ગુજરાત રાજ્યના શિક્ષણવિભાગના પત્ર-ક્રમાંક
મશબ/1215/178/છ, તા. 24-11-2016-થી મંજૂર

STATISTICS

(Part 1)

Standard 12

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.

રાજ્ય સરકારની વિનામૂલ્યે યોજના હેઠળનું પુસ્તક

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

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
Shri Yatin A. Parikh

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

Reviewers
Shri Lalita K. Patel
Shri Haresh N. Doshi
Shri Girish C. Naik
Smt. Aparna S. Pathak
Shri Chirag S. Shah
Smt. Deepa K. Lalwani
Shri C. P. Rajpurohit
Shri Manoj G. Naidu

Language Reviewer
Dr. Chetan Mewada

Artist
Media Graphics

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

Preparation and Planning
Dr. Kamlesh N. Parmar
(Dy. Director : Academic)

Lay-out and Planning
Shri Haresh S. Limbachiya
(Dy. Director : Production)

PREFACE

Gujarat State Board of School Textbooks has prepared new textbooks as per the new curricula developed by the Gujarat State Secondary and Higher Secondary Education Board and which has been sanctioned by the Education Department of the Government of Gujarat. A panel of experts from Universities/Colleges, Teachers Training Colleges and Schools have put lot of efforts in preparing the manuscript of the subject. It is then reviewed by another panel of experts to suggest changes and filter out the mistakes, if any. The suggestions of the reviewers are considered thoroughly and necessary changes are made in the manuscript. Thus, the Textbook Board takes sufficient care in preparing an error-free manuscript. The Board is vigilant even while printing the textbooks.

The Board expresses the pleasure to publish the Textbook of **Statistics (Part 1)** for **Std. 12** which is a translated version of Gujarati. The Textbook Board is thankful to all those who have helped in preparing this textbook. However, we welcome suggestions to enhance the quality of the textbook.

H. N. Chavda	**Dr. Nitin Pethani**
Director	Executive President
Date : 15-02-2017	Gandhinagar

First Edition : 2017

Published by : H. N. Chavda, Director, on behalf of Gujarat State Board of School Textbooks, 'Vidyayan', 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;

(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.

*Constitution of India : Section 51-A

CONTENTS

•

1

Index Number

Contents :

1.1 Definition and Meaning of Index Number

Price of an item, national income, supply, production, employment, unemployment, investment, import-export, cost of living, population of a country, birth rate and death rate vary continuously with time. Generally, the proportion and direction of these variations also keep changing. It is important to study the variations in the price and quantity of an item with respect to change in time. The planning for the future can be suitably done from the knowledge of these changes. The changes taking place in the values of the variable at two different time periods can be measured by the following methods :

(1) Method of absolute measure (difference) and (2) Method of relative measure (ratio)

We will understand this concept by the following illustration :

Suppose the data regarding the average price per kilogram of two items, wheat and rice, for a month in the year 2015 and year 2016 are as follows :

Item	**Price per kilogram ₹**	
	Year 2015	**Year 2016**
Wheat	24	30
Rice	40	46

Let us understand the comparision of variations in the prices of wheat and rice using the two methods stated above.

(1) Method of Absolute Measure (difference) : The price of wheat in the year 2015 was ₹ 24 which increased to ₹ 30 in the year 2016. Thus, the price per kilogram increased in the year 2016 by ₹ 6 with respect to the year 2015. Similarly, the price of rice has also increased by ₹ 6. This is obtained by the absolute difference. Thus, it can be said that there is same rise in price in both the items. But this is not true in reality because the prices per unit of these items are not same in the year 2015. Thus, the base for comparative study of prices in 2016 is different. Hence, this method is not appropriate to compare the variations in a variable. We shall now study the method of relative measure which is used in such situations.

(2) Method of Relative Measure (ratio) : A ratio of price of the commodity in the year 2016 is obtained with the price in the year 2015 in this method to find the relative changes of the price of the item in the year 2016.

Thus, ratio of prices of wheat $= \frac{30}{24} = 1.25$

ratio of prices of rice $= \frac{46}{40} = 1.15$

It can be known from these ratios that the relative increase in the prices of wheat and rice in the year 2016 is not same. The price of wheat in the year 2016 is 1.25 times the price in the year 2015, whereas the price of rice is 1.15 times its price in the year 2015. Thus, it can be said that change in the price of wheat is more than the change in the price of rice.

