size 7 is the combination of these selected 3 boys and 4 girls.

7.6.2.2 Advantages and Disadvantages of Stratified Random Sampling

Advantages:

- (1) In this method, sample is selected from each stratum so stratified sample provides representative sample.
- (2) Generally, stratified sampling provides estimates with increased accuracy.
- (3) This method enables to maintain predetermined level of accuracy for each of the stratum.
- (4) Administrative convenience increases in this method.
- (5) Separate enumerators can be appointed to select samples from different strata.

Disadvantages :

- (1) At times, it is difficult task to divide population into homogeneous strata.
- (2) If stratification is not proper, accuracy of the results obtained by sample inquiry decreases.
- (3) The procedure of estimating population parameters in this method is more complicated as compared to simple random sampling.

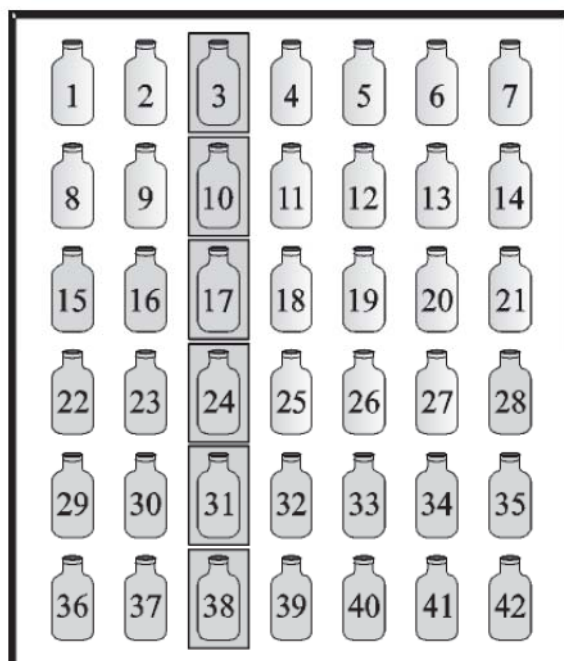
7.6.3 Systematic Sampling

7.6.3.1 Meaning

In this method, the first sample unit is randomly selected and the remaining units are automatically selected in definite sequence at uniform interval from one another. This method of sampling is usually recommended if the complete list of population units is available and the units are arranged in some systematic order such as alphabetical, chronological, geographical, etc. Systematic sample units are uniformly distributed over the population.

Suppose that the N units in the population are arranged in some systematic order and are numbered 1 to N . We want to select a sample of n units from it such that $N = n \times k$ or $k = N/n$ where k is usually called the **sampling interval**. Here we will assume k to be an integer value. We select a unit at random, from the first units and select every k -th unit thereafter. Set of such selected units is called **systematic sample** and the method of obtaining such a sample is called **systematic sampling**.

Suppose we want to select 6 plastic bottles from 42 plastic bottles manufactured in a factory. We will first assign the numbers 1 to 42 to bottles manufactured in the factory. Here population size (N) is 42, sample size (n) is 6, so $k = N/n = 42/6 = 7$. Select a bottle from 1 to 7 using simple random sample, say bottle number 3 is selected. There after, every 7th (k -th) bottle, that is the 10th, 17th, 24th, 31th and 38th bottle will be selected in the sample.



Suppose, for a study, we want a sample of 10 houses from a society of 120 houses. Here, $k = \frac{120}{10} = 12$.

So, every 12th house is chosen after a random starting point between 1 and 12. If house with number 8 is randomly selected, from the first 12 houses then the houses with following numbers will be selected:

8, 20, 32, 44, 56, 68, 80, 92, 104 and 116.

(If every selected house in the sample was a "corner house" then this corner pattern could destroy the randomness of the sample.)

Additional information for understanding

When $N \neq nk$, one of the following procedures may be adopted to obtain systematic sample.

- Drop one unit at random if sample has $(n + 1)$ units.
- Eliminate some units so that $N = nk$.
- Use circular systematic sampling scheme.
- Round off the fractional interval k .

Illustration 6 : There are 40 flats in a multi-storied apartment. 5 flats are to be selected using systematic sampling from 40 flats for decoration in a cultural program. Explain how the sample can be selected ?

Here, total number of flats is 40 i.e. $N = 40$ and we have to select 5 flats i.e. $n = 5$. So, the sampling interval $k = N/n = 40/5 = 8$. Now, every 8th flat will be selected after randomly selecting one flat from flat number 1 to 8. Suppose flat number 3 is selected from first 8 flats then flats with following numbers are selected:

3, 11, 19, 27, 35.

Illustration 7 : An owner of a service station has a list of 1000 car owners who have got their car repaired at his service station. He wishes to conduct satisfaction survey of 50 car owners.

Explain how a sample of 50 car owners can be selected using systematic sampling.

Here, total number of car owners is 1000 i.e. $N = 1000$ and we have to select 50 car owners i.e. $n = 50$. So, the sample interval $k = N/n = 1000/50 = 20$. Now, every 20th car owner will be selected after randomly selecting one car owner from 1 to 20. Suppose car owner number 11 is selected from first 20 car owners then the car owners with following numbers will be selected in systematic sample for conducting satisfaction survey 11, 31, 51, 71, ..., 991.

Illustration 8 : Select all possible systematic samples of size 3 from a population of size 12.

Here population size (N) is 12 and sample size (n) = 3. So the sample interval $k = N/n = 12/3 = 4$. Now, every 4th unit will be selected after randomly selecting the first unit from 1 to 4. Thus we will get 4 possible samples.

	1	2	3	4
	5	6	7	8
	9	10	11	12
SAMPLE NO.	I	II	III	IV

Thus, possible samples are:

Sample I: 1, 5, 9

Sample II : 2, 6, 10

Sample III : 3, 7, 11

Sample IV : 4, 8, 12

7.6.3.2 Advantages and Disadvantages of Systematic Sampling :

Advantages :

- (1) It is easier to draw a sample and often easier to execute without mistakes.
- (2) Sample is evenly spread over the entire population.
- (3) It requires less time and labour compared to simple random and stratified random sampling.

Disadvantages :

- (1) It works well only if complete list of population units is available.
- (2) Arranging all population units in a systematic order is time consuming, tedious and at times, not feasible.
- (3) Systematic sample is not completely random sample.
- (4) Sample may be biased if hidden periodicity in population coincides with that of selection of sample.

Summary

- A set or group of all units or items under study is called **population**.
- A set or group of units selected from population on basis of some definite criteria is called a **sample**.
- The inquiry in which information is collected from each and every unit of the population is called **population inquiry** or **census survey**.
- The inquiry in which information is collected from a few units selected from the population is called **sample inquiry** or **sample survey**.
- The procedure of selecting a sample from a population is called **sampling**.
- The measures such as mean, standard deviation, etc. calculated from the numerical data obtained from sample units are called **sample statistics**. All these measures of the population data are called **parameters**.
- Method of selecting sample from the population is called **sampling method**.
- The sampling in which units are selected independent of each other in such a way that each unit belonging to the population has an equal chance of being a part of the sample is called **simple random sampling**.
- When the population is heterogeneous, the use of stratified random sampling is more advantageous than simple random sampling.
- A random sample is independently drawn from each stratum and random samples so obtained from all strata are combined to get a sample which is called a **stratified random sample** and the method of selecting such a sample is called stratified random sampling method.
- In systematic sampling, the first sample unit is randomly selected and the remaining units are automatically selected in definite sequence at uniform interval from one another.

EXERCISE 7

Section A

For the following multiple choice questions choose the correct option :

1. A sample selected from a population consists which of the following ?
(a) All units of the population (b) Only 50% of the units of the population
(c) Only 15% of the units of the population (d) Some units of the population
2. Which of the following statements is true ?
(a) A sample in which a unit is selected after replacing the unit selected earlier in the population is called a sample without replacement.
(b) If a unit is to be destroyed during an inquiry then sample inquiry is not only necessary but also inevitable.
(c) In any sampling method, the sample size is larger than the population size.
(d) Stratified random sampling is best if the population is homogeneous.
3. Which of the following statements is true ?
(a) In stratified random sampling, all the units of the population have equal chance of being selected in the sample.
(b) In simple random sampling all units of the population have equal chance of being selected in the sample.
(c) In any sampling method, the sample size does not depend on population.
(d) In systematic sampling, all units of the population have equal chance of being selected in the sample.
4. A parameter and statistic respectively are characteristics of which of the following ?
(a) Population and sample (b) Sample and population
(c) Sample and sample (d) Population and population
5. Which sampling is affected the most if there is hidden periodicity in population ?
(a) Simple random sampling (b) Stratified random sampling
(c) Systematic sampling (d) Both (b) & (c)
6. Suppose we are using stratified sampling for a particular population and have divided it into strata of different sizes. How can we now make sample selection ?
(a) Select an equal number of units from each stratum at random.
(b) Draw unequal number of units from each stratum and weigh the results.
(c) Draw number of units from each stratum proportional to their size in the population.
(d) None of the above.
7. A security checkpoint that checks every vehicle entering into the mall is an example of which of the following ?
(a) Census inquiry (b) Stratified random sampling
(c) Systematic sampling (d) Simple random sampling

Section B

State whether the following statements are 'true' or 'false' :

1. A sampling plan that selects units from a population at uniform intervals in time, value or position is called stratified sampling.
2. A statistic is a characteristic of a population.
3. Units in the sample should be selected within the same time duration.
4. When properties of the units of the population have more dissimilarity, the use of stratified random sampling method is advantageous.
5. In simple random sampling method, each unit of the population has an equal chance of being included in the sample.
6. A sampling method that divides the population into homogeneous groups from which random samples are drawn is known as systematic sampling.
7. Each unit of the populations is examined in census inquiry.

Give answer in one sentence of the following questions :

1. What is the process by which inference about a population is made from sample information ?
2. Which sampling should be used when each group considered has small variation within itself but there is wide variation between different groups ?
3. In which sampling method, units are selected from the population at uniform intervals ?
4. Which table of random numbers is most commonly used ?
5. Which type of inquiry involves more errors ?
6. What is meant by population inquiry ?
7. What do you mean by a sample without replacement ?
8. When is the use of stratified random sampling considered to be favourable or suitable ?
9. When can the systematic random sample be biased ?
10. Define heterogeneous population.
11. Give an example an inquiry where units are destroyed during inspection.
12. If the three - digit random numbers are given and population size is of two digits, how will random numbers be used for selecting the sample ?
13. If the two-digit random numbers are given and population size is of three digits, how will random numbers be used for selecting the sample ?
14. Define parameters of the population.
15. Define sample statistics.

Section C

Answer the following questions :

1. When is sample inquiry undertaken ?
2. What is sampling ?
3. State the methods of selecting a simple random sample.
4. Name various methods of sampling.
5. State the strategies used for deciding the number of units to be selected from each stratum in stratified random sampling.
6. Explain sample interval in systematic random sampling.
7. Explain the process of stratification.
8. Define stratum in stratified random sampling.

Section D

Answer the following questions :

1. Explain the meaning of population inquiry and sample inquiry with an illustration.
2. Differentiate population inquiry and sample inquiry.
3. State the characteristics of an ideal sample.
4. State the points to be considered while determining the sample size.
5. State the advantages of simple random sampling.
6. Write a note on simple random sampling.
7. Write a note on stratified random sampling.
8. State the disadvantages of stratified random sampling.
9. Write a note on systematic sampling.
10. State the advantages of systematic random sampling.
11. Why is population inquiry usually not feasible in practice ?
12. State advantages of sampling.
13. Use the following random numbers to select a random sample of 5 ATMs without replacement from a total of 100 ATMs of a bank.
018, 502, 153, 096, 027, 007, 118, 245, 012, 054, 444, 211, 323, 428, 137.
14. There are 70 students in a class-room. A teacher wants to select 7 students for 7 activities. Obtain a random sample with replacement using the following random numbers.
274, 323, 923, 599, 667, 320, 910, 484, 786, 253, 009, 885, 115.
15. Three digit random numbers are given below :
170, 111, 352, 002, 563, 203, 405, 545, 111, 446, 776, 691, 816, 233, 616, 300, 250, 816, 010.
Using the random numbers, select a 2% random sample with and without replacement from a population of 350 units.

16. Draw a random sample of 2 percent students without replacement from 600 students of a particular college for giving their feedback on faculty members. There are 200 students in each of the three years (F.Y., S.Y. and T.Y.). Use the following three – digit random numbers:
 For F.Y. : 158, 092, 411, 745, 009, 724, 674, 550, 716, 359, 419, 696, 200, 458
 For S.Y. : 384, 019, 679, 131, 390, 057, 299, 786, 006, 206, 729, 344, 543, 309
 For T.Y. : 227, 483, 741, 766, 027, 070, 648, 956, 198, 912, 200, 058, 696, 500
17. To study the usages of fertilizer, randomly select 10 farmers without replacement from 30 small farm owners and 20 large farm owners. There should be 6 small farm owners and 4 large farm owners in the randomly selected 10 farmers.
 Random numbers for small farm owners :
 12, 95, 18, 96, 20, 84, 56, 11, 52, 03, 10, 45
 Random numbers for large farm owners :
 04, 40, 34, 11, 72, 11, 50, 55, 08, 13, 76, 18.
18. There are 60 employees in the office of an I.T. company. 5 employees are to be selected using systematic random sampling for a trial of ‘work from home’ concept. Explain how can a sample be selected ?
19. Select all possible samples of size 4 using systematic sampling from a population of 20 units.
20. A Teacher wants to check home-work of 10 students out of 30 students of Standard XI of a school. How many random samples can be obtained using systematic sampling ?



W. G. Cochran
 (1909 - 1980)

Prof. Cochran began his career at Rothamsted Research without a Ph.D., Cochran published 18 papers while at Rothamsted and attended the lectures of R.A. Fisher before leaving England for the United States. He was tasked with developing the graduate program in Statistics within the Mathematics Department. During this time, Cochran also worked on the advisory panel to the U.S. Census.

Later, he moved to Johns Hopkins University’s Department of Biostatistics, where his work shifted from agricultural issues to medical applications of Statistics. While at Johns Hopkins, he wrote Sampling Techniques and Experimental Designs. From 1957 until his retirement in 1976, Cochran worked at Harvard. His last position was Professor Emeritus.



“In practice, most of the interdependent relationships among characteristics can not be termed as function.”

– Unknown



Function

Contents :

8.1 Definition

8.2 Domain, codomain, range

8.3 Notations of functions

8.4 Types of function

8.4.1 One - one function

8.4.2 Many - one function

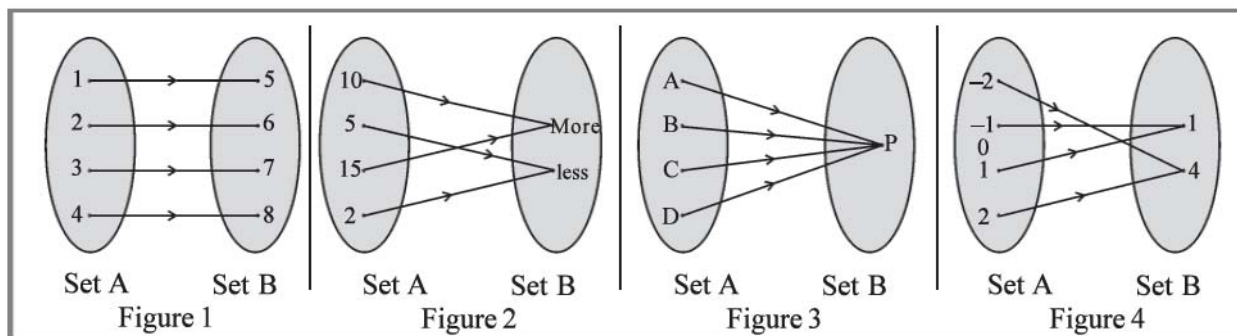
8.4.3 Constant function

8.5 Equal functions

8.6 Real function

8.1 Definition

While studying set theory we know that, there may exist a relationship between the elements of two different sets. Some such relations are known as functions. To understand this, let us consider the following figures :



In figure 1, the elements of set A are $\{1, 2, 3, 4\}$ and the elements of set B are $\{5, 6, 7, 8\}$. The rule between them is to add 4 in the elements of set A and hence we get the elements of set B.

In figure 2, elements of set A are $\{10, 5, 15, 2\}$ which shows the daily demand of a commodity and set B indicates its type $\{\text{more, less}\}$. Now, if it is decided to bifurcate the demand as “more” if it is 10 or more otherwise it is called “less” demand, then from figure 2 it is clear that there is a relation between the elements of set A and set B.

In figure 3, set A indicates the names of students of a class and set B indicates name of class teacher, then there exists an association between the elements of these sets.

In figure 4, the elements of set A are $\{-2, -1, 0, 1, 2\}$ and that of set B are $\{1, 4\}$. Some of the elements of set A are related with the elements of set B. But for the element “0” of set A, there is no corresponding element in set B.

From the above illustrations, it is clear that there exists various types of relations between the elements of two different sets. Can we say that all these relations are functions? To understand this, let us first understand the definition of a function.