The ratios of prices of wheat and rice given here indicate the changes in prices at two different time periods. It is also called as relative change or price relative. Generally, the ratios are expressed as percentages to facilitate comparison. Hence,

Percentage change in the price of wheat $= 1.25 \times 100 = 125$ and

Percentage change in the price of rice $= 1.15 \times 100 = 115$

This is the relative percentage measure for the changes. Such a relative measure is called index number.

Thus, the percentage change in the value of a variable associated with any item for the given (current) period compared to its value in a fixed (base) period is called an index number.

Now, we shall obtain a relative measure for the collective change in the prices of these two mutually related items. The absolute method is not useful to find a measure for the overall change because many times the units expressing the prices of these two items may be different and it is not possible to combine the changes in these prices. The method of relative measure is used in such a situation. Since the relative measure is free from the unit of measurement, it is possible to combine the changes in the prices of the two items and it is convenient to find a mathematical measure for these changes. Now, we shall take a relative measure for the overall change from the changes in prices of grains, wheat and rice. We shall denote the price for the year 2016 as p_1 and the price of the year 2015 by p_0. p_0 is called base year price and p_1 is called the current year price. The ratio $\frac{p_1}{p_0}$ is called the price relative of that item.

We shall present this in a tabular form :

Item	Price of base year 2015 (₹) p_0	Price of current year 2016 (₹) p_1	Price relative or Relative change $= \frac{p_1}{p_0}$	Percentage of Price relative $= \frac{p_1}{p_0} \times 100$
Wheat	24	30	$\frac{30}{24} = 1.25$	125
Rice	40	46	$\frac{46}{40} = 1.15$	115
Total			**2.40**	**240**

The index obtained by multiplying the average of price relatives of the current year for these two items by 100 is called the price index number of the items for the current year. It is denoted by *I*. Thus,

Price index number of wheat and rice for the current year

$$= \frac{\text{Price relative of wheat} + \text{Price relative of rice}}{\text{No. of items}} \times 100$$

$$= \frac{1.25 + 1.15}{2} \times 100$$

$$= 120$$

Hence, price index number of wheat and rice for the current year $I = 120$. The price index number of wheat and rice $I = 120$ indicates that there is an overall rise of 20 percent in the prices of the two items in the year 2016 as compared to the year 2015. Index number is a relative measure based on ratio. Similarly, measure for the overall change can be obtained using the relative method by combining changes in the values of more than two variables. We can define the general index number for a group as follows.

"The average of the percentage change in the value of a variable associated with one or more items for the given (current) period compared to its value in the fixed (base) period is called a general index number for the group."

General index number for the group $I = \frac{\Sigma\left[\frac{p_{1i}}{p_{0i}}\right]}{n} \times 100$

Where, general index number I = Index number of current period with respect to comparison period

p_{1i} = Value of variable i for current period $(i = 1, 2, 3, ..., n)$

p_{0i} = Value of variable i for comparison period $(i = 1, 2, 3, ..., n)$

n = Number of values of the variable

The simple mean is used in the definition of general index number for the group of n items. But the weighted mean or geometric mean can also be used in the definition of general index number, which will be discussed later in this chapter.

In practice, several mutually related items are to be included and the data regarding their prices should be obtained to find a price index number. For example, wheat, rice, pulses, oil, ghee, jaggery, spices, vegetables are included in the category of food items. Thus, price index number for food is an index associated with the relative change or price relative for the prices of several related items.

Now, if we take a group of n such mutually related items then an index is found using relative change in price of each item in that group. An average measure obtained from them is called the price index numbers for the group. It can be written as the following formula :

General price index number $I = \frac{\Sigma\left[\frac{p_{1i}}{p_{0i}}\right]}{n} \times 100$

Where, p_{1i} = price of item i in current period $(i = 1, 2, 3, ..., n)$

p_{0i} = price of item i in base period $(i = 1, 2, 3, ..., n)$

n = number of items

Further, if we take a group of n such mutually related items then an index is found using relative change in the quantity of each item in that group. An average measure obtained from them is called the quantity index number for the group.

Note : The index number for production, import, export, unemployment, industrial output, etc. can be obtained by the above formula.

1.2 Characteristics of Index Number

Some of the characteristics of index number deduced from its definition are as follows :

(1) Index number is free from the unit as it is a relative measure.

(2) The changes in the values of the variable having different units can be compared using index number. Hence, index number is a comparative measure.

(3) Index number is a relative measure showing percentage change.

(4) Index number is a special average. It has all the characteristics of an average.

(5) The situation at two different periods can be compared by ratio with the standard (base) period using an index number.

1.3 Uses of Index Number

A general notion about the index number is, an index number is only used to find a measure for changes in the value of a variable or the price level. But now its use is not limited to the study of change in the price level. The index number is used in various fields in the current revolutionary age. Index number is a useful statistical tool to study the challenges in the given economic, political, social and industrial activities. Index number provides important guidance for planning the economic development of a country as it gives a comparative study of economic and industrial scenario of the country. Some of the uses of index number are as follows :

(1) Index Number for Trade : This index number provides useful guidance to study the general situation of economic activities of business and trade in the country.

(2) Wholesale Price Index Number : This index number measures the changes in the general price level in the country. This index number is useful to the government, producers and businessmen to take policy decisions such as knowing the demand and supply of items in the economy, estimating the future values and planning the future. The Reserve Bank of India uses this index number to take necessary steps to control inflation by studying the changes in price levels. Using the wholesale price index number, the rate of inflation is found as follows.

$$\text{Rate of inflation} = \frac{\left(\begin{array}{c}\text{Wholesale index}\\ \text{number of current year}\end{array}\right) - \left(\begin{array}{c}\text{Wholesale index number}\\ \text{of previous year}\end{array}\right)}{\text{Wholesale index number of previous year}} \times 100$$

(3) Cost of Living Index Number : This index number is useful to study the changes in the cost of living of people of different sections. This index number helps to determine purchasing power of money, salary to employees, dearness allowance, bonus, to calculate real wage and to devise tax policies by the government.

(4) Index Number of Human Development : This index number is useful to determine the state of human resource, standard of living, life expectancy and level of education and it gives information about human resource development.

(5) Index Number of National Income : This index number is useful to evaluate economic condition of the country and to determine targets for the five year plans by the government. The suggestions for increasing the national income, production and per capita income of the country can also be given using this index number by studying the changes in the national income of the country.

(6) Index Number of Industrial Production : This index number is very useful to study the changes in the production in industrial and craft fields. This index number is helpful to increase the rate of development of the country, planning industrial and trading activities.

(7) Index Number of Agricultural Production : This index number is useful to study the changes in the prices of agricultural production. The government plans agricultural policies using this index number. Moreover, this index number is helpful in forming policy to give proper support price to the farmers for their production.

(8) Index Number of Import-Export : This index number is useful to determine import-export policy, exchange rate, foreign exchange requirement and the rate of excise on goods and to provide necessary suggestions.

(9) Index Number for Employment : This index number provides the picture of employment, unemployment prevailing in the country. This shows the problems of unemployment which facilitates human resource planning.

(10) Index Number of Capital Investment : The changes in prices of shares and stocks, debentures, government securities and flow of capital investment can be studied by this index number. It also helps to estimate the trend of prices of shares and stocks.

(11) Index Number of Raw Material : This index number provides necessary guidance to traders, businessmen, economists, etc. for the policies of production-sales.

As barometer is used to predict weather, air pressure, cyclone and rain, the index number is a necessary tool for the measurement and comparative study of changes in the economic, business and social activities of the country. Hence, an index number is called the barometer of the economy of a country.

1.4 Base Year

In the construction of index numbers, the value of a variable for the current period is compared with the value of the variable with a fixed period (usually from the past). This fixed period or year is called the base year. The fixed year from the past can be the preceding year or any year before that. The period or year for which the value is to be compared with the base period or year is called the current period or year. For example, if the price of an item in the year 2016 is to be compared with the price of the same item in the year 2015, the year 2015 is called base year and the year 2016 is called the current year.

The year selected as base year should be standard or normal. It should be free from natural calamities like floods, draught, earthquake, abnormal man-made events like war, revolt, riot, strike, agitation, political events, economic disturbance or any unusual events. It is also necessary that the base year should not be from a distant past. If the base year selected is an unusual year and the values of the variable are unusually high or low then the value of index number could be misleading and it will not reveal the realistic picture of the current situation. Thus, the base year should be carefully selected while constructing index number.