Definition : If A and B are any two non-empty sets and each element of set A is related with one and only one element of set B by some rule, relation or correspondence then it is called a function from set A to set B and is denoted by f, g, h, k , etc.

In the above figure 1, the rule ‘add 4 to elements of set A’ is called function and is denoted by $f(x) = x + 4, x \in A$.

In the figure 2, the relation ‘bifurcate the demand as “more” if it is 10 or more otherwise demand is “less”’, is also called function.

In figure 3, the correspondence between name of a student and the name of the class teacher is also called function.

Observing the figure 4, we notice that set A is $\{-2, -1, 0, 1, 2\}$ and set B is $\{1, 4\}$. The elements $-2, -1, 1, 2$ have corresponding images $1, 4$ in set B but the element (0) of A does not have an image in B. Hence, it is not a function.

Illustration 1 : Verify whether relations between the elements of the sets given below are functions or not.

(1) $A = \{1, 2, 3, 4\}, B = \{3, 5, 7, 9\}$, and relation is $f(x) = 2x + 1, x \in A$

(2) $P = \{-\frac{1}{2}, 0, 1\}, S = \{10\}$, and rule is $k(x) = 10, x \in P$

(3) $A = \{2, 5, 6\}, B = \{1, \frac{3}{2}, \frac{9}{5}, \frac{11}{7}, \frac{13}{6}\}$, and rule is $y = \frac{2x-1}{x+1}, x \in A$

(4) $B = \{-1, 0, 1, 3\}, C = \{-5, -3, -1, 1, 3\}$, and rule is $h(x) = 2x + 3, x \in B$

- (1) **We shall determine the image for all $x \in A$.**

The relation is $f(x) = 2x + 1$

For $x = 1$, $f(1) = 2(1) + 1 = 3$

For $x = 2$, $f(2) = 2(2) + 1 = 5$

For $x = 3$, $f(3) = 2(3) + 1 = 7$

For $x = 4$, $f(4) = 2(4) + 1 = 9$

Thus, for each element of set A, there exists one and only one element in set B i.e. each element of set A is related with one and only one element of set B by the relation $f(x) = 2x + 1$. Hence, $f(x) = 2x + 1$ is a function.

- (2) **We shall determine the image for all $x \in P$ for $k(x) = 10$**

For $x = -\frac{1}{2}$, $k\left(-\frac{1}{2}\right) = 10$

For $x = 0$, $k(0) = 10$

For $x = 1$, $k(1) = 10$

So, for every $x \in P$, there exists an element "10" in set S.

Hence the relation $k(x) = 10$ is a function.

- (3) **We shall determine the image for all $x \in A$, for $y = \frac{2x-1}{x+1}$**

For $x = 2$, $y = \frac{2(2)-1}{2+1} = \frac{4-1}{3} = \frac{3}{3} = 1$

For $x = 5$, $y = \frac{2(5)-1}{5+1} = \frac{10-1}{6} = \frac{9}{6} = \frac{3}{2}$

For $x = 6$, $y = \frac{2(6)-1}{6+1} = \frac{12-1}{7} = \frac{11}{7}$

So for every $x \in A$, there exists a unique corresponding element in set B.

Hence, the relation $y = \frac{2x-1}{x+1}$ is a function.

- (4) **We shall determine the image for all $x \in B$. The relation is $h(x) = 2x + 3$**

For $x = -1$, $h(-1) = 2(-1) + 3 = -2 + 3 = 1$

For $x = 0$, $h(0) = 2(0) + 3 = 0 + 3 = 3$

For $x = 1$, $h(1) = 2(1) + 3 = 2 + 3 = 5$, which not an element of set C

So, there is no corresponding element for $x = 1$ in set C.

Hence, the relation $h(x) = 2x + 3$ is not a function.

8.2 Domain, Co-domain and Range

We know that, function is a unique relationship between the elements of two different non-empty sets. A unique image for elements of set A is obtained in set B. The set A is called domain and set B is called co-domain. A set of images or functional values of all the elements of set A is called range of the function. Domain of the function is denoted by D_f and range of the function is denoted by R_f . In notations, $R_f = f(A) = \{f(x) \mid x \in A\}$. Range of a function is a subset of co-domain or co-domain itself. For the illustration 1 above, in (1) set A is domain, set B is co-domain and it can be easily seen that the range is also co-domain B itself. Whereas in (2), domain is P and co-domain and range is S. In (3), domain is A, co-domain is B and range $R_f = \left\{1, \frac{3}{2}, \frac{11}{7}\right\}$ which is subset of co-domain B.

8.3 Notations of Functions

If f is a function of set A to set B then in notation it is shown as $f: A \rightarrow B$ and in words it is said that f is a function from set A to set B . Set A is called the domain and set B is called the co-domain.

From illustration 1, the relationship can be expressed as under :

- (1) $f: A \rightarrow B$, $A = \{1, 2, 3, 4\}$, $B = \{3, 5, 7, 9\}$ and $f(x) = 2x + 1$, $x \in A$
- (2) $k: P \rightarrow S$, $P = \{-\frac{1}{2}, 0, 1\}$, $S = \{10\}$ and $k(x) = 10$, $x \in P$
- (3) $f: A \rightarrow B$, $A = \{2, 5, 6\}$, $B = \{1, \frac{3}{2}, \frac{9}{5}, \frac{11}{7}, \frac{13}{6}\}$, $y = f(x) = \frac{2x-1}{x+1}$, $x \in A$
- (4) The relationship given here is not a function.

Illustration 2 : Obtain domain, co-domain and range for the following functions :

- (1) $f: A \rightarrow B$, $A = \{-1, 0, 1\}$, $B = \{1, 2, 3, 4, 5, 6, 7\}$, $f(x) = 2x + 5$, $x \in A$
- (2) $g: A \rightarrow N$, $A = \{-1, 2, 3, 4\}$, $g(x) = 3x + 5$, $x \in A$
- (3) $h: P \rightarrow S$, $P = \{-2, -1, 0, 1\}$, $S = \{-4, -3, -2, -1\}$, $h(x) = x - 2$, $x \in P$
- (4) $k: A \rightarrow Z$, $A = \{-\frac{1}{2}, 0, \frac{1}{2}\}$, $k(x) = 4x^2 + 3$, $x \in A$

- (1) Here, Domain $A = D_f = \{-1, 0, 1\}$

Co-domain $B = \{1, 2, 3, 4, 5, 6, 7\}$

Now, for every $x \in A$, we find $f(x) = 2x + 5$

For $x = -1$, $f(-1) = 2(-1) + 5 = -2 + 5 = 3$

For $x = 0$, $f(0) = 2(0) + 5 = 0 + 5 = 5$

For $x = 1$, $f(1) = 2(1) + 5 = 2 + 5 = 7$

\therefore Range of the function $R_f = \{f(-1), f(0), f(1)\} = \{3, 5, 7\}$

- (2) Here, Domain $A = D_f = \{-1, 2, 3, 4\}$

Co-domain $B = N$

Now, for every $x \in A$, we find $g(x) = 3x + 5$.

For $x = -1$, $g(-1) = 3(-1) + 5 = 2$

For $x = 2$, $g(2) = 3(2) + 5 = 11$

For $x = 3$, $g(3) = 3(3) + 5 = 14$

For $x = 4$, $g(4) = 3(4) + 5 = 17$

\therefore Range of the function $R_f = \{2, 11, 14, 17\}$ which is subset of co-domain.

- (3) Here, Domain $A = P = D_f = \{-2, -1, 0, 1\}$

Co-domain $B = S = \{-4, -3, -2, -1\}$

Now, for every $x \in P$, we find $h(x) = x - 2$.

For $x = -2$, $h(-2) = -2 - 2 = -4$

For $x = -1$, $h(-1) = -1 - 2 = -3$

For $x = 0$, $h(0) = 0 - 2 = -2$

For $x = 1$, $h(1) = 1 - 2 = -1$

\therefore Range of the function $R_f = \{-4, -3, -2, -1\}$

Here, the range and co-domain are same.

(4) Here, Domain $A = D_f = \left\{-\frac{1}{2}, 0, \frac{1}{2}\right\}$

Co-domain $B = Z$

Now, for every $x \in A$, we find $k(x) = 4x^2 + 3$.

For $x = -\frac{1}{2}$, $k\left(-\frac{1}{2}\right) = 4\left(-\frac{1}{2}\right)^2 + 3 = 1 + 3 = 4$

For $x = 0$, $k(0) = 4(0)^2 + 3 = 0 + 3 = 3$

For $x = \frac{1}{2}$, $k\left(\frac{1}{2}\right) = 4\left(\frac{1}{2}\right)^2 + 3 = 1 + 3 = 4$

\therefore Range of the function $R_f = \{3, 4\}$. Here, the range is a subset of co-domain.

Illustration 3 : Find domain, co-domain and range for the following functions :

(1) $f: A \rightarrow N, f(x) = x^2 + 1, A = \{x \mid -2 \leq x < 1, x \in Z\}$

(2) $f: Z \rightarrow N, f(x) = x^2 + 2, x \in Z$

(3) $f: N \rightarrow N, f(x) = 4x, x \in N$

(1) Here, Domain $D_f = A = \{-2, -1, 0\}$

Co-domain $B = N$

function $f(x) = x^2 + 1$

For $x = -2$, $f(-2) = 4 + 1 = 5$

For $x = -1$, $f(-1) = 1 + 1 = 2$

For $x = 0$, $f(0) = 0 + 1 = 1$

\therefore Range of the function $R_f = \{5, 2, 1\}$

(2) Domain $D_f = A = Z$

Co-domain $B = N$

For the range of the function $f(x) = x^2 + 2$,

$R_f = \{\dots, f(-2), f(-1), f(0), f(1), f(2), \dots\}$

$= \{\dots, 2, 3, 6, 11, \dots\}$

(3) Domain $D_f = A = N$

Co-domain $B = N$,

The range of function $f(x) = 4x$

$R_f = \{f(1), f(2), f(3), f(4), \dots\}$

$= \{4, 8, 12, 16, \dots\}$

Activity

For the last cricket match that you have seen, write a set indicating names of bowlers. Also write a set indicating the possible wickets that they may get. From this information can you say that the relation between the set of bowlers' names and the set of possible wickets is a function? If yes, then find the domain, co-domain and range of the function.

8.4 Types of Function

Functions are of many types. The main three types among them are as under :

- (1) One-One function (2) Many-One function (3) Constant function.

8.4.1 One – One function

Suppose $f : A \rightarrow B$. If for any two different elements of domain A, their images or functional values are different then the function f is called one – one function.

i.e. for function $f : A \rightarrow B$, $a_1 \neq a_2$, $a_1, a_2 \in A$ and $f(a_1) \neq f(a_2)$ then function f is called one-one function.

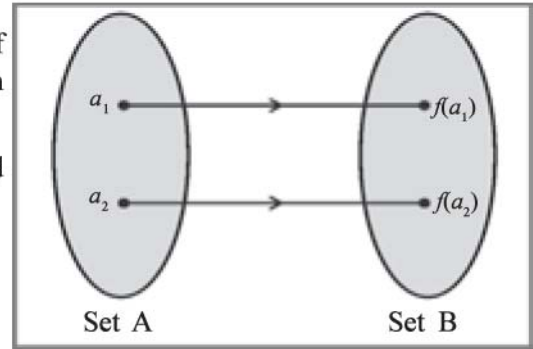
e.g., $g : X \rightarrow Y$ where $X = \{-2, -1, 0\}$,
 $Y = \{0, 1, 2, 3\}$ and $g(x) = x + 2$

Now, for $x = -2$, $g(-2) = 0$

for $x = -1$, $g(-1) = 1$

for $x = 0$, $g(0) = 2$

Thus, for $a_1 \neq a_2$, $a_1, a_2 \in A$, $g(a_1) \neq g(a_2)$ i.e. for two different values of the domain, there are two different images in co-domain. Hence, the given function g is one–one function.



8.4.2 Many–One function

Suppose $f : A \rightarrow B$. If for any two different elements of domain A, their images or functional values are same, then function f is called many–one function.

i.e. for function $f : A \rightarrow B$, $a_1 \neq a_2$, $a_1, a_2 \in A$ and $f(a_1) = f(a_2)$ then function f is called many–one function.

e.g., $f : A \rightarrow B$, $A = \{-2, -1, 1, 2\}$ and $B = \{2, 5\}$,
 $f(x) = x^2 + 1$. Then the functional value for each value of the domain is obtained as under :

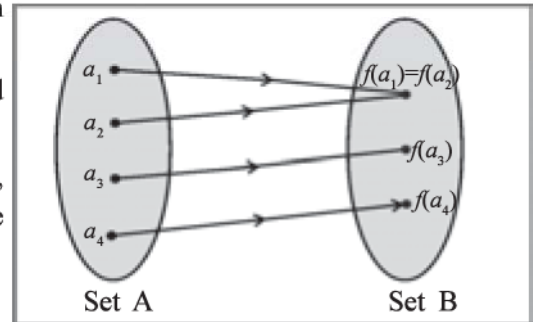
for $x = -2$, $f(-2) = (-2)^2 + 1 = 5$

for $x = -1$, $f(-1) = (-1)^2 + 1 = 2$

for $x = 1$, $f(1) = (1)^2 + 1 = 2$

for $x = 2$, $f(2) = (2)^2 + 1 = 5$

Thus, for $a_1 \neq a_2$, $a_1, a_2 \in A$, $f(a_1) = f(a_2)$ i.e. for two different values of the domain, there are two same images in co-domain. Hence, the given function f is many–one function.



8.4.3 Constant function

Suppose $f : A \rightarrow B$. If for each element of domain A, the image or functional value is same then function f is called constant function.

e.g., $f = \{1, 2, 3\} \rightarrow \{4, 5, 6\}$ and $f(x) = 5$. Then, for values 1, 2 and 3 of the variable x , we get $f(x) = 5$. Thus, f is a constant function.

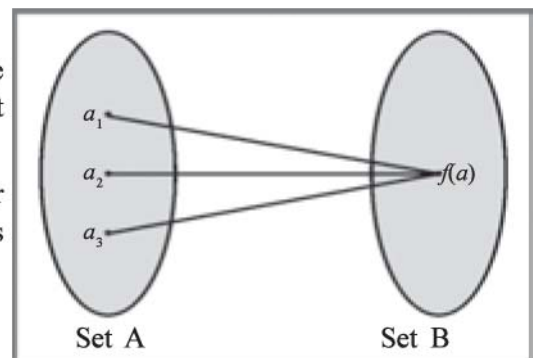


Illustration 4 : State the type of function $f : \mathbf{N} \rightarrow \mathbf{N}$, $f(x) = x^2$

Domain $D_f = \mathbf{N} = \{1, 2, 3, \dots\}$

Co-domain $B = \mathbf{N} = \{1, 2, 3, \dots\}$

$f(x) = x^2$. By substituting the values of $x = 1, 2, 3, \dots$ the images are 1, 4, 9,...

Thus, for two different values of domain their images in co-domain are different. Hence, the given function is one – one function.

Illustration 5 : $f : \mathbf{Z} \rightarrow \mathbf{N} \cup \{0\}$, $f(x) = x^2$ then state the type of function f .

Domain $D_f = \mathbf{Z} = \{\dots -3, -2, -1, 0, 1, 2, 3, \dots\}$

Co-domain $B = \mathbf{N} \cup \{0\} = \{0, 1, 2, 3, \dots\}$

By substituting the values of $x = -2, -1, 0, 1, 2$ the images are 4, 1, 0, 1, 4

Thus, for two different values $\{-2, 2\}$ and $\{-1, 1\}$ of domain their images in co-domain are same. Hence, the given function is many – one function.

Illustration 6 :

(i) $f : \mathbf{Z} \rightarrow \mathbf{R}$ and $f(x) = 100$ then state the type of function f .

Domain $A = D_f = \mathbf{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ and co-domain $B = \mathbf{R}$. For every value of the domain, the image in B is same, which is “100”.

Hence, the given function is constant function.

(ii) State the type of the function between dates of a month and days of the week.

In a month, there are four weeks. So, the four different days are repeated four times with different dates of a month.

Hence, it is a many–one function.

Activity

Decide the type of function formed between the elements of sets indicating names of your five friends and the number of their family members.

8.5 Equal Functions

If two different functions f and g satisfy the following conditions then they are said to be equal functions. It is denoted by $f = g$.

(1) Both the functions must have same domain i.e. both the functions must be defined on same domain.

(2) For each element x of the domain, $f(x) = g(x)$ i.e. for each value of the domain their images must be same.

In notation, the above definition can be written as under :

If $f : A \rightarrow B$ and $g : A \rightarrow C$ and for each, $x \in A$, $f(x) = g(x)$ then we can say that $f = g$.

Illustration 7 : If $f : \mathbf{N} \rightarrow \mathbf{N}$, $f(x) = x^2$ and $g : \mathbf{Z} - \{0\} \rightarrow \mathbf{N}$, $g(x) = x^2$, can we say that f and g are equal functions ?