The base year can be selected in two ways : (1) Fixed Base Method (2) Chain Base Method

1.4.1 Fixed Base Method

In this method, a stable period or year with usual events or situation is selected as a normal year or base year. But sometimes it becomes difficult to select a normal or base year. In this case, an average value of certain years is taken as the value of the variable for the base year. Index number is obtained by comparing value of the variable in the current year with the value of variable for the base year. The base year should be changed periodically so that it does not become a year of the distant past. The index number by fixed base method is obtained from the following formula :

$$\text{Index number } I = \frac{\text{Value of the variable in current year (period)}}{\text{Value of the variable in base year (period)}} \times 100$$

$$= \frac{p_1}{p_0} \times 100$$

Where, p_1 = Value of the variable in current year (period)

p_0 = Value of the variable in base year (period)

Illustration 1 : The data about wholesale prices of wheat in a region are as follows. Taking the year 2005 as the base year, prepare the index numbers for the price of the item for the remaining years. State the percentage increase in the price of wheat in the year 2013 from these index numbers.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013
Price per Quintal (₹)	1650	1690	1730	1750	1810	1850	1870	1900	1950

We will find the fixed base index number as the year 2005 is to be taken as the base year. The index number for the price of wheat in the year 2005 will be taken as 100.

Year	**Price of wheat per Quintal (₹)**	**Index number** $= \frac{p_1}{p_0} \times 100$
2005	1650	$\frac{1650}{1650} \times 100 = 100$
2006	1690	$\frac{1690}{1650} \times 100 = 102.42$
2007	1730	$\frac{1730}{1650} \times 100 = 104.85$
2008	1750	$\frac{1750}{1650} \times 100 = 106.06$
2009	1810	$\frac{1810}{1650} \times 100 = 109.70$
2010	1850	$\frac{1850}{1650} \times 100 = 112.12$
2011	1870	$\frac{1870}{1650} \times 100 = 113.33$
2012	1900	$\frac{1900}{1650} \times 100 = 115.15$
2013	1950	$\frac{1950}{1650} \times 100 = 118.18$

It can be said that the increase in the price of wheat in the year 2013 is (118.18 – 100) = 18.18 % with respect to the year 2005.

Illustration 2 : The prices per unit (₹) of six food items in the year 2014 and 2015 are given in the following table. Taking 2014 as the base year, compute the general index number for the price of food items and state the overall rise in prices of these food items.

Item	Unit	Price per unit (₹) of the item	
		Year 2014	Year 2015
Bread	Packet	25	28
Eggs	Dozen	30	35
Ghee	Tin	375	380
Milk	Litre	36	40
Cheese	Kilogram	440	500
Butter	Kilogram	265	300

A general index number for the price of these items for the current year 2015 is to be obtained with the base year 2014. We will find price relatives $\frac{p_1}{p_0}$ by taking base year price as p_0 and current year price as p_1. The calculation is shown in the following table :

Item	Price of item (₹)		Price relative $= \frac{p_1}{p_0}$
	p_0	p_1	
Bread	25	28	$\frac{28}{25} = 1.1200$
Eggs	30	35	$\frac{35}{30} = 1.1666$
Ghee	375	380	$\frac{380}{375} = 1.0133$
Milk	36	40	$\frac{40}{36} = 1.1111$
Cheese	440	500	$\frac{500}{440} = 1.1364$
Butter	265	300	$\frac{300}{265} = 1.1321$
Total			**= 6.6795**

General index number of six food items I $= \dfrac{\Sigma\left[\frac{p_1}{p_0}\right]}{n} \times 100$

$= \dfrac{6.6795}{6} \times 100$

$= 111.33$

$\therefore$ General price index number of six food items is $I = 111.33$.

It can be seen from the value of the index number I that there is an overall rise in prices of food items by (111.33 – 100) = 11.33 % in the year 2015 as compared to the year 2014.

Illustration 3 : The data about sugar production of a sugar manufacturing company from the year 2008 to 2015 are as follows. Prepare index number by fixed base method from these data by taking average production of the years 2009, 2010 and 2011 as the production of the base year.

Year	2008	2009	2010	2011	2012	2013	2014	2015
Production (thousand tons)	186	196	202	214	229	216	226	230

The average production of the years 2009, 2010 and 2011 $= \dfrac{196 + 202 + 214}{3} = \dfrac{612}{3} = 204$

Year	**Production (thousand tons)**	**Index number by fixed base method** $= \dfrac{p_1}{p_0} \times 100$
2008	186	$\frac{186}{204} \times 100 = 91.18$
2009	196	$\frac{196}{204} \times 100 = 96.08$
2010	202	$\frac{202}{204} \times 100 = 99.02$
2011	214	$\frac{214}{204} \times 100 = 104.90$
2012	220	$\frac{220}{204} \times 100 = 107.84$
2013	216	$\frac{216}{204} \times 100 = 105.88$
2014	226	$\frac{226}{204} \times 100 = 110.78$
2015	230	$\frac{230}{204} \times 100 = 112.75$

Merits and limitations of fixed base method

Merits : (1) Uniformity is maintained in calculation and comparison of the relative changes in the values of the variable as the base year is constant in this method.