Domain of function f is $D_f = A = \mathbf{N}$

Domain of function g is $D_g = A = \mathbf{Z} - \{0\}$

Thus, both the functions are defined on different domains. So, they are not equal functions.

Illustration 8 : $f : A \rightarrow B$, $A = \{1, 3\}$, $B = \{1, 4, 9, 16\}$, $f(x) = x^2$ and $g : A \rightarrow B$, $A = \{1, 3\}$,

$B = \{1, 4, 7, 9, 11\}$ $g(x) = 4x - 3$. Check the equality of the functions f and g .

Both the functions given here are defined on the same domain A .

$$\begin{aligned} \text{For } x = 1, f(1) &= 1^2, & g(1) &= 4(1) - 3 \\ &= 1 & &= 1 \\ x = 3, f(3) &= 3^2, & g(3) &= 4(3) - 3 \\ &= 9 & &= 9 \end{aligned}$$

Thus, for each $x \in A$, $f(x) = g(x)$. Therefore, f and g are equal functions.

Illustration 9 : $f : A \rightarrow B$, $f(x) = 1 - 4x$ and $g : A \rightarrow C$, $g(x) = 6x + 1$ and $A = \{0, 1, 2\}$. Are f and g equal functions ?

Both the functions given here are defined on the same domain.

$$\text{For } x = 0, f(0) = 1 - 4(0) = 1 \text{ and } g(0) = 6(0) + 1 = 1$$

$$\text{For } x = 1, f(1) = 1 - 4(1) = -3 \text{ and } g(1) = 6(1) + 1 = 7$$

Thus, $f(1) \neq g(1)$. Therefore, f and g are not equal functions.

8.6 Real Functions

If $f : A \rightarrow B$ where $A \subset \mathbb{R}$ then f is called function of real variable and if the range of f is also defined on real set \mathbb{R} then f is called real function. e.g., for $f : Z \rightarrow \mathbb{R}$, domain is Z which is the set of integers and the co-domain is the real set \mathbb{R} itself. Therefore, function f is called real function. In short, a function for which domain and co-domain are defined on real set \mathbb{R} or on any subset of it, is called real function.

Illustration 10 : If $f(x) = \frac{x^3 + 1}{x^2 - 2x + 1}$ where $x \in Z - \{1\}$ then find $f(-2)$, $f(-1)$ and $f(0)$.

$$f(x) = \frac{x^3 + 1}{x^2 - 2x + 1}$$

$$f(-2) = \frac{(-2)^3 + 1}{(-2)^2 - 2(-2) + 1}; \quad f(-1) = \frac{(-1)^3 + 1}{(-1)^2 - 2(-1) + 1}$$

$$= \frac{-8 + 1}{4 + 4 + 1} \qquad = \frac{-1 + 1}{1 + 2 + 1}$$

$$= -\frac{7}{9} \qquad = \frac{0}{4}$$

$$= 0$$

$$f(0) = \frac{0^3 + 1}{0^2 - 2(0) + 1} = \frac{1}{1} = 1$$

Illustration 11 : If $f(x) = x^3 + 3^x - x^2 - 2^x$ then find $f(3) - 6f(2)$.

$$f(x) = x^3 + 3^x - x^2 - 2^x$$

$$f(3) = (3)^3 + 3^3 - (3)^2 - (2)^3 \quad \text{and} \quad f(2) = 2^3 + 3^2 - 2^2 - 2^2$$

$$= 27 + 27 - 9 - 8 \qquad = 8 + 9 - 4 - 4$$

$$= 37 \qquad = 9$$

$$\begin{aligned}\therefore f(3) - 6f(2) &= 37 - 6(9) \\ &= 37 - 54 \\ &= -17\end{aligned}$$

Illustration 12 : $f(x) = x(2x - 7)$ where $x \in \mathbb{R}$. If $f(x) = 15$ then find the value of x .

$$\begin{aligned}f(x) &= 15 \\ \therefore x(2x - 7) &= 15 \\ \therefore 2x^2 - 7x - 15 &= 0 \\ \therefore 2x^2 + 3x - 10x - 15 &= 0 \\ \therefore x(2x + 3) - 5(2x + 3) &= 0 \\ \therefore (2x + 3)(x - 5) &= 0 \\ \therefore (2x + 3) = 0 \text{ or } (x - 5) &= 0 \\ \therefore x = -\frac{3}{2} \text{ or } x = 5\end{aligned}$$

Illustration 13 : If $f(x) = x^2 - 4x + 8$ then for which value of x , is $f(2x) = 2f(x)$?

$$\begin{aligned}f(x) &= x^2 - 4x + 8 \\ \therefore f(2x) &= (2x)^2 - 4(2x) + 8 \\ &= 4x^2 - 8x + 8 \\ \text{Now, } f(2x) &= 2f(x) \\ \therefore 4x^2 - 8x + 8 &= 2x^2 - 8x + 16 \\ \therefore 2x^2 - 8 &= 0 \\ \therefore x^2 - 4 &= 0 \\ \therefore x^2 &= 4 \\ \therefore x &= \pm 2 \\ \therefore \text{When } x = \pm 2, f(2x) &= 2f(x).\end{aligned}$$

Summary

- Unique relation between elements of two non-empty sets is called function.
- If $f: A \rightarrow B$ is a function then set A is called domain and B is called co-domain.
- A set of all images obtained for each value of the domain is called range of a function.
- Range of a function is subset of the co-domain or is the co-domain itself.
- If for any two different values of the domain of a function, their images in co-domain are also different then it is called one - one function.
- If for any two different values of the domain of a function, their images in co-domain are same then it is called many - one function.
- If for each value of the domain of a function, their images in co-domain are same then it is called constant function.
- For equality of two different functions, they must be defined on same domain and their functional values for each value of the domain must be same.
- If the domain of a function is a subset of real set \mathbb{R} then the function f is called a function of real variable.
- If the domain and co-domain of a function are defined on real set \mathbb{R} then it is called real function.

EXERCISE 8

Section A

Find the correct option for the following multiple choice questions :

- Which of the following statement is true ?
 - $f: \{1, 2, 3, 4\} \rightarrow \{3, 4, 5\}$, the rule 'add 2 to the elements of domain' is not a function.
 - $f: A \rightarrow B$, $A = \{-2, -1, 0, 1, 2\}$, $B = \{0, 1, 2, 3, 4\}$, $f(x) = x^2$ is not a function.
 - $g: P \rightarrow Q$, $P = \{-1, 0, 1\}$, $Q = \{-\frac{1}{3}, -1, 3\}$, $g(x) = \frac{x+2}{x-2}$, then g is called function.
 - $g: \{2, 3, 4, 5\} \rightarrow \{-1, 0, 1\}$ and $g(x) = 4x - 3$ is a function.
- Which of the following statements is true for the range of function $f: A \rightarrow B$?
 - $f(A) = \{f(x) \mid x \in A\}$
 - It is not a co-domain or subset of co-domain.
 - Domain itself is the range
 - $f(A) = \{f(x) \mid x \in B\}$
- Which of the following statements is true for the relation $g: X \rightarrow Y$, $X = \{-1, 0\}$, $Y = \{2, 4\}$, $g(x) = 4 - 2x$?
 - g is called a function.
 - g is not a function.
 - X is called function.
 - Y is called function.
- What is the type of function $f: A \rightarrow B$, wherein, for two different values of domain their functional values are same ?
 - One – one function
 - Many – one function
 - One – many function
 - Many – many function
- What is the type of function $f: A \rightarrow B$, where each value of domain A has the same image in set B ?
 - Not a function
 - One – one function
 - Constant function
 - Many – one function
- Which of the following statements is true for a one–one function ?
 - Only for two values of the domain, their images should be different.
 - For any two values of the domain, their images are same.
 - For any two different values of the domain, their images are different.
 - For each value of the domain, their images are same.
- What is the type of function $f: Z - \{0\} \rightarrow N$ and $f(x) = x^2$, $x \in Z - \{0\}$?
 - One–one function
 - Many–one function
 - Constant function
 - None of the above
- Which of the following is a sufficient condition for two different functions to be equal ?
 - Domains of both the functions must be same.
 - Ranges of both the functions must be same.
 - (a) and (b)
 - (a) or (b)

Section B

Answer the following questions in one sentence :

1. Give the necessary condition for defining a function.
2. $f: A \rightarrow B$, $A = \{-3, -1, 1, 3\}$, $B = \{1, 0, 9\}$, $f(x) = x^2$. Is f a function ?
3. $g: \mathbb{N} \rightarrow \mathbb{N}$, 'subtract 2 from the elements of the domain'. Can this rule be called a function ?
4. Define one-one function in notations.
5. Define many-one function in notations.
6. Define constant function in notations.
7. $f: \{1, 2, 3\} \rightarrow \mathbb{N}$, $g: \{2, 3, 4\} \rightarrow \mathbb{N}$, $f(x) = 2x + 1$ and $g(x) = x - 1$. Can these two functions f and g be equal functions? Why ?
8. $f: \mathbb{Z} \rightarrow \mathbb{N}$, $f(t) = t^2 + 1$ $t \in \mathbb{Z}$. Determine the type of function f .
9. $f: \mathbb{N} \rightarrow \mathbb{N}$, $f(t) = t^2 + 1$ $t \in \mathbb{N}$. Determine the type of function f .
10. Define a function of real variable.

Section C

Answer the following questions :

1. Give definition of a function.
2. Define domain and co-domain of a function.
3. Define range of a function.
4. $g: A \rightarrow \mathbb{N}$, $A = \{x \mid x \in \mathbb{N}, 1 < x \leq 4\}$, $g(x) = x + 1$. Find range of function g .
5. $k: X \rightarrow Y$, $X = \{t \mid t \in \mathbb{Z}; -3 \leq t \leq 3\}$, $Y = \{a \mid a \in \mathbb{N}, 1 \leq a \leq 20\}$, $k(t) = t^2 + 2$. State the type of function k .
6. $h: A \rightarrow B$, $A = \{1, 2, 3\}$, $B = \{3, 4, 5, 6, 7, 8\}$, $h(x) = x + 5$. State the type of function h .
7. If $P: A \rightarrow B$, $P(x) = 2x - 3$ and $R_f = \{-2, -1, 0\}$ then find domain of the function.
8. If $f(x) = 1 - \frac{1}{1-x^2}$, $x \in \mathbb{R} - \{-1, 1\}$ then find $f(2) - f(-2)$.
9. If domain of $f(x) = \frac{x-3}{x+4}$ is $\{0, 3, 6\}$ then find its range.
10. If $f(x) = \frac{x^2(x+1)^2}{4}$ is a real function then find the value of $f(3) - f(2)$.
11. If $f: \mathbb{R} \rightarrow \mathbb{R}$ and $f(x) = x^2 + 2x - 1$ then state the type of function f .
12. If $f(x) = \frac{2x-4}{x+7}$ is a real function then for which value of x the image is zero ?
13. $f: \mathbb{Z} - \{2\} \rightarrow \mathbb{Z}$, $f(x) = \frac{x^2+x-6}{x-2}$. State the type of the function.
14. For a real function, $f(x) = 6x^3 - 5x + 15$, find the value of $f(0)$.
15. If $f(x) = x^3 - 2x + \frac{1}{x}$ is a real function then find the value of $f(3) + f(-3)$.

Section D

Answer the following questions :

1. For $f: A \rightarrow B$, $A = \{10, 20, 30\}$, $B = \{18, 48, 98, 128, 148\}$, $f(x) = 5x - 2$, obtain domain, co-domain and range.
2. Obtain domain, co-domain and range for $f: P \rightarrow Q$, $P = \left\{-\frac{1}{2}, 1, \frac{1}{2}, \frac{3}{2}\right\}$, $Q = \left\{-\frac{1}{5}, 1, \frac{1}{3}, 3\right\}$, $f(x) = \frac{x}{2-x}$.
3. If $f: \mathbb{R} - \{0\} \rightarrow \mathbb{R}$, $f(x) = \frac{1}{x} \left(1 + \frac{1}{x}\right) - 1$ then find the value of $f(-1)$, $f(-2)$ and $f\left(\frac{1}{2}\right)$.
4. For the function $f: A \rightarrow B$, $f(x) = 4x - 3$, $R_f = \{9, 13, 17, 25\}$ then find D_f .
5. For the real function $f(x) = 2x^2 - 5x + 4$ find the value of x , for which $f(3x) - 3f(x) + 5 = 0$.
6. If $f: A \rightarrow M$, $A = \{x \mid x \in \mathbb{N}, 1 \leq x < 5\}$ and $M = \{x \mid x \in \mathbb{N}, 1 \leq x \leq 20\}$ and $f(x) = x^2 + 1$ then find the range of f .
7. If $f(x) = \frac{x^2 - 4}{x - 2}$ where $x \in \mathbb{Z} - \{2\}$ then find the value of $f(0) + f(1) - f(-2)$.
8. If the domain of a function $f: A \rightarrow \mathbb{N} \cup \{0\}$, $f(x) = \sqrt{x^2 - 16}$ is $A = \{4, 5\}$ then find its range.
9. If $f(x) = x^2$ and $g(x) = 5x - 6$ where $x \in \{2, 3, 4\}$, check the equality of the functions.
10. If $k: \mathbb{R} \rightarrow \mathbb{R}$, $k(x) = x^2 + 3x - 12$ then determine the type of the function k .
11. If $f(x) = x(3x - 2)$, $g(x) = x^3$ and $x \in \{0, 1, 2\}$ then prove that f and g are equal functions.
12. If $f(x) = \frac{2x + 3}{5x + 2}$, $x \in \mathbb{R} - \left\{\frac{-2}{5}\right\}$ then find the value of $f(2) \cdot f\left(\frac{1}{2}\right)$.
13. If $f(x) = 2x^2 + \frac{1}{x}$, $x \in \mathbb{R} - \{0\}$ obtain the value of $f(3) + f(-3)$.
14. If $f(x) = 15x^3 - 4x^2 + x + 10$, $x \in \mathbb{R}$, obtain the value of $\frac{f(2)}{f(1)}$.
15. If $f: A \rightarrow \mathbb{N} \cup \{0\}$, $A = \{500, 1000, 1300, 1400\}$, $f(x) = \sqrt{5600 - 4x}$, $x \in \mathbb{R}$ then find the value of $f(x)$ for $x = 1000$. Also, for which value of x , $f(x) = 20$?



Pierre-Simon Laplace
(1749 - 1827)

Pierre-Simon Laplace : He was an influential French scholar whose work was important to the development in the fields of mathematics, statistics, physics, and astronomy. His work translated the geometric study of classical mechanics to the one based on calculus, opening up a broader range of problems in statistics. He was a pioneer in the development of classical probability theory. The Bayesian interpretation of probability was developed by Laplace.

Geometric Progression in Nature

Bacteria such as Shewanella Oneidensis multiply by doubling their population in size after as little as 40 minutes. A geometric progression such as this, where each number is double the previous number produces a rapid increase in the population in a very short time.



Geometric Progression

Contents :

- 9.1 Meaning
- 9.2 Formula of obtaining n th term
- 9.3 Meaning of series
- 9.4 Three consecutive terms

9.1 Meaning

A sequence is an ordered arrangement of numbers according to some definite rule. There are different types of sequences. We have already learnt the Arithmetic Progression (A.P.). A sequence in which the difference between any two consecutive terms is a non-zero constant is called an arithmetic progression. e.g. 3, 6, 9, 12,...

We shall now learn about another sequence called Geometric Progression (G.P.).

Let us take a sequence as 3, 6, 12, 24,...

The first term of the sequence is 3. If the first term 3 is multiplied by 2, we get the number $3 \times 2 = 6$ which is the second term of the sequence. If the second term 6 is multiplied by 2, we get the number $6 \times 2 = 12$, which is the third term of the sequence. In other words, the following ratios of consecutive terms of the sequence remain constant:

$$\frac{\text{second term}}{\text{first term}} = 2, \frac{\text{third term}}{\text{second term}} = 2, \frac{\text{fourth term}}{\text{third term}} = 2, \text{ and so on}$$

Consider another sequence 4, -12, 36, -108,...

The first term of the sequence is 4. If the first term 4 is multiplied by -3 , we get the number $4 \times (-3) = -12$, which is the second term of the sequence. If the second term (-12) is multiplied by (-3) , we get the number $(-12) \times (-3) = 36$, which is the third term of the sequence. In other words, the following ratios of consecutive terms of the sequence remain constant :

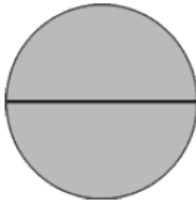
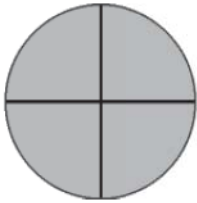
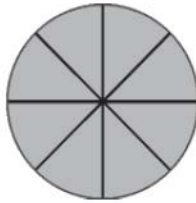

$$\frac{\text{second term}}{\text{first term}} = -3, \quad \frac{\text{third term}}{\text{second term}} = -3, \quad \frac{\text{fourth term}}{\text{third term}} = -3, \text{ and so on.}$$

Consider one more sequence, $2, 1, \frac{1}{2}, \frac{1}{4}, \dots$

The first term of the sequence is 2. If the first term 2 is multiplied by $\frac{1}{2}$, we get the number $2 \times \frac{1}{2} = 1$, which is the second term of the sequence. If the second term 1 is multiplied by $\frac{1}{2}$, we get the number $1 \times \frac{1}{2} = \frac{1}{2}$, which is the third term of the sequence. In other words, the following ratios of consecutive terms of the sequence remain constant :

$$\frac{\text{second term}}{\text{first term}} = \frac{1}{2}, \quad \frac{\text{third term}}{\text{second term}} = \frac{1}{2}, \quad \frac{\text{fourth term}}{\text{third term}} = \frac{1}{2}, \text{ and so on}$$

In each of the above three sequences, it can be seen that for any $n \geq 1$, the ratio of the $(n+1)$ th term to the n th term of the sequence is a non-zero constant. Such a sequence is called **Geometric Progression**.