(2) This method is useful to compare the long term changes in the values of the variable.

(3) This method is easy to understand and compute.

Limitations : (1) The taste, habits and fashion of consumers change with time and hence there is a change in the items used by the consumer. The items with reduced usage which were used in the past can not be removed in this method.

(2) It is not always possible to have a standard year with normal conditions as the base year. Therefore, selection of the base year is difficult.

The reliability of the index number reduces if the base year is not selected appropriately.

(3) This method is not suitable to compare the short term changes in the value of the variable.

(4) The quality of selected items keeps changing. It is not possible to make necessary change in their weights in this method.

(5) If the base year is a year of very remote past, the comparison can not be considered to be appropriate.

1.4.2 Chain Base Method

A fixed year or period is not taken as a base year or period in this method. For every current year, its preceding year is taken as a base year. For example, the year 2015 is taken as a base year for the index number of the year 2016. The base year keeps changing in this method. Since the base year is repeatedly changed, this method is called chain base method. The current situation is compared with the recent past situation in this method. The index number by this method is found using the following formula :

$$\text{Index number} = \frac{\text{Value of the variable for current year (period)}}{\text{Value of the variable for preceding year (period)}} \times 100$$

$$\therefore \quad I = \frac{p_1}{p_0} \times 100$$

Illustration 4 : The data about bi-monthly closing prices of shares of a company in the year 2014 are given. Compute the chain base index numbers from these data.

Month	January	March	May	July	September	November
Price (₹)	22	21.20	22	23	24.70	26.00

The price for the month before January is not given here. Hence, we will take the index number for January, 2014 as 100. The calculation of index numbers for remaining months using the chain base method are shown in the following table.

Month	Price of share (₹)	Chain base index number $= \frac{\text{Value of the variable in current month}}{\text{Value of the variable in preceding month}} \times 100$
January	22.00	= 100
March	21.20	$\frac{21.20}{22.00} \times 100$ = 96.36
May	22.00	$\frac{22.00}{21.20} \times 100$ = 103.77
July	23.00	$\frac{23.00}{22.00} \times 100$ = 104.55
September	24.70	$\frac{24.70}{23.00} \times 100$ = 107.39
November	26.00	$\frac{26.00}{24.70} \times 100$ = 105.26

Merits and limitations of chain base method

Merits : (1) The problem of selecting the base year does not arise in this method because at any given time the preceding year (period) is taken as the base year (period).

(2) As the comparison is with the preceding year, new items can be included according to the taste and choice of the consumers. It is possible to remove the items not in use.

(3) This method is useful in the fields of economics, trade and commerce as the value of the variable in the current period is compared with the period in the recent past.

Limitations : (1) This method is suitable only for short term comparison of the value of the variable in the current period as the preceding year is taken as the base year. The method is not very convenient for long term comparison.

(2) If there is an error in the calculation of index number by this method then the effect of that error continues in the interpretation of the index number of the succeeding year.

(3) There is no uniformity in the computation of index numbers obtained by this method.

(4) If the information for a year is not available then the index number for the next year can not be obtained.

Illustration 5 : The data about the purchase of groundnut by an edible oil mill from the year 2008 to 2015 are as follows. Prepare the index numbers by fixed base method with the year 2008 as the base year, with chain base and by taking the average quantity purchased in the year 2010 and 2011 as the purchase for the base year.