				
No. of diameters	$1 = 2^0$	$2 = 2^1$	$4 = 2^2$	$8 = 2^3$
No. of parts of circle	$2 = 2^1$	$4 = 2^2$	$8 = 2^3$	$16 = 2^4$

In the above figure, the number of diameters are 1, 2, 4, 8... respectively. They form a G.P.

Also, the number of parts of circle are 2, 4, 8, 16... respectively. They also form a G.P.

Let us understand few practical examples of geometric progression.

Assume that a rich person's wealth is estimated to be ₹ 200 crore and it doubles every five years.

The current wealth is ₹ 200 crore, it doubles every five years. So, the wealth after first five years will be $200 \times 2 = 400$ crore rupees. Now it again doubles at the end of the next five years, so the wealth becomes $400 \times 2 = 800$ crore rupees. Thus, his wealth respectively will be 200, 400, 800 (Crore) which is a geometric progression with ratio 2.





A person deposits an amount of ₹ 10,000 with a bank. The bank offers him an annual return of 10 % compound interest on his investment, i.e. the deposited sum will increase by an interest of 10 % at the end of each year.

The amount deposited is ₹ 10,000. Bank offers 10% interest, so that amount after 1 year will be $10000 + 1000$ (10% of 10000) = 11000. Now the bank counts 10% interest on ₹ 11,000. The amount after the 2nd year will be $11,000 + 1100$ (10 % of 11000) = 12,100 ₹. The amount at the end of 3rd year will be $12100 + 1210$ (10% of 12100) = 13310 ₹. Thus 10000, 11000, 12100, 13310... forms a geometric progression with ratio 1.1.

A crude oil reserve at a particular place is 5 million metric tonnes and is diminishing by 10 % each year.

The current oil reserve is 50,00,000 metric tonnes. Oil reserve diminishes by 10% every year, so that oil reserve at the end of first year is $5000000 - 500000$ (10% of 5000000) = 4500000 metric tonnes. The oil reserve after 2nd year $4500000 - 450000$ (10 % of 4500000) = 4050000 metric tonnes. Thus oil reserves respectively will 5000000, 4500000, 4050000... which is a geometric progression with ratio 0.9.



9.2 Formula for obtaining n th term

If a and r are non-zero real numbers, the sequence whose n th term is $T_n = ar^{n-1}$ for an integer $n \geq 1$ is called the **geometric progression**. The real numbers a and r are called the first term and the common ratio respectively of the geometric progression.

From the definition of the G.P., it can be seen that the consecutive terms of G.P. are a, ar, ar^2, ar^3, \dots and its n th term, $T_n = ar^{n-1}$ for $n \geq 1$ is called the general formula (or term) of the G.P.

Illustration 1 : If the first term and the common ratio of a G.P. are 7 and 2 respectively, find its sixth term.

In a given G.P., the first term $a = 7$, common ratio $r = 2$ and we require the sixth term of G.P. i.e. $n = 6$.

Putting values of a, r and n in the general term $T_n = ar^{n-1}$, we get,

$$\begin{aligned} T_6 &= 7 \times (2)^{6-1} \\ &= 7 \times (2)^5 \\ &= 7 \times 32 \\ &= 224 \end{aligned}$$

Hence, the sixth term of the G.P. is 224.

Illustration 2 : The common ratio and the fifth term of a G.P. are 3 and 324 respectively, find the first term of the G.P.

In the given G.P., the common ratio $r = 3$, the fifth term is 324 i.e. $T_5 = 324$ and we require the first term of the G.P. i.e. a .

Here, $T_5 = 324$

$$\therefore ar^{5-1} = 324 \quad (\because T_n = ar^{n-1})$$

$$\therefore a(3)^4 = 324 \quad (\because r = 3)$$

$$\therefore a \times 81 = 324$$

$$\therefore a = \frac{324}{81} = 4$$

Hence, the first term of the G.P. is 4.

Illustration 3 : The first term and the fourth term of a G.P. are 5 and 40 respectively; find the common ratio of a G.P.

In this G.P., the first term $a = 5$, the fourth term is 40 i.e. $T_4 = 40$ and we want to find the common ratio of the G.P. i.e. r

Here, $T_4 = 40$

$$\therefore ar^{4-1} = 40 \quad (\because T_n = ar^{n-1})$$

$$\therefore 5 \times r^3 = 40 \quad (\because a = 5)$$

$$\therefore r^3 = \frac{40}{5} = 8$$

$$\therefore r = 2$$

Hence, the common ratio of the G.P. is 2.

Illustration 4 : The first term and the common ratio of a G.P. are 4 and -2 respectively. If its n th term is -128 , find the value of n .

For the given G.P. first term $a = 4$, common ratio $r = -2$ and the n -th term is -128 i.e. $T_n = -128$.

Here, $T_n = -128$

$$\therefore ar^{n-1} = -128 \quad (\because T_n = ar^{n-1})$$

$$\therefore 4 \times (-2)^{n-1} = -128 \quad (\because a = 4 \text{ and } r = -2)$$

$$\therefore (-2)^{n-1} = \frac{-128}{4} = -32$$

$$\therefore (-2)^{n-1} = (-2)^5$$

Equating the powers on both the sides, we get

$$n - 1 = 5$$

$$\therefore n = 6$$

Illustration 5 : If the fourth term and the seventh term are $\frac{3}{4}$ and $\frac{3}{32}$ respectively for a G.P., find its tenth term.

Let a and r be the first term and common ratio respectively of the given G.P.

We are given $T_4 = \frac{3}{4}$ and $T_7 = \frac{3}{32}$

Hence, $\frac{T_7}{T_4} = \frac{ar^6}{ar^3} = r^3$ and $\frac{T_7}{T_4} = \frac{\frac{3}{32}}{\frac{3}{4}} = \frac{3}{32} \times \frac{4}{3} = \frac{1}{8}$

$$\therefore r^3 = \frac{1}{8} = \left(\frac{1}{2}\right)^3$$

$$\therefore r = \frac{1}{2}$$

Now putting $r = \frac{1}{2}$ in $T_4 = ar^3 = \frac{3}{4}$, we get

$$\therefore a \left(\frac{1}{2}\right)^3 = \frac{3}{4}$$

$$\therefore a \times \frac{1}{8} = \frac{3}{4}$$

$$\therefore a = \frac{3}{4} \times 8$$

$$\therefore a = 6$$

We now have to find the tenth term, so $n = 10$.

Putting values of a , r and n in the general term $T_n = ar^{n-1}$, we get

$$\begin{aligned} T_{10} &= 6 \times \left(\frac{1}{2}\right)^{10-1} \\ &= 6 \times \left(\frac{1}{2}\right)^9 \\ &= 6 \times \frac{1}{512} \\ &= \frac{3}{256} \end{aligned}$$

Hence, the tenth term of the G. P. is $\frac{3}{256}$

Illustration 6 : Find the 5th term of a G.P. 9, -6, 4, ...

Here, $a = 9$ and $r = \frac{-6}{9} = \frac{-2}{3}$. The 5th term is to be found, so $n = 5$

Putting values of a , r and n in the general term $T_n = ar^{n-1}$, we get

$$\begin{aligned} T_5 &= 9 \times \left(\frac{-2}{3}\right)^{5-1} \\ &= 9 \times \left(\frac{-2}{3}\right)^4 \\ &= 9 \times \frac{16}{81} \\ &= \frac{16}{9} \end{aligned}$$

Hence, the 5th term of the G. P. is $\frac{16}{9}$.

Illustration 7 : Find the 8th term of the G. P. $\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, \dots$

Here $a = \frac{1}{8}$ and $r = \frac{\frac{1}{4}}{\frac{1}{8}} = 2$. Now, 8th term is to be found, so $n = 8$

Putting the values of a, r and n in the general term $T_n = ar^{n-1}$, we get

$$\begin{aligned}T_8 &= \frac{1}{8} \times (2)^{8-1} \\ &= \frac{1}{8} \times (2)^7 \\ &= \frac{1}{8} \times 128 \\ &= 16\end{aligned}$$

Hence, the 8th term of the G. P. is 16.

Illustration 8 : If the second term of a G.P. is 4 then find the product of the first three terms of the G.P.

Let a and r be the first term and common ratio.

Here, the second term is 4. i.e. $T_2 = 4$.

Now putting $n = 1, n = 2$ and $n = 3$ in the general term $T_n = ar^{n-1}$, we get

$$T_1 = a, T_2 = ar \text{ and } T_3 = ar^2$$

Hence, $T_1 \times T_2 \times T_3 = a \times ar \times ar^2$

$$\begin{aligned}&= a^3 \times r^3 \\ &= (ar)^3 \\ &= (4)^3 \qquad (\because T_2 = ar = 4) \\ &= 64\end{aligned}$$

Hence, the product of the first three terms of the G.P. is 64.

Illustration 9 : The first term and the product of the first three terms of a G.P. are 3 and 216 respectively. Find the 7th term of the G.P.

The first term $a = 3$.

Now, putting $n = 1, n = 2$ and $n = 3$ in the general term $T_n = ar^{n-1}$, we get $T_1 = a, T_2 = ar$ and $T_3 = ar^2$

It is given that product of first three terms is 216.

$$\text{i.e. } T_1 \times T_2 \times T_3 = 216$$

Now, $T_1 \times T_2 \times T_3 = 216$

$$\therefore a \times ar \times ar^2 = 216$$

$$\therefore a^3 \times r^3 = 216$$

$$\therefore 3^3 \times r^3 = 216 \qquad (\because a = 3)$$

$$\therefore 27 \times r^3 = 216$$

$$\therefore r^3 = \frac{216}{27} = 8$$

$$\therefore r = 2$$

The 7th term of the G.P. is $T_7 = ar^6$

$$\begin{aligned}\therefore T_7 &= 3 \times (2)^6 \\ &= 3 \times 64 \\ &= 192\end{aligned}$$

Hence, the 7th term of the G. P. is 192.

Illustration 10 : The numbers 2, G, 50 are in G.P. Find the value of G.

Here $T_1 = 2$, $T_2 = G$ and $T_3 = 50$

The common ratio of the G. P. is $= \frac{T_2}{T_1} = \frac{T_3}{T_2}$

$$\therefore \frac{G}{2} = \frac{50}{G}$$

$$\therefore G^2 = 100$$

$$\therefore G = \pm 10$$

$$\therefore G = 10 \text{ or } -10$$

Illustration 11 : For the numbers a, b, c, d , if $\frac{b}{a} = \frac{c}{b} = \frac{d}{c}$, show that the numbers a, b, c, d are in G.P.

Suppose $\frac{b}{a} = \frac{c}{b} = \frac{d}{c} = r$, where r is a non-zero constant.

$$\therefore b = ar, c = br, d = cr$$

$$\therefore c = (ar)r = ar^2 \text{ and } d = (ar^2)r = ar^3$$

If we take $T_1 = a$, $T_2 = b = ar$, $T_3 = c = ar^2$, $T_4 = d = ar^3$

then we see that T_1, T_2, T_3 and T_4 are in G.P. Hence a, b, c, d are in G.P.

Illustration 12 : Which term of the G.P. 0.008, 0.016, 0.032, ... is 4.096 ?

Here, the first term $a = 0.008$ and the common ratio $r = \frac{0.016}{0.008} = 2$

Now, $T_n = 4.096$

$$\therefore ar^{n-1} = 4.096$$

$$\therefore 0.008 \times (2)^{n-1} = 4.096$$

$$\therefore 2^{n-1} = \frac{4.096}{0.008}$$

$$\therefore 2^{n-1} = 512$$

$$\therefore 2^{n-1} = 2^9$$

Equating the powers on both the sides, we get

$$n - 1 = 9$$

$$\therefore n = 10$$

Hence, 4.096 is the 10th term of the given G.P.

Illustration 13 : If the third term of a G.P. is the square of the first term and the fourth term is 243, find the sequence.

Let a and r be the first term and common ratio respectively of the given G. P. Here, the third term is square of the first i.e. $T_3 = a^2$

$$\therefore T_3 = ar^2 = a^2$$

$$\therefore r^2 = a$$

Also, the fourth term is 243 i.e. $T_4 = 243$

$$\therefore T_4 = ar^3 = 243$$

$$\therefore r^2 \times r^3 = 243 \quad (\because r^2 = a)$$

$$\therefore r^5 = 243$$

$$\therefore r^5 = 3^5$$

$$\therefore r = 3$$

$$\text{Now, } a = r^2$$

$$\therefore a = 3^2$$

$$\therefore a = 9$$

Taking $a = 9$ and $r = 3$ the sequence is 9, 27, 81, 243, ...

Illustration 14 : A person deposits ₹10,000 in a bank in the year 2009, ₹ 20,000 in the year 2010, ₹ 40,000 in the year 2011. The amount of deposits in any given year is twice the amount of deposit of the previous year. What would be the amount of deposit in the year 2014 ?

A person deposits ₹ 10,000 in the first year i.e. 2009, so the first term $a = 10000$. Every year the amount of deposits is twice the previous year's amount, so the common ratio (r) = 2. We want to find the amount deposited in the year 2014 i.e. the sixth year, so $n = 6$.

Putting the values of a , r and n in the general term $T_n = ar^{n-1}$, we get

$$\begin{aligned} T_6 &= 10000 (2)^{6-1} \\ &= 10,000 (2)^5 \\ &= 10,000 \times 32 \\ &= 3,20,000 \end{aligned}$$

Thus the amount deposited in the year 2014 is ₹ 3,20,000.

Illustration 15 : A water tank of a capacity of 50,000 litres is fully filled with water. Every week, the water level reduces to half of the previous level due to leakage. What will be the level of water after five weeks ? (No new water is added in the tank)

A water tank of a capacity of 50,000 litres is fully filled with water $\therefore a = 50000 = T_1$.

Every week, the water level reduces to half of the previous level \therefore common ratio $r = \frac{1}{2}$.

Water level after first week = $50000 \times \frac{1}{2} = 25000 = T_2$. We want the water level after five weeks, so $n = 6$.

Putting the values of a , r and n in the general term $T_n = ar^{n-1}$, we get

$$\begin{aligned} T_6 &= 50,000 \left(\frac{1}{2}\right)^{6-1} \\ &= 50,000 \left(\frac{1}{2}\right)^5 \\ &= 50,000 \times \frac{1}{32} = 1562.5 \end{aligned}$$

Thus the level of water after five weeks will be 1562.5 litres.

Illustration 16 : Population of a city is 20 lakhs. If the population increases at the rate of 3% every year, find the population of the city after 6 years.

Current population of the city is 20,00,000 i.e. $a = 2000000 = T_1$.

Population increases at the rate of 3% every year

i.e. $r = 1.03$

As the population increases at the rate of 3%,

$$\therefore r = \frac{100+3}{100} = 1.03$$

Population after the first year = $2000000 \times 1.03 = 2060000 = T_2$. We need to find the population after 6 years, so $n = 7$.

Putting the values of a , r and n in the general term $T_n = ar^{n-1}$, we get

$$\begin{aligned} T_7 &= 20,00000 (1.03)^{7-1} \\ &= 20,00000 (1.03)^6 \end{aligned}$$

$$\therefore T_7 = 2388104.5930$$

Thus, population after 6 years will be 23,88,105.

Illustration 17 : Government decides to fix the depreciation rate of a machine to 15 % per year. If the purchase price of a machine is ₹ 50,000. Find the value of the machine after 7 years.

Here purchase price of a machine is ₹ 50,000 i.e. $a = 50,000 = T_1$. Machine depreciates at the rate of 15% every year i.e. $r = 0.85$

As the machine depreciates, its price decreases at the rate of 15% $\therefore r = \frac{100-15}{100} = 0.85$

Value of machine after the first year = $50,000 \times 0.85$

$$= 42,500 = T_2$$

We need to find the value of machine after 7 years, so $n = 8$.

Putting the values of a , r and n in the general term $T_n = ar^{n-1}$,

$$\begin{aligned} T_8 &= 50,000 (0.85)^{8-1} \\ &= 50,000 (0.85)^7 \end{aligned}$$

$$= 16028.8544$$

$$\approx 16028.85$$

Thus, the value of machine after 7 years will be ₹ 16,028.85

9.3 Meaning of Series

If the geometric progression is taken as a, ar, ar^2, ar^3, \dots , then the geometric series is given as $a + ar + ar^2 + ar^3 + \dots$,

e.g. If the geometric progression is 2, 6, 18, 54, ... then its corresponding geometric series is $2 + 6 + 18 + 54 + \dots$

The sum of the first n terms of the G.P. is denoted by the symbol S_n .

Then $S_n = T_1 + T_2 + T_3 + \dots + T_n$

where $T_n = ar^{n-1}$, $n = 1, 2, 3, \dots$

Hence, $S_1 = T_1$, $S_2 = T_1 + T_2$, $S_3 = T_1 + T_2 + T_3$ and so on.

If we write all the terms of S_n in terms of a and r then we have

$$S_n = a + ar + ar^2 + \dots + ar^{n-1}$$

We shall accept the following results relating to the geometric progression without proof :

(1) $\frac{T_{n+1}}{T_n} = r = \text{common ratio for any positive integer } n.$

(2) $T_{n+1} = S_{n+1} - S_n$ where $n = 1, 2, 3, \dots$

(3) $S_n = na$ where $r = 1$

(4) $S_n = \frac{a(r^n - 1)}{(r - 1)}$ where $r \neq 1$

(5) $S_n = \frac{rT_n - a}{(r - 1)}$ where $r \neq 1$

Illustration 18 : Find the sum of the first five terms of a G.P. 5, 15, 45, ...

Here, the first term $a = 5$ and the common ratio $r = \frac{15}{5} = 3$.

Sum of the first five terms is required i.e. $n = 5$.

Putting the values of a , r and n in $S_n = \frac{a(r^n - 1)}{(r - 1)}$, we get

$$\begin{aligned} S_5 &= \frac{5(3^5 - 1)}{(3 - 1)} \\ &= \frac{5(243 - 1)}{2} \\ &= \frac{5(242)}{2} \\ &= 605 \end{aligned}$$

Thus, the sum of the first five terms of the G. P. is 605.

Illustration 19 : Find the sum of the first six terms of a G.P. 8, 4, 2, ...

Here, the first term $a = 8$ and the common ratio $r = \frac{4}{8} = \frac{1}{2}$

Sum of the first six terms is required i.e. $n = 6$.

Putting the values of a , r and n in $S_n = \frac{a(r^n - 1)}{(r - 1)}$, we get

$$\begin{aligned} \therefore S_6 &= \frac{8\left[\left(\frac{1}{2}\right)^6 - 1\right]}{\left(\frac{1}{2} - 1\right)} \\ &= \frac{8\left[\frac{1}{64} - 1\right]}{\left(-\frac{1}{2}\right)} \\ &= \frac{8\left[-\frac{63}{64}\right]}{\left(-\frac{1}{2}\right)} \\ &= 8 \times \left(-\frac{63}{64}\right) \times \left(\frac{2}{-1}\right) \\ &= \frac{63}{4} \end{aligned}$$

Thus, the sum of the first six terms of the G. P. is $\frac{63}{4}$.

Illustration 20 : Find the sum of the first four terms of the G.P. whose first term is 3 and common ratio is 2.

Here, the first term $a = 3$ and the common ratio $r = 2$. Sum of the first four terms is required i.e. $n = 4$.

Putting the values of a , r and n in $S_n = \frac{a(r^n - 1)}{(r - 1)}$, we get

$$\begin{aligned}
S_4 &= \frac{3[2^4-1]}{(2-1)} \\
&= \frac{3(16-1)}{1} \\
&= 3 \times 15 \\
&= 45
\end{aligned}$$

Thus, the sum of the first four terms of the G. P. is 45.

Illustration 21 : The sum of the first three terms of a G.P. with common ratio 0.2 is 0.496. Find the first term of the G.P.

Here, the common ratio $r = 0.2$ and sum of the first three term is 0.496. i.e. $S_3 = 0.496$

Putting the values of a , r and n in $S_n = \frac{a(r^n-1)}{(r-1)}$, we get

$$0.496 = \frac{a[0.2^3-1]}{0.2-1}$$

$$\therefore 0.496 = \frac{a(0.008-1)}{(-0.8)}$$

$$\therefore 0.496 = \frac{a(-0.992)}{(-0.8)}$$

$$\therefore a = \frac{0.496 \times (-0.8)}{(-0.992)}$$

$$\therefore a = 0.4$$

Thus, the first term of the G. P. is 0.4.

Illustration 22 : Sum of how many terms of a G.P. 800,400,200, ... is 1500 ?

Here, the first term $a = 800$ and the common ratio $r = \frac{400}{800} = 0.5$. Sum of n terms is 1500

i.e. $S_n = 1500$

Putting the values of a , r and n in $S_n = \frac{a(r^n-1)}{(r-1)}$, we get

$$S_n = \frac{800[(0.5)^n-1]}{(0.5-1)}$$

$$\therefore 1500 = \frac{800[(0.5)^n-1]}{(-0.5)}$$

$$\therefore (0.5)^n - 1 = \frac{1500 \times (-0.5)}{800}$$

$$\therefore (0.5)^n = 1 - 0.9375$$

$$\therefore (0.5)^n = 0.0625$$

$$(0.5)^n = (0.5)^4$$

Equating the powers on both the sides, we get

$$n = 4$$

Thus, the sum of the first four terms of the G. P. is 1500.

Illustration 23 : In a G.P., the first term is 27 and sum of the first three terms is 189. Find the common ratio of the G.P.

Here the first term $a = 27$ and sum of the first three terms is 189 i.e. $S_3 = 189$.

Putting the values in $S_n = \frac{a(r^n - 1)}{(r - 1)}$, we get

$$S_3 = \frac{27(r^3 - 1)}{(r - 1)}$$

$$\therefore 189 = \frac{27(r - 1)(r^2 + r + 1)}{(r - 1)}$$

$$\therefore 189 = 27(r^2 + r + 1)$$

$$\therefore r^2 + r + 1 = 7$$

$$\therefore r^2 + r - 6 = 0$$

$$\therefore (r + 3)(r - 2) = 0$$

$$\therefore r = -3 \text{ OR } r = 2$$

Thus, the common ratio of the G. P. is -3 or 2 .

Illustration 24 : The first term and the sum of the first five terms of the G.P. are equal to 1 each.

Find the common ratio of the G. P.

Here, the first term $a = 1$ and sum of the first five terms is 1 i.e. $S_5 = 1$.

Putting the values in $S_n = \frac{a(r^n - 1)}{(r - 1)}$, we get

$$S_5 = \frac{1(r^5 - 1)}{(r - 1)}$$

$$\therefore 1 = \frac{(r^5 - 1)}{(r - 1)}$$

$$\therefore r^5 - 1 = r - 1$$

$$\therefore r^5 = r$$

$$\therefore r^4 = 1$$

$$\therefore r = \pm 1$$

If we take $r = 1$ and $a = 1$ then the first five terms of the G. P. will be 1, 1, 1, 1, 1 whose sum is 5.

If we take $r = -1$ and $a = 1$ then the first five terms of the G. P. will be 1, -1 , 1, -1 , 1 whose sum is 1.

But, in the data, the sum of the first five terms is given as 1, so the only possible value of r is -1 .

Thus the common ratio of the G. P. is -1 .

Illustration 25 : Find the maximum value of n such that the sum of the first n terms of a G. P.

2, 4, 8, 16..... does not exceed 5000.

Here, the first term $a = 2$ and the common ratio $r = \frac{4}{2} = 2$.

Sum of the first n terms should not exceed 5000 i.e. $S_n \leq 5000$.

Putting the values of a , r and n in $S_n = \frac{a(r^n - 1)}{(r - 1)}$, we get

$$S_n = \frac{2(2^n - 1)}{(2 - 1)}$$

$$= 2(2^n - 1)$$

Since $S_n \leq 5000$, we have

$$2(2^n - 1) \leq 5000$$

$$\therefore 2^n - 1 \leq 2500$$

$$\therefore 2^n \leq 2501$$

We now tabulate values of 2^n for different positive values of n as shown in the following table and we shall take the maximum value of n for which $2^n \leq 2501$

n	4	5	6	7	8	9	10	11	12
2^n	16	32	64	128	256	512	1024	2048	4096

We can see from the table that for $n = 10$, $2^n = 1024$, for $n = 11$, $2^n = 2048$ and for $n = 12$, $2^n = 4096$ which exceeds 2501 so the maximum value of n for which 2^n does not exceed 2501 is 11.

$$\therefore n = 11$$

Illustration 26 : Find the minimum value of n such that the sum of the first n terms of a G. P. 1, 3, 3^2 , 3^3 , ... is greater than or equal to 3000.

Here, the first term $a = 1$ and the common ratio $r = \frac{3}{1} = 3$. Sum of the first n terms should be greater than or equal to 3000 i.e. $S_n \geq 3000$.

Putting the values of a , r and n in $S_n = \frac{a(r^n - 1)}{(r - 1)}$, we get,

$$\begin{aligned} S_n &= \frac{1(3^n - 1)}{(3 - 1)} \\ &= \frac{3^n - 1}{2} \end{aligned}$$

Since, $S_n \geq 3000$, we have

$$\frac{3^n - 1}{2} \geq 3000$$

$$\therefore 3^n - 1 \geq 6000$$

$$\therefore 3^n \geq 6001$$

We now tabulate values of 3^n for different positive values of n as shown in the following table and we shall take minimum value of n for which $3^n \geq 6001$.

n	4	5	6	7	8	9
3^n	81	243	729	2187	6561	19683

We can see from the table that when we take $n = 8, 9, \dots, 3^n$ exceeds 6001

Thus, the minimum value of n for which 3^n is greater than or equal to 6001 is 8.

$$\therefore n = 8$$

Illustration 27 : If for a G.P., $S_8 = 10S_4$ find 'r'.

Here, $S_8 = 10S_4$

$$\therefore \frac{a(r^8-1)}{(r-1)} = 10 \left[\frac{a(r^4-1)}{(r-1)} \right]$$

$$\therefore r^8 - 1 = 10(r^4 - 1)$$

$$\therefore (r^4 - 1)(r^4 + 1) = 10(r^4 - 1)$$

$$\therefore r^4 + 1 = 10$$

$$\therefore r^4 = 9$$

$$\therefore r^2 = 3$$

$$\therefore r = \pm \sqrt{3}$$

Illustration 28 : A person gives ₹ 5 to his son on 1st March, ₹ 10 on 2nd March, ₹ 20 on 3rd March and so on. Thus each day he gives double the amount than that of the previous day. Find the total amount he has given to his son upto 10th of March.

The person gives ₹ 5 on 1st March, so the first term $a = 5$. Every day, the amount given to his son is twice the previous day's amount, so the common ratio $r = 2$.

We want to find the total amount given by him to his son upto 10th March i.e. S_{10} , so $n = 10$.

Putting the values of a , r and n in $S_n = \frac{a(r^n-1)}{(r-1)}$, we get,

$$\begin{aligned} S_{10} &= \frac{5(2^{10}-1)}{(2-1)} \\ &= \frac{5(1024-1)}{1} \\ &= 5 \times 1023 \\ &= 5115 \end{aligned}$$

Hence, the total amount the person gives to his son upto 10th March is ₹ 5115.

Illustration 29 : A person wants to donate ₹ 2,42,000 in five months such that every month he donates one-third of the amount he donated in the previous month. Find the amount he donated in the first month.

The person wants to donate ₹ 242000 in five months i.e. $S_5 = 242000$ and $n = 5$. Every month, he donates one-third of the amount he donated in the previous month, so the common ratio $r = \frac{1}{3}$. We want to find the amount donated in the first month i.e. a .

Putting the values of a , r and n in $S_n = \frac{a(r^n-1)}{(r-1)}$, we get,

$$S_5 = \frac{a \left[\left(\frac{1}{3} \right)^5 - 1 \right]}{\left(\frac{1}{3} - 1 \right)}$$

$$\therefore 242000 = \frac{a\left(\frac{1}{243}-1\right)}{\left(-\frac{2}{3}\right)}$$

$$\therefore 242000 = \frac{a\left(-\frac{242}{243}\right)}{\left(-\frac{2}{3}\right)}$$

$$\therefore 242000 = a \times \left(\frac{-242}{243}\right) \times \left(\frac{3}{-2}\right)$$

$$\therefore 242000 = a \times \frac{121}{81}$$

$$\therefore a = \frac{242000 \times 81}{121}$$

$$\therefore a = 162000$$

Thus, the amount donated in the first month is ₹ 1,62,000.

Illustration 30 : A person deposits ₹ 20,000 in a bank at the compound interest rate of 8% per annum. Find the amount the person receives after 5 years.

The person deposits ₹ 20,000 i.e. $a = 20,000 = T_1$

Amount received after first year $20,000 \times 1.08 = 21,600 = T_2$

The rate of interest is 8% so the common ratio $r = 1.08$

We want to find the amount received after 5 years,

so $n = 6$

Putting the values of a , r and n in the general term $T_n = ar^{n-1}$, we get,

$$\begin{aligned} T_6 &= 20,000 \times (1.08)^{6-1} \\ &= 20,000 \times (1.08)^5 \\ &= 29386.5615 \\ &\approx 29386.56 \end{aligned}$$

Thus, amount received after 5 years will be ₹ 29,386.56.

As the amount increases at the rate of 8%,

$$r = \frac{100+8}{100} = 1.08$$

Illustration 31 : If three positive numbers $k + 1$, $3k - 1$, $5k + 1$ are in G.P., find the value of k .

Here, $k + 1$, $3k - 1$, $5k + 1$ are in G. P., so $\frac{T_2}{T_1} = \frac{T_3}{T_2} =$ common ratio r

$$\therefore \frac{3k-1}{k+1} = \frac{5k+1}{3k-1}$$

$$\therefore (3k - 1)^2 = (5k + 1)(k + 1)$$

$$\therefore 9k^2 - 6k + 1 = 5k^2 + 6k + 1$$

$$\therefore 4k^2 - 12k = 0$$

$$\therefore 4k(k - 3) = 0$$

$$\therefore 4k = 0 \text{ or } k - 3 = 0$$

$$\therefore k = 0 \text{ or } k = 3$$

But $k = 0$ is not possible because for $k = 0$, the value of $(3k - 1)$ will be -1 which a negative number.

Hence, $k = 3$

Illustration 32 : If 15, x , 240, y are in G.P., find the values of x and y .

Here, 15, x , 240, y are in G. P. so $\frac{T_2}{T_1} = \frac{T_3}{T_2} = \frac{T_4}{T_3} = \text{common ratio } (r)$

Let us take $\frac{T_2}{T_1} = \frac{T_3}{T_2}$,

$$\therefore \frac{x}{15} = \frac{240}{x}$$

$$\therefore x^2 = 15 \times 240$$

$$= 3600$$

$$\therefore x = \pm 60$$

$$\therefore x = 60 \text{ or } -60$$

Also, $\frac{T_3}{T_2} = \frac{T_4}{T_3}$

$$\therefore \frac{240}{x} = \frac{y}{240}$$

$$\therefore xy = 240 \times 240$$

$$\therefore xy = 57600$$

Now, if we take $x = 60$

$$60y = 57600$$

$$\therefore y = \frac{57600}{60}$$

$$\therefore y = 960$$

and if we take $x = -60$

$$-60y = 57600$$

$$\therefore y = \frac{57600}{-60}$$

$$\therefore y = -960$$

Thus, $x = 60$ and $y = 960$ or $x = -60$ and $y = -960$.

Illustration 33 : If for a G.P., $T_n = 80$, $S_n = 157.5$ and $r = 2$, find a and n .

Here, values of T_n and S_n are given, so we shall use the formula

$$S_n = \frac{rT_n - a}{r - 1}$$

Putting the values of T_n , r and S_n in $S_n = \frac{rT_n - a}{r - 1}$

$$\therefore 157.5 = \frac{2 \times 80 - a}{2 - 1}$$

$$\therefore 157.5 = 160 - a$$

$$\therefore a = 160 - 157.5$$

$$\therefore a = 2.5$$

Now $T_n = ar^{n-1}$

$$\therefore 80 = 2.5 \times (2)^{n-1}$$

$$\therefore 2^{n-1} = 32$$

$$\therefore 2^{n-1} = 2^5$$

Equating the powers on both the sides, we get

$$n - 1 = 5$$

$$\therefore n = 6$$

Thus, $a = 2.5$ and $n = 6$

Illustration 34 : If for a G.P., $T_n = 2^{n+1}$, obtain S_4 .

Here, $T_n = 2^{n+1}$

$$\therefore T_1 = 2^{1+1} = 4, T_2 = 2^{2+1} = 8, T_3 = 2^{3+1} = 16, \dots$$

Here, the first term $a = 4$ and the common ratio $r = \frac{8}{4} = 2$. Sum of the first four terms is required

i.e. $n = 4$.

Putting the values of a , r and n in $S_n = \frac{a(r^n - 1)}{(r - 1)}$, we get

$$\begin{aligned} S_4 &= \frac{4(2^4 - 1)}{(2 - 1)} \\ &= \frac{4(16 - 1)}{1} \\ &= 4 \times 15 \\ &= 60 \end{aligned}$$

Thus, the sum of the first four terms of the G. P. is 60.

Illustration 35 : If in a G. P., $S_n = \frac{2}{3}(4^n - 1)$, obtain T_{n+1} .

$$\begin{aligned} \text{We know } T_{n+1} &= S_{n+1} - S_n \\ &= \frac{2}{3}[4^{n+1} - 1] - \frac{2}{3}[4^n - 1] \\ &= \frac{2}{3}[(4^{n+1} - 1) - (4^n - 1)] \\ &= \frac{2}{3}[4^{n+1} - 1 - 4^n + 1] \\ &= \frac{2}{3}[4^{n+1} - 4^n] \\ &= \frac{2}{3} \times 4^n [4 - 1] \end{aligned}$$

$$\therefore T_{n+1} = 2(4^n)$$

Illustration 36 : If in a G. P., $S_n = \frac{4}{3}(3^n - 1)$, find T_3 .

$$\begin{aligned} \text{We know } T_{n+1} &= S_{n+1} - S_n \\ \therefore T_3 &= S_3 - S_2 \\ &= \frac{4}{3}(3^3 - 1) - \frac{4}{3}(3^2 - 1) \\ &= \frac{4}{3}[(27 - 1) - (9 - 1)] \\ &= \frac{4}{3}[26 - 8] \\ &= \frac{4}{3}(18) \\ \therefore T_3 &= 24 \end{aligned}$$

9.4 Three consecutive terms of geometric progression

At times, the sum and the product of some consecutive terms of a G.P. are given and these consecutive terms are to be found. These terms can be found if we take the terms of the G.P. as a , ar , ar^2 , ar^3 , The calculations of the terms selected in this way may, most likely, become cumbersome. The mathematical forms of the terms of the G.P. are to be assumed in such a way that the calculation of the terms becomes simple. The assumptions of such forms of the terms of the G.P., for some selected values of n , are given below:

For $n = 3$, three consecutive terms are : $\frac{a}{r}$, a , ar

For $n = 4$, four consecutive terms are : $\frac{a}{r^3}$, $\frac{a}{r}$, ar , ar^3

For $n = 5$, five consecutive terms are : $\frac{a}{r^2}$, $\frac{a}{r}$, a , ar , ar^2

Note : We shall restrict our study of finding the unknown terms of the G.P. for three terms only.

Illustration 37 : The sum and the product of the three consecutive terms of a G.P. are 26 and 216 respectively. Find the three terms of the G.P.

Let us assume three consecutive terms of G.P. as $\frac{a}{r}$, a , ar

Here, the product of the terms = 216

$$\therefore \frac{a}{r} \times a \times ar = 216$$

$$\therefore a^3 = 216$$

$$\therefore a = 6$$

Now, the sum of the terms = 26

$$\therefore \frac{a}{r} + a + ar = 26$$

$$\therefore a \left(\frac{1}{r} + 1 + r \right) = 26$$

$$\therefore 6 \left(\frac{1+r+r^2}{r} \right) = 26 \quad (\because a = 6)$$

$$\therefore 3(1 + r + r^2) = 13r$$

$$\therefore 3 + 3r + 3r^2 = 13r$$

$$\therefore 3r^2 - 10r + 3 = 0$$

$$\therefore (r - 3)(3r - 1) = 0$$

$$\therefore r = 3 \text{ or } r = \frac{1}{3}$$

Now, if we take $a = 6$ and $r = 3$ then three consecutive terms will be $\frac{6}{3} = 2$, 6 , $6 \times 3 = 18$.

Hence, three consecutive terms are 2, 6, 18.

Now, if we take $a = 6$ and $r = \frac{1}{3}$ then three consecutive terms will be $\frac{6}{1/3} = 18$, 6 , $6 \times \frac{1}{3} = 2$

Hence, three consecutive terms are 18, 6 and 2.

Illustration 38 : The sum and the product of the three consecutive terms of a G.P. are 9.5 and 27 respectively. Find the three terms of the G.P.

Let us assume three consecutive terms of G.P. as $\frac{a}{r}$, a , ar

Here the product of the terms = 27

$$\therefore \frac{a}{r} \times a \times ar = 27$$

$$\therefore a^3 = 27$$

$$\therefore a = 3$$

Now the sum of the terms = 9.5

$$\therefore \frac{a}{r} + a + ar = 9.5$$

$$\therefore a \left(\frac{1}{r} + 1 + r \right) = 9.5$$

$$\therefore 3 \left(\frac{1+r+r^2}{r} \right) = 9.5 \quad (\because a = 3)$$

$$\therefore 3 + 3r + 3r^2 = 9.5r$$

$$\therefore 3r^2 - 6.5r + 3 = 0$$

Multiplying by 2 on both the sides, we get

$$\therefore 6r^2 - 13r + 6 = 0$$

$$\therefore (2r - 3)(3r - 2) = 0$$

$$\therefore r = \frac{3}{2} \text{ or } r = \frac{2}{3}$$

Now, if we take $a = 3$ and $r = \frac{3}{2}$ then three consecutive terms will be $\frac{3}{\left(\frac{3}{2}\right)} = 2$, 3 , $3 \times \frac{3}{2} = \frac{9}{2}$

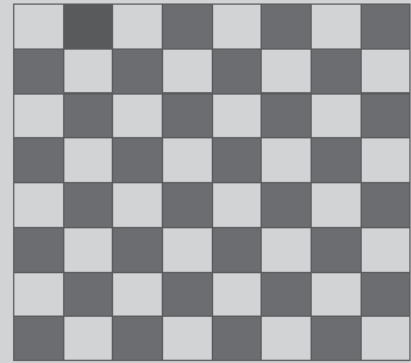
Hence, three consecutive terms are 2 , 3 , $\frac{9}{2}$

Now, if we take $a = 3$ and $r = \frac{2}{3}$ then three consecutive terms will be $\frac{3}{\frac{2}{3}} = \frac{9}{2}$, 3 , $3 \times \frac{2}{3} = 2$

Hence, three consecutive terms are $\frac{9}{2}$, 3 , 2 .

Activity

- (1) Place one grain on 1st square of the chess board, 2 grains on 2nd square of the chess board, 4 grains on 3rd square, 8 grains on 4th square and so on. How many grains will have to be kept on the last square of the chess board ?



(2)

$$\bigcirc \begin{array}{c} 1 \end{array} + \bigcirc \begin{array}{c} 0.1 \end{array} + \bigcirc \begin{array}{c} 0.01 \end{array} + \bigcirc \begin{array}{c} 0.001 \end{array} + \bigcirc \begin{array}{c} 0.0001 \end{array} = \bigcirc \begin{array}{c} 1.1111 \end{array}$$

Verify using Geometric progression.

Hint : Here, the sequence is 1, 0.1, 0.01, 0.001, 0.0001

Summary and Formulae

- For any $n \geq 1$ the ratio of the $(n+1)$ th term to the n -th term of the sequence is a non-zero constant then such a sequence is called **Geometric Progression**.
- If the geometric progression is given as a, ar, ar^2, ar^3, \dots then the **geometric series** is given as $a + ar + ar^2 + ar^3 + \dots$
- n th term $T_n = ar^{n-1}$ for $n \geq 1$ is called the general formula (or term) of the G.P.
- $\frac{T_{n+1}}{T_n} = r =$ common ratio for any positive integer n
- $T_{n+1} = S_{n+1} - S_n ; n = 1, 2, 3, \dots$
- When $r = 1$ then $S_n = na$
- When $r \neq 1$ then $S_n = \frac{a(r^n - 1)}{(r - 1)}$
- When $r \neq 1$ then $S_n = \frac{rT_n - a}{r - 1}$
- Three consecutive terms in G. P. are : $\frac{a}{r}, a, ar$

EXERCISE 9

Section A

For the following multiple choice questions choose the correct option :

- Find the 6th term of a G. P. 0.2, 1, 5, ...
(a) 25 (b) 0.5 (c) 0.1 (d) 625
- If the first term of a G. P. is 'a' and the common ratio is 'b'. Find the $(n + 1)$ th term.
(a) ab^n (b) ar^n (c) ab^{n-1} (d) ar^{n-1}
- For a G. P. 1, $\sqrt{3}$, 3, $3\sqrt{3}$, ..., find the 5th term ?
(a) 9 (b) $9\sqrt{3}$ (c) 27 (d) $\frac{(\sqrt{3})^5 - 1}{(\sqrt{3} - 1)}$
- For a G. P., $T_1 = a$ and $T_5 = \frac{1}{a}$, where $a > 0$ then obtain its third term.
(a) a^2 (b) 1 (c) $\frac{1}{a^2}$ (d) a
- For a G. P. $\frac{1}{9}$, $\frac{1}{3}$, 1, ..., find the seventh term.
(a) 6561 (b) 243 (c) 81 (d) $\frac{1}{81}$
- Find the common ratio of a G.P. whose n th term is $3(2^{n-1})$.
(a) 3 (b) 2 (c) 6 (d) 1
- For a G. P. 0.4, 0.04, 0.004, ..., find the common ratio.
(a) 10 (b) 0.4 (c) 4 (d) 0.1
- If x , 10, -25 are in G.P. then find the value of x .
(a) 4 (b) -25 (c) -4 (d) 2
- The common ratio of a G.P. is -1 and its first term is -1 then find the sum of the first six terms of the G.P.
(a) 0 (b) -1 (c) 1 (d) 6
- The common ratio of a G.P. is 1 and $S_{10} = 40$ then find the first term.
(a) 0 (b) 10 (c) 4 (d) 400

Section B

Give answer in one sentence for the following questions :

- What is the n th term of the G.P. ar , ar^2 , ar^3 , ...?
- Find the common ratio of the G.P. 0.1, 0.01, 0.001, ...
- Find the sum of twenty terms of the G.P. 7, 7, 7, ...
- If in a G.P. the n th term is given as $T_n = 2^{n+1}$ find the common ratio.
- The numbers 4, 1, y are in G.P. Find the value of y .
- For a G.P., sum of any two consecutive terms is zero then what will be the common ratio?
- For a G.P., if $S_7 = 15$ and $S_6 = 11$ then find the seventh term of the G.P.
- State whether the statement "if a , b , c , d are in G.P. then $ad = bc$ " is true or false.
- State whether the statement " $T_1 = S_1$ " is true or false in G.P.

Section C

Answer the following questions :

1. Define Geometric progression.
2. Define Geometric series.
3. If in a G.P., common ratio is 1 and $S_8 = 24$, find the first term of the G.P.
4. For a G.P., $T_1 = 2$ and the product of the first three terms is 1000. Find the common ratio.
5. For a G.P., $a = 2$ and $r = 3$ then find the sum of first four terms.
6. Which term of the G.P. 4, 12, 36, ... is 324 ?
7. For a G.P., $a = \frac{4}{9}$ and $r = \frac{-3}{2}$. Find T_3 .
8. If the common ratio of a geometric progression is 2, find the ratio of its 7th and 3rd terms.
9. Find the required term of the following sequence using sequence formula :
(1) 2, 10, 50, ... (6th term) (2) 100, 50, 25, ... (7th term)
(3) $\frac{1}{3}, \frac{2}{9}, \frac{4}{27}, \dots$ (8th term) (4) 2, $2\sqrt{2}$, 4, ... (5th term)

Section D

Solve the following questions :

1. For a given G.P., if $T_5 = 405$ and $T_7 = 3645$ then find T_4 .
2. Find T_5 and S_4 of the geometric progression if the first term is $\frac{27}{16}$ and common ratio is $\frac{2}{3}$.
3. For a given G.P., $a = 4$ and $T_5 = \frac{1}{4}$ then find T_7 .
4. For a given G.P., if $T_2 = 9$ and $T_5 = 243$ then find S_4 .
5. The first term of a G.P. is 10 and $T_4 = 0.08$. Find sum of first three terms.
6. For a geometric progression, $T_1^2 = T_2$ and $T_3 = 64$. Write the sequence.
7. If 5, m , 20, t are in geometric progression find m and t .
8. For a given G.P., if $a = 10$, $r = 0.1$ and $T_n = 0.01$ then find n .
9. For a given G.P., if $a = 1$, $r = 3$ and $S_n = 121$ then find n .
10. If $S_n = \frac{2}{3}(2^n - 1)$, find T_4 .
11. If $S_n = 4(3^n - 1)$, find T_{n+1} .
12. Find the product of the first three terms of a G.P. whose second term is 5.
13. How many terms of a geometric progression 2, 4, 8, 16, ... would add to 126 ?
14. If for a G.P., $T_n = 324$, $S_n = 484$ and $r = 3$, find a and n .
15. Find the sum of required terms for the following sequence using series formula :
(1) 4, 16, 64, ... (first 4 terms) (3) 100, 20, 4, ... (first 5 terms)
(2) 2, 3, $\frac{9}{2}$, ... (first 5 terms) (4) $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$ (first 10 terms)

Section E

Solve the following :

1. If the three positive numbers $k + 4$, $4k - 2$ and $7k + 1$ are in G.P., find k .
2. Find the maximum value of n such that the sum of the first n terms of a G.P. $1, 3, 3^2, 3^3, \dots$ does not exceed 365.
3. Find the minimum value of n such that the sum of the first n terms of a G.P. $1, 2, 2^2, 2^3, \dots$ is greater than or equal to 2000.
4. The sum of the first five terms of the G.P. $y, \frac{y}{3}, \frac{y}{9}, \dots$ (where $y > 0$) is 121. Find y .
5. For a geometric progression, $S_4 = 10 S_2$. Find the common ratio.
6. For a geometric progression, the ratio of sum of the fifth and the third term to the difference of the fifth and the third term is 5:3. Find r .
7. The sum and the product of the three consecutive numbers in a geometric progression are 31 and 125 respectively. Find the three numbers of the G.P.
8. The sum and the product of the three consecutive terms of G.P. are 6 and -64 respectively. Find the three terms of the G.P.
9. A construction company offers a scheme on a flat to attract customers. In this scheme, customer has to pay ₹ 10,000 as the first installment and has to pay double the amount of the preceding installment in the subsequent annual installments. What is the total amount that the customer has to pay upto 10 installments ?
10. A banker counts 128 notes in the first minute and there after he counts half the number of notes he counted in the previous minute. How many notes he would count in five minutes?
11. Population of a village is 5000. Population increases at the rate of 2% every year. What will be the population of the village after 10 years ?
12. A car depreciates at the rate of 10% every year. If the cost price of the car is ₹ 5,00,000, what will be the value of the car after 6 years ?



Aryabhata
(476 – 550)

Aryabhata was a famous Indian mathematician and astronomer. His notable contributions to the world of science and mathematics includes the theory that the earth rotates on its axis, explanations of the solar and lunar eclipses, solving of quadratic equations, place value system with zero, and approximation of pie (π). Aryabhata had defined sine, cosine, and inverse sine back in his era, influencing the birth of trigonometry. His calendar calculation has been in continuous use in India, on which the present day Panchangam is based. His studies are also base for the national calendars of Iran and Afghanistan today.

The ISRO (Indian Space Research Organization) named its first satellite after the genius mathematician and astronomer.

Answers

Exercise 1

Section A

1. (c) 2. (b) 3. (a) 4. (d) 5. (a)
 6. (b) 7. (c) 8. (a) 9. (b) 10. (c)

Exercise 2.1

1. A discrete distribution of 'number of children' in 50 families

Number of children (x)	0	1	2	3	Total
Number of families (f)	6	16	21	7	50

2. The exclusive continuous frequency distribution showing ages (in full years) for 60 employees

Age (years)	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	Total
No. of employees	3	8	10	11	12	7	6	3	60

3. An inclusive continuous frequency distribution of 'number of mobile phones produced' in a factory during 60 days

Number of mobile phones	100 - 199	200 - 299	300 - 399	400 - 499	500 - 599	600 - 699	700 - 799	800 - 899	900 - 999	1000 - 1099	Total
Number of days	2	4	9	7	10	8	9	4	3	4	60

'less than' type cumulative frequency distribution

Number of mobile phones	99.5	199.5	299.5	399.5	499.5	599.5	699.5	799.5	899.5	999.5	1099.5
Number of days	0	2	6	15	22	32	40	49	53	56	60

'more than' type cumulative frequency distribution

Number of mobile phones more than or equal to	99.5	199.5	299.5	399.5	499.5	599.5	699.5	799.5	899.5	999.5	1099.5
Cumulative Frequency	60	58	54	45	38	28	20	11	7	4	0

4.

Class	0 - 99	100 - 299	300 - 499	500 - 749	750 - 899	900 - 999
Mid value	49.5	199.5	399.5	624.5	824.5	949.5
Class length	100	200	200	250	150	100
Frequency	10	12	14	16	8	10

5. **'less than' type cumulative frequency distribution**

x or less errors	0	1	2	3
Cumulative Frequency	140	250	370	400

'more than' type cumulative frequency distribution

x or more errors	0	1	2	3
Cumulative Frequency	400	260	150	30

6. **Inclusive continuous frequency distribution**

Class	45 - 49	50 - 54	55 - 59	60 - 64	65 - 69	70 - 74	75 - 79	Total
Frequency	30	80	100	50	150	80	10	500

7. **Exclusive continuous frequency distribution**

Class	30 - 35	35 - 40	40 - 45	45 - 50	50 - 55	55 - 60	60 - 65	65 - 70	Total
Frequency	17	8	15	8	6	3	2	1	60

8.

Class	0 - 50	50 - 160	160 - 300	300 - 500	500 - 800	800 - 1000	Total
Frequency	10	30	40	60	80	30	250

9.

Class	7 - 16	17 - 26	27 - 36	37 - 46	47 - 56	Total
Frequency	160	120	43	40	2	365

10.

Class	0.9875 - 1.4875	1.4875 - 1.9875	1.9875 - 2.4875	2.4875 - 2.9875	2.9875 - 3.4875	3.4875 - 3.9875	Total
Frequency	5	10	20	20	10	5	70

Exercise 2.2

1. Classification of college students according to their gender and year of study

Year of study	Gender		Total
	Boys	Girls	
First	330	220	550
Second	225	225	450
Third	300	100	400
Total	855	545	1400

2. C

Marital status	Gender		Total
	Men	Women	
Married	715	485	1200
Unmarried	205	195	400
Total	920	680	1600

3. A table showing designation, gender and marital status of the applicants of job at bank

Designation	Married			Unmarried			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Manager									
Clerk									
Cashier									
Peon									
Total									

4. Classification of number of women according to their work experience, residence area and marital status

Residence area	Married			Unmarried			Total		
	Experienced	Inexperienced	Total	Experienced	Inexperienced	Total	Experienced	Inexperienced	Total
Labour Area	250	93	343	163	43	206	413	136	549
Other Area	87	400	487	14	800	814	101	1200	1301
Total	337	493	830	177	843	1020	514	1336	1850

5. A table showing number of skilled and unskilled workers in a company during year 2011 to 2014

Year	Skilled			Unskilled			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
2011	1170	80	1250	260	140	400	1430	220	1650
2012	1300	175	1475	200	50	250	1500	225	1725
2013	1460	240	1700	40	10	50	1500	250	1750
2014	1670	290	1960	30	10	40	1700	300	2000

Exercise 2

Section A

1. (d) 2. (d) 3. (b) 4. (a) 5. (c)
 6. (a) 7. (b) 8. (a) 9. (c) 10. (d)
 11. (c) 12. (a) 13. (c) 14. (b) 15. (b)

Section C

7. 4.5, 17, 37, 62, 87.5 8. 10, 15, 25, 25, 26

9. 'less than' discrete cumulative frequency distribution

Observation or less than that	10	20	30	40	50
Cumulative Frequency	10	40	70	90	100

10. A table showing demand of an item during a year

Demand	Good	Moderate	Weak	Total
Number of weeks	12	22	18	52

- 11.

Year	Attribute A			Attribute B			Total		
	Sub data-1	Sub data-2	Total	Sub data-1	Sub data-2	Total	Sub data-1	Sub data-2	Total
2014	200	100	300	100	100	200	300	200	500
2015	150	400	550	150	300	450	300	700	1000

Section D

- 8.

Class	200 - 300	300 - 400	400 - 500	500 - 600	600 - 700	700 - 800	Total
Frequency	20	80	80	40	60	20	300

9. A table showing marital status of 40 employees

Gender	Married	Unmarried	Total
Male	8	8	16
Female	15	9	24
Total	23	17	40

10. An inclusive continuous frequency distribution of 'monthly income' of workers

Monthly Income (₹)	2400 - 2900	2900 - 3400	3400 - 3900	3900 - 4400	4400 - 4900	4900 - 5400	5400 - 5900	5900 - 6400	Total
No. of workers	3	9	18	25	23	10	7	5	100

11. An inclusive continuous frequency distribution of 'marks' of 200 students

Marks	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100	Total
No. of students	20	40	50	35	25	22	6	1	1	200

Class	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50
Frequency	12	18	16	22	14	10	6	2

13. A table showing use of buses as public transport in Ahmedabad city

Type of transport	Types of bus		Total
	Air-conditioned	Non Air-conditioned	
BRTS	250	100	350
AMTS	150	500	650
Total	400	600	1000

14. Classification of college students according to their gender and stream.

Stream	Gender		Total
	Boys	Girls	
Science	250	350	600
Commerce	650	250	900
Total	900	600	1500

19. Multiple bar diagram.

21. Circle diagram.

Section E

1. An inclusive continuous distribution of 'number of mangoes' on different mango trees during 30 days

No. of mangoes	90 - 94	95 - 99	100 - 104	105 - 109	110 - 114	115 - 119	120 - 124	125 - 129	Total
No. of days	2	3	6	4	4	4	4	3	30

2. An inclusive continuous distribution of 'daily income' of 40 rickshaw drivers of a city

Daily income (₹)	200 - 219	220 - 239	240 - 259	260 - 279	280 - 299	300 - 319	320 - 339	340 - 359	Total
No. of rickshaw drivers	4	6	4	6	4	5	6	5	40

3. An exclusive continuous distribution of 'water consumption' of 50 residence of an area

Water consumption	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 55	55 - 60	Total
Number of houses	2	10	9	7	10	5	3	4	50

4. An exclusive continuous distribution of 'price' of 50 shops

Price of an item	60 - 65	65 - 70	70 - 75	75 - 80	80 - 85	85 - 90	Total
No. of shops	5	11	11	11	7	5	50

5. 'less than' type cumulative frequency distribution

Upper boundary point	24.5	29.5	34.5	39.5	44.5	49.5	54.5	59.5
Cumulative Frequency	0	3	11	21	26	41	49	50

'more than' type cumulative frequency distribution

Lower boundary point and above	24.5	29.5	34.5	39.5	44.5	49.5	54.5	59.5
Cumulative Frequency	50	47	39	29	24	9	1	0

6. A discrete frequency distribution of 'absence of workers' in a factory during 30 days.

No. of absent workers	0	1	2	3	4	5	6	Total
No. of days	5	7	5	6	4	2	1	30

'less than' type cumulative frequency distribution

Number of absent workers or less than that	0	1	2	3	4	5	6
Number of days	5	12	17	23	27	29	30

7. A table showing classification of 850 students of a school according to their gender and class

Class	Gender		Total
	Boys	Girls	
10	255	145	400
11	125	125	250
12	150	50	200
Total	530	320	850

8. A classification of students, according to their gender and residence from the year 2013 to 2015

Year	Residing in hostel			Not residing in hostel			Total		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
2013	600	350	950	200	50	250	800	400	1200
2014	700	420	1120	260	100	360	960	520	1480
2015	840	520	1360	260	100	360	1100	620	1720

9. Classification of applicants according to their qualification, gender and marital status

Education	Male			Female			Total		
	Married	Unmarried	Total	Married	Unmarried	Total	Married	Unmarried	Total
Graduate	150	450	600	160	240	400	310	690	1000
Post graduate	192	288	480	160	160	320	352	448	800
Other professional	70	70	140	36	24	60	106	94	200
Total	412	808	1220	356	424	780	768	1232	2000

10. (1) increase of 50%
 (2) 20%
 (3) Male increase by 53.33% and female by 40%

Section F

1. An exclusive continuous frequency distribution of 'thickness of lenses'

Thickness of lenses	1.505 - 1.510	1.510 - 1.515	1.515 - 1.520	1.520 - 1.525	1.525 - 1.530	Total
No. of lenses	5	3	6	7	4	25

Percentage of defective lenses = 36 %

2. An exclusive continuous frequency distribution of 'variation in a price of a share' in stock market during 30 days

Price of share (₹)	10.5 - 12.5	12.5 - 14.5	14.5 - 16.5	16.5 - 18.5	18.5 - 20.5	20.5 - 22.5	Total
No. of days	2	6	8	4	8	2	30

- (i) ₹ 17.5 (ii) 16 days (iii) 6 days

3. An exclusive continuous frequency distribution of 'deviation in production of mixers' during 40 days

Deviation in number of mixers produced	-12 to -6	-6 to 0	0 to 6	6 to 12	12 to 18	18 to 24	Total
No. of Days	2	5	12	10	6	5	40

'less than' type cumulative frequency distribution.

Upper boundary point	-12	-6	0	6	12	18	24
Cumulative Frequency	0	2	7	19	29	35	40

'more than' type cumulative frequency distribution.

Lower boundary point	-12	-6	0	6	12	18	24
Cumulative Frequency	40	38	33	21	11	5	0

4. An inclusive continuous frequency distribution of 'height' of 30 students.

Height (cm)	140 - 144	145 - 149	150 - 154	155 - 159	160 - 164	165 - 169	Total
No. of students	2	8	8	4	6	2	30

'less than' type cumulative frequency distribution.

Upper boundary point	139.5	144.5	149.5	154.5	159.5	164.5	169.5
Cumulative Frequency	0	2	10	18	22	28	30

'more than' type cumulative frequency distribution

Lower boundary point	139.5	144.5	149.5	154.5	159.5	164.5	169.5
Cumulative Frequency	30	28	20	12	8	2	0

(i) 8 students (ii) 12 students (iii) 149 cm.

5. A table showing classification of students according to their stream and gender

Stream	Boys	Girls	Total
Engineering	7750	3150	10,900
Doctor	6000	4000	10,000
Science	7000	1000	8000
Arts	2800	6800	9600
Commerce	450	1050	1500
Total	24,000	16,000	40,000



Exercise 3.1

- (1) Mean = 1.78 cm
 (2) corrected mean = 24.5 years, can participate
 (3) Mean = 41 mm
 (4) Mean = 35.36 marks
 (5) Mean = 7.49 min
 (6) Mean = ₹ 18.76 lakh
 (7) Mean = 40 units

Exercise 3.2

- (1) Combined mean = ₹ 198.75
(2) Weighted mean = 118.09 percent change
(3) Weighted mean = ₹ 506.67
(4) mean = 82 marks

Exercise 3.3

- (1) Geometric mean = 2.61 books
(2) Average depreciation = 6.05% (Geometric mean)
(3) Geometric mean = 61.73 km

Exercise 3.4

- Quartiles $Q_1 = 4$ marks, $Q_2 = 6$ marks, $Q_3 = 7$ marks
- Median = 229.17 km. The travelling on 50 % days will be 229.17 km or less.
 $Q_3 = 291.67$ km. The maximum distance in the least travelled 75 % days will be 291.67 km.
 $D_8 = 314.29$ km. The maximum distance in the least travelled 80 % days will be 314.29 km.
 $P_{62} = 259.17$ km. The maximum distance in the least travelled 62 % days will be 259.17 km.
- Median = 19 years. Age of 50 % students will be 19 years or less.
 $Q_1 = 18$ years, Age of 25 % students will be 18 years or less.
 $D_4 = 19$ years, Age of 40 % students will be 19 years or less.
 $P_{32} = 18.92$ years, Age of 32 % students will be 18.92 years or less.
- Median = ₹ 21.78 thousand, Lower limit of the richest 20 % employees is ₹ 26.84 thousand
- Median = ₹ 495
- Median from raw data = 4 days
Median from grouped data = 4.17 days
Both the values are almost same.

Exercise 3.5

- Mode = 138
- Mode = 13 cakes
- For empirical formula, Mean = 68.85 years, Median = 67.11 years and hence Mode = 63.63 years
- Mode cannot be found using its definition but it can be found using empirical formula.
- Mode using formula = 152.35 gm
Mode using graph = 153 gm
- Mode = ₹ 22 thousand

Exercise 3

Section A

- | | | | | |
|---------|---------|---------|---------|---------|
| 1. (a) | 2. (d) | 3. (b) | 4. (c) | 5. (a) |
| 6. (b) | 7. (c) | 8. (d) | 9. (c) | 10. (b) |
| 11. (a) | 12. (b) | 13. (c) | 14. (b) | 15. (d) |

Section B

- | | | |
|-----------------------|--------------------------|-----------------|
| 2. Weighted mean | 4. $M_o = 3M - 2\bar{x}$ | 7. Karl Pearson |
| 8. Median = 55 | 9. mean = 13 | 10. Mode = 10 |
| 11. Second number = 4 | 12. $Q_1 = 4$ | 13. Median |
| 14. $P_{75} = 25.75$ | 15. Median = 150 | |

Section C

7. Median = 5.8 8. Second number = 16 9. mean = 293
10. $a = 10$ 11. Combined mean = 81 marks 12. Median = 2 vehicles
13. Weighted mean = 1090

Section D

7. Average growth rate = 2.87 % Geometric mean
8. $D_7 = 8$ phones. Sale of phones will be 8 or less on 70% days.
 $P_{15} = 6$ phone. Sale of phones will be 6 or less on 15% days.
9. mean = 30.07 ml. Machine is working properly.
10. mean = 61.62 marks
11. New mean = 34.69
12. Median = 54 marks
13. mean = 138.9 units. Sales have increased after advertisement.

Section E

- | | |
|---|---|
| 1. Median = 362.5 unit | 2. Mode = ₹ 2.86 thousand |
| 3. $Q_1 = 34.21$ marks, $D_4 = 36.69$ marks | 4. mean = 164.97 cm |
| 5. Median = ₹ 15.2 thousand | 6. Mode = 23 thousand |
| 7. mean = 24.46 marks | 8. $Q_1 = 6.66$ hours, $Q_3 = 7.64$ hours |

Section F

1. (i) $D_3 = 25$, Maximum marks among failing students is 25 hence 26 marks would be required for passing.
(ii) $P_{95} = 60.83$, Minimum marks among the highest scoring 5% students is 61.
2. Mean for A = 22.33 km.
Mean for B = 23.5 km.
 \therefore Brand B tyres are better.
3. For empirical formula, Mean = 16.71 cars, Median = 16.72 cars
Hence Mode = 16.74 cars
4. Mean = 35.93 quintals
Median = 35.11 quintals
5. Mode = 34 years
6. Mode = 72.5 units. Value of mode has increased.
7. Median for $x = 13.26$ tins
Median for $y = 10.7$ tins
Company x has higher sale.
8. Mode = 25.5 years



Exercise 4.1

1. Range = 40 cm, Coefficient of range = 0.12
2. Range = 35, Coefficient of range = 0.90
3. Range = 60 marks, Coefficient of range = 0.6
4. Range = ₹ 29 thousand, Coefficient of range = 0.74

Exercise 4.2

1. $Q_d = 7.88$ mm, Coefficient of quartile deviation = 0.29
2. $Q_d = 10$ marks, Coefficient of quartile deviation = 0.33
3. $Q_d = ₹ 38.54$, Coefficient of quartile deviation = 0.32

Exercise 4.3

1. Mean Deviation = 5 cm
2. Mean Deviation = 2.8 bearings, Coefficient of mean deviation = 0.35
3. Mean Deviation = 3.33 minutes, Coefficient of mean deviation = 0.46
4. Mean Deviation = 15 TV, Coefficient of quartile deviation = 0.25
5. Mean Deviation = 13.18 boxes

Exercise 4.4

1. $s = 2.67$ marks
2. $s = 2.65$ cars
3. $s = 6.71$ units, Coefficient of standard deviation = 0.35
4. $s = 12.89$ (Lakh ₹)
5. $s = 19.76$ years, Coefficient of standard deviation = 0.56

Exercise 4.5

1. For share A : $\bar{x} = ₹ 321$, $s = ₹ 2.65$ Coefficient of Variation = 0.83 %
For share B : $\bar{x} = ₹ 140$, $s = ₹ 7.14$ Coefficient of Variation = 5.1 %. Price of share B has more variation
2. Coefficient of variation of company A and B are 5 % and 4 %. Company B is more stable
3. The means of two series are 50 and 36 respectively.

Exercise 4.6

1. $\bar{x}_c = 53.45$ marks, $s_c = 12.64$ marks
2. $\bar{x}_c = 21$ min, $s_c = 5.22$ min

Exercise 4

Section A

1. (b)
2. (a)
3. (d)
4. (c)
5. (a)
6. (c)
7. (a)
8. (c)
9. (b)
10. (c)
11. (a)
12. (a)

Section B

3. Relative Measures
4. Standard Deviation
5. (Centimeter)²
6. Range = 100 cm
7. $Q_d = 15.91$
8. $s = 0$
9. Mean Deviation = 2

Section C

4. Mean Deviation and Standard Deviation
5. Range = 14, Coefficient of range = 1.75
6. $Q_d = 6$, Coefficient of quartile deviation = 0.67
7. Mean Deviation = 2.4
8. Variance = 25
9. $s = 1.41$
10. Coefficient of Variation of A = 20 %, Coefficient of Variation of B = 25 %. Factory A is stable with respect to production.
11. Coefficient of quartile deviation = 0.29

Section D

9. $Q_d = 3$ flowers
10. Mean Deviation = 0.75 Goals
11. $\bar{x} = 4.25$, $s = 1.63$, Coefficient of Variation = 38.35 %
12. $s_c = 7.43$
13. $\bar{x} = 8$, $s = 4$, Coefficient of Variation = 50 %

Section E

1. $\bar{x} = 25.17$ marks, Mean Deviation = 3.81 marks
2. $Q_1 = ₹ 16.5$ thousand, $Q_3 = ₹ 43.75$ thousand, $Q_d = ₹ 13.63$ thousand
3. $s = 15.94$ runs
4. $Q_1 = 14.5$ marks, $Q_3 = 34.5$ marks, $Q_d = 10$ marks
5. For team A : $\bar{x} = 1.45$ goals, $s = 1.48$ goals,
Coefficient of Variation = 102.07 %
For team B : $\bar{x} = 1.07$ goals, $s = 1.33$ goals
Coefficient of Variation = 124.3 %
Team A is more consistent
6. Corrected Mean = 39.3
Corrected standard deviation = 10.24
7. For total cost y : Range = 150, Quartile deviation = 15, Mean deviation = 24 and
Standard deviation = 30

Section F

1. Range = 32 visits, Coefficient of Variation = 0.84
Quartile deviation = 6 visits, Coefficient of quartile deviation = 0.33
Mean deviation = 5.91 visits, Coefficient of mean deviation = 0.33
2. $\bar{x} = 15.54$ days, $s = 1.45$ days, $\bar{x} \pm s = 14.09$ days to 16.99 days, 55 %
3. Q_d is an appropriate measure, $Q_1 = 17.5$ marks, $Q_3 = 29$ marks, $Q_d = 5.75$ marks,
Coefficient of quartile deviation = 0.25
4. $s = ₹ 14.84$ thousand
5. $\bar{x} = ₹ 42.6$, Mean deviation = ₹ 14.99
6. $\bar{x} = ₹ 404.35$, $s = ₹ 172.58$, Coefficient of Variation = 42.68 %
7. For student A : $\bar{x} = 62$ marks, $s = 11.49$ marks
Coefficient of Variation = 18.53 %
For student B : $\bar{x} = 60.5$ marks, $s = 8.62$ marks
Coefficient of Variation = 14.25 %
Student B is more consistent.
8. For group A : $\bar{x} = 46.29$ kg, $s = 11.57$ kg
Coefficient of Variation = 25 %
For group B : $\bar{x} = 46.43$ kg, $s = 10.93$ kg
Coefficient of Variation = 23.54 %
Group A has greater relative variation.

Exercise 5.1

1. $\bar{x} = 4.27$ packets of milk; $M_o = 4$ packets of milk; $s = 1.65$ packets of milk; $j = 0.16$
2. $\bar{x} = 14.01$ inches; $M = 14$ inches; $s = 0.87$ inches, $j = 0.03$, Positive skewness
3. $\bar{x} = 14.22$ min.; $M_o = 12.28$ min; $s = 5.33$ min.; $j = 0.36$, Positive skewness
4. $\bar{x} = ₹ 9.92$ lakh; $M_o = ₹ 9.8$ lakh; $s = ₹ 2.37$ lakh; $j = 0.05$, Positive skewness
5. $\bar{x} = ₹ 20.12$ lakh; $M = ₹ 21.5$ lakh; $s = ₹ 7.98$ lakh; $S_k = ₹ -4.14$ lakh, $j = -0.52$, Negative skewness
6. $\bar{x} = 10.31$ thousand bales; $M = 9.86$ thousand bales; $S_k = 1.35$ thousand bales, $s = 6.33$ thousand bales; $j = 0.21$, Positive skewness
7. $\bar{x} = 9.5$ Celsius; $M = 8.6$ Celsius; $s = 7.27$ Celsius; $S_k = ₹ 2.7$ Celsius, $j = 0.37$, Positive skewness

Exercise 5.2

1. $Q_1 = 20$ years; $M = 22$ years; $Q_3 = 25$ years; $j = 0.2$, Positive skewness
2. $Q_1 = ₹ 335$ lakh; $M = ₹ 490$ lakh; $Q_3 = ₹ 912.5$ lakh; $S_k = ₹ 267.5$ lakh; $j = 0.46$, Positive skewness
3. $Q_1 = 40$ thousand tonnes; $M = 48$ thousand tonnes; $Q_3 = 68.75$ thousand tonnes; $S_k = 12.75$ thousand tonnes, $j = 0.44$, Positive skewness
4. $Q_1 = ₹ 18.27$ thousand; $M = ₹ 20.53$ thousand; $Q_3 = ₹ 22.44$ thousand; $j = -0.08$

Exercise 5

Section A

- | | | | | | | |
|--------|--------|---------|---------|---------|---------|---------|
| 1. (c) | 2. (b) | 3. (c) | 4. (c) | 5. (a) | 6. (d) | 7. (c) |
| 8. (d) | 9. (a) | 10. (b) | 11. (c) | 12. (a) | 13. (d) | 14. (b) |

Section B

- | | |
|----------------------------|----------------------------|
| 13. Negative skewness | 14. Negative skewness |
| 15. Negative skewness | 16. Symmetric distribution |
| 17. Symmetric distribution | |

Section C

- | | | |
|------------------------|-------------------------|-----------------|
| 5. $\bar{x} = 46$ | 6. $M = 69$ | 7. $M = 32.50$ |
| 8. $s = 12$ | 9. $j = 0.33$ | 10. $j = -0.15$ |
| 11. $M = 42, j = -0.4$ | 13. $j = -0.24$ | |
| 14. $s = 4, s^2 = 16$ | 15. $M_o = 38; j = 0.5$ | |

Section D

5. coefficient of skewness for group A : $j = -0.17$; coefficient of skewness for group B : $j = -0.40$
Group A is closer to symmetry.
6. For group A, $Q_1 = 36$; $M = 48$; $Q_3 = 72$; j for group A = 0.33; j for group B = 0.39; Group B is more skewed than group A.
7. $S_k = 1.6$; $j = 0.07$
8. $s = 8$; $j = -0.025$
9. $\bar{x} = 32$; $s = 4$, $j = -0.15$
10. $M_o = 66$; $M = 62$
11. $s = 12$; $M_o = 56$; $M = 61.33$; C.V. = 18.75
12. $j = -0.75$
13. $\bar{x} = 36$; $M_o = 24$; $M = 32$; $j = 0.19$ by Karl Pearson's method; $j = -0.4$ by Bowley's method

Section E

6. Firm A : Coefficient of skewness : Karl Pearson's method $j = 0.69$; Bowley's method $j = -0.25$
 Firm B : Coefficient of skewness : Karl Pearson's method $j = 1.58$; Bowley's method $j = 0.5$

The data for firm B has more skewness than firm A in Karl Pearson's method. Firm B is more skewed than firm A in Bowley's method.

7. $\bar{x} = 18.9$ dozen; $M = 18$ dozen; $s = 4.44$ dozen; $j = 0.61$
 8. $\bar{x} = 21.14$ staplers; $M_o = 20$ staplers; $s = 1.65$ staplers; $j = 0.69$
 9. $\bar{x} = 240$, $s = 7.5$; $j = -2.4$, Negative skewness

Section F

1. $Q_1 = 2$ hours; $M = 3$ hours; $Q_3 = 4$ hours; $j = 0$; Bowley's; $j = 0$
 2. $\bar{x} = 14.89$ Celsius; $M = 15.12$ Celsius; $s = 7.95$ Celsius; $j = -0.09$, Negative skewness
 3. $\bar{x} = 31.42$ marks; $M = 31.32$ marks; $s = 11.68$ marks; $j = 0.026$, Positive skewness
 4. $Q_1 = ₹ 17.5$ lakh; $Q_3 = ₹ 34.38$ lakh; $M = ₹ 26$ lakh; $j = -0.007$ Negative skewness
 5. $\bar{x} = 9.28$ units; $M = 8$ units; $s = 6.66$ units; $S_k = 3.84$, $j = 0.58$
 6. $Q_1 = 4.19$ mm; $Q_3 = 4.46$ mm; $M = 4.32$ mm; $j = 0.037$ Positive skewness
 7. $\bar{x} = 9.23$ packets; $M = 4.75$ packets; $s = 10.22$ packets; $j = 1.32$, Positive skewness
 8. $Q_1 = 2.95$ sq. m; $Q_3 = 5.95$ sq. m; $M = 4.55$ sq. m; $j = -0.067$; Negative skewness
 9. $\bar{x} = 180$ sq. m; $M_o = 180$ sq. m; $s = 41.63$ sq.m.; $j = 0$. Symmetric distribution
 10. $Q_1 = 23.75$ units; $Q_3 = 35.63$ units; $M = 30.5$ units; $j = -0.14$



Exercise 6.1

- | | | | |
|---------------|------------|------------|--------------|
| 1. (1) 720 | (2) 2450 | (3) 40,320 | (4) 3,62,880 |
| 2. $n = 11$ | 3. $r = 4$ | 4. $n = 7$ | 5. 24 |
| 7. 2880 | 8. 576 | 9. 72 | 10. 24 |
| 11. (1) 50400 | (2) 151200 | (3) 90720 | 12. 2:1 |
| 14. (1) 49 | (2) 12 | (3) 83 | (4) 93 |
| 15. 240 | 16. 720 | | 13. 9072 |

Exercise 6.2

- | | | | |
|----------------|------------------------|-----------|--------------|
| 1. (1) 330 | (2) 1 | (3) 300 | (4) 1 |
| 2. (1) $n = 8$ | (2) $r = 8$ or $r = 5$ | (3) | $n = 6$ |
| 3. 28 | 4. 10 | 5. 120 | 6. 2184 |
| 7. (1) 11 | (2) 15 | 8. (1) 78 | (2) 16 |
| 9. (1) 56 | (2) 20 | 10. 55 | 11. 63 |
| 12. (1) 35 | (2) 21 | 13. 127 | 14. 560, 126 |
| 15. $n = 8$ | 16. $r = 4$ | | |

Exercise 6.3

1. (1) $27a^3 + 108a^2b + 144ab^2 + 64b^3$
 (2) $1 + 7x + 21x^2 + 35x^3 + 35x^4 + 21x^5 + 7x^6 + x^7$
 (3) $\frac{81}{x^4} - \frac{144}{x^2} + 96 - \frac{256x^2}{9} + \frac{256x^4}{81}$
 (4) $\frac{x^3}{729} + \frac{2x^2}{27} + \frac{5x}{3} + 20 + \frac{135}{x} + \frac{486}{x^2} + \frac{729}{x^3}$
 (5) $\frac{a^5}{32} - \frac{5a^4b}{48} + \frac{5a^3b^2}{36} - \frac{5a^2b^3}{54} + \frac{5ab^4}{162} - \frac{b^5}{243}$
2. (1) 352 (2) 198 (3) 248

Exercise 6

Section A

- | | | | | | |
|--------|--------|--------|---------|---------|--------|
| 1. (d) | 2. (a) | 3. (b) | 4. (d) | 5. (b) | 6. (c) |
| 7. (b) | 8. (a) | 9. (d) | 10. (a) | 11. (d) | |

Section B

5. 1,6,15,20,15,6,1
- | | | | | | |
|-------|------|------|----------|--------|---------|
| 7. 60 | 8. 6 | 9. 7 | 10. 5039 | 11. 60 | 12. 120 |
|-------|------|------|----------|--------|---------|

Section C

- | | | | | | |
|--------|--------|-------------|--------|----------|-------|
| 2. 720 | 3. 144 | 4. 12441600 | 5. 96 | 6. 12 | 7. 72 |
| 8. 18 | 9. 1:1 | 10. 6 | 11. 12 | 12. 5200 | |

13. $8x^3 + 36x^2y + 54xy^2 + 27y^3$
14. $x^3 - 3x + \frac{3}{x} - \frac{1}{x^3}$
15. $y^5 + 5y^4k + 10y^3k^2 + 10y^2k^3 + 5yk^4 + k^5$

Section D

- | | | |
|--------------------------|------------------------|------------------------------------|
| 1. (1) 120 (2) 72 (3) 24 | 2. 1152 | 3. (1) 48 (2) 12 (3) 36 |
| 4. 4370 | 5. (1) 4 (2) 16 (3) 12 | 6. (1) 9 (2) 12 (3) 6 |
| 7. (1) 1 (2) 36 (3) 12 | 8. (1) 34 (2) 50 | 9. 44 10. 416 |



Exercise 7

Section A

- | | | | | | | |
|--------|--------|--------|--------|--------|--------|--------|
| 1. (d) | 2. (b) | 3. (b) | 4. (a) | 5. (c) | 6. (c) | 7. (a) |
|--------|--------|--------|--------|--------|--------|--------|

Section B

- | | | | | | |
|--------------|---------------|---------------|-----------|----------------------|--|
| 1. (1) False | (2) False | (3) True | (4) True | | |
| (5) True | (6) False | (7) True | | | |
| 2. Sampling | 3. Stratified | 4. Systematic | 5. Tippet | 6. Population survey | |

Section D

For Examples 13 to 17, random sample other than the one in the answer can be obtained.

13. 018, 096, 027, 007, 012
14. 27, 32, 59, 66, 32, 48, 25
15. With replacement : 170, 111, 002, 203, 111, 233, 300
Without replacement : 170, 111, 002, 203, 233, 300, 250
16. First years : 158, 092, 009, 200
Second years : 019, 131, 057, 006
Third years : 027, 070, 198, 200
17. Wheat producing farmers : 12, 18, 20, 11, 03, 10
Rice producing farmers : 04, 11, 08, 13
19. $N = 20, n = 4, k = N/n = 20/4 = 5$
Sample 1 : 1, 6, 11, 16 Sample 4 : 4, 9, 14, 19
Sample 2 : 2, 7, 12, 17 Sample 5 : 5, 10, 15, 20
Sample 3 : 3, 8, 13, 18
20. $N = 30, n = 10, k = N/n = 30/10 = 3$
Sample 1 : 1, 4, 7, 10, 13, 16, 19, 22, 25, 28
Sample 2 : 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
Sample 3 : 3, 6, 9, 12, 15, 18, 21, 24, 27, 30



Exercise 8

Section A

1. (a) 2. (a) 3. (b) 4. (b) 5. (c) 6. (c) 7. (b) 8. (c)

Section B

1. Domain A and co-domain B should be non-empty. 2. Yes 3. No
7. No. Domains of two functions are different 8. Many-one 9. one-one

Section C

4. $R_f = \{3, 4, 5\}$ 5. Many-one 6. One-one 7. $D_f = \{\frac{1}{2}, 1, \frac{3}{2}\}$
8. 0 9. $R_f = \{-\frac{3}{4}, 0, \frac{3}{10}\}$ 10. 27 11. Many-one
12. $x = 2$ 13. One-one 14. 14 15. 0

Section D

1. $D_f = \{10, 20, 30\}, B = \{18, 48, 98, 128, 148\}, R_f = \{48, 98, 148\}$
2. $D_f = \{-\frac{1}{2}, 1, \frac{1}{2}, \frac{3}{2}\}, B = \{-\frac{1}{5}, 1, \frac{1}{3}, 3\}, R_f = \{-\frac{1}{5}, 1, \frac{1}{3}, 3\}$

3. $f(-1) = -1, f(-2) = -\frac{5}{4}, f\left(\frac{1}{2}\right) = 5$ 4. $D_f = \{3, 4, 5, 7\}$ 5. $x = \pm \frac{1}{2}$
 6. $R_f = \{2, 5, 10, 17\}$ 7. 5 8. $\{0, 3\}$
 9. Unequal functions 10. Many-one 12. $\frac{14}{27}$
 13. 36 14. $\frac{58}{11}$ 15. 40, 1300

•

Chapter 9

Section A

1. (d) 2. (a) 3. (a) 4. (b) 5. (c)
 6. (b) 7. (d) 8. (c) 9. (a) 10. (c)

Section B

1. ar^n 2. 0.1 3. 140 4. 2 5. $\frac{1}{4}$
 6. -1 7. 4 8. True 9. True

Section C

3. $a = 3$ 4. $r = 5$ 5. 80 6. Fifth 7. 1
 8. 16 9. (1) 6250 (2) $\frac{25}{16}$ (3) $\frac{128}{6561}$ (4) 8

Section D

1. ± 135 2. $T_5 = \frac{1}{3}$ and $S_4 = \frac{65}{16}$ 3. $\frac{1}{16}$ 4. 120
 5. 12.4 6. 4, 16, 64, ... 7. $m = \pm 10, t = \pm 40$
 8. $n = 4$ 9. $n = 5$ 10. $\frac{16}{3}$ 11. $8(3^n)$
 12. 125 13. 6 14. $a = 4$ and $n = 5$
 15. (1) 340 (2) $\frac{211}{8}$ (3) 124.96 (4) $\frac{1023}{1024}$

Section E

1. $k = 5$ 2. $n = 6$ 3. $n = 11$
 4. 81 5. $r = \pm 3$ 6. $r = \pm 2$
 7. 1, 5, 25 or 25, 5, 1 8. 2, -4, 8 or 8, -4, 2 9. $S_{10} = ₹ 1,02,30,000$
 10. 248 notes 11. 6095 12. ₹ 2,65,720.50

• • •