Year	2008	2009	2010	2011	2012	2013	2014	2015
Purchase of groundnut (ton)	230	250	230	250	270	280	300	300

Year	Quantity Purchase of groundnut (ton)	Index number with base year 2008 $= \frac{\text{Value of variable in current year}}{\text{Value of variable in base year}} \times 100$	Chain base Index number $= \frac{\text{Value of variable in current year}}{\text{Value of variable in preceding year}} \times 100$	Index number by taking average of quantity in year 2010 and 2011 $= \frac{230 + 250}{2} = 240$ as base year quantity
2008	230	= 100	= 100	$\frac{230}{240} \times 100 = 95.83$
2009	250	$\frac{250}{230} \times 100 = 108.70$	$\frac{250}{230} \times 100 = 108.70$	$\frac{250}{240} \times 100 = 104.17$
2010	230	$\frac{230}{230} \times 100 = 100$	$\frac{230}{250} \times 100 = 92.00$	$\frac{230}{240} \times 100 = 95.83$
2011	250	$\frac{250}{230} \times 100 = 108.70$	$\frac{250}{230} \times 100 = 108.70$	$\frac{250}{240} \times 100 = 104.17$
2012	270	$\frac{270}{230} \times 100 = 117.39$	$\frac{270}{250} \times 100 = 108$	$\frac{270}{240} \times 100 = 112.5$
2013	280	$\frac{280}{230} \times 100 = 121.74$	$\frac{280}{270} \times 100 = 103.70$	$\frac{280}{240} \times 100 = 116.67$
2014	300	$\frac{300}{230} \times 100 = 130.43$	$\frac{300}{280} \times 100 = 107.14$	$\frac{300}{240} \times 100 = 125$
2015	300	$\frac{300}{230} \times 100 = 130.43$	$\frac{300}{300} \times 100 = 100$	$\frac{300}{240} \times 100 = 125$

Illustration 6 : The data about sale of three grain flour wheat, bajri and chana at a flour mill from the year 2011 to 2015 are as follows. Compute the general index number using simple average with (i) fixed base method (taking base year 2011) and (ii) Chain base method.

Year → / Grain flour	Sale (lakh ₹)				
	2011	**2012**	**2013**	**2014**	**2015**
Wheat flour	40	46	50	56	64
Bajri flour	20	30	36	42	54
Chana flour	50	64	80	96	112

(i) Fixed base method

$$\text{Fixed base index number } I = \frac{\text{Value of variable in current year (period)}}{\text{Value of variable in base year (period)}} \times 100$$

Year / Grain flour	2011	2012	2013	2014	2015
Wheat flour	100	$\frac{46}{40} \times 100 = 115$	$\frac{50}{40} \times 100 = 125$	$\frac{56}{40} \times 100 = 140$	$\frac{64}{40} \times 100 = 160$
Bajri flour	100	$\frac{30}{20} \times 100 = 150$	$\frac{36}{20} \times 100 = 180$	$\frac{42}{20} \times 100 = 210$	$\frac{54}{20} \times 100 = 270$
Chana flour	100	$\frac{64}{50} \times 100 = 128$	$\frac{80}{50} \times 100 = 160$	$\frac{96}{50} \times 100 = 192$	$\frac{112}{50} \times 100 = 224$
Total	300	393	465	542	654
General index number of sale $= \frac{\text{Total}}{3}$	$\frac{300}{3}$ $= 100$	$\frac{393}{3}$ $= 131$	$\frac{465}{3}$ $= 155$	$\frac{542}{3}$ $= 180.67$	$\frac{654}{3}$ $= 218$

(ii) General index number by chain base method :

$$\text{Chain base index number } I = \frac{\text{Value of variable in current year (period)}}{\text{Value of variable in preceding year (period)}} \times 100$$

Year / Grain flour	2011	2012	2013	2014	2015
Wheat flour	100	$\frac{46}{40} \times 100 = 115$	$\frac{50}{46} \times 100 = 108.70$	$\frac{56}{50} \times 100 = 112$	$\frac{64}{56} \times 100 = 114.29$
Bajri flour	100	$\frac{30}{20} \times 100 = 150$	$\frac{36}{30} \times 100 = 120$	$\frac{42}{36} \times 100 = 116.67$	$\frac{54}{42} \times 100 = 128.57$
Chana flour	100	$\frac{64}{50} \times 100 = 128$	$\frac{80}{64} \times 100 = 125$	$\frac{96}{80} \times 100 = 120$	$\frac{112}{96} \times 100 = 116.67$
Total	300	393	353.7	348.67	359.53
Aggregate index number $= \frac{\text{Total}}{3}$	$\frac{300}{3}$ $= 100$	$\frac{393}{3}$ $= 131$	$\frac{353.7}{3}$ $= 117.90$	$\frac{348.67}{3}$ $= 116.22$	$\frac{359.53}{3}$ $= 119.84$

Illustration 7 : The following data are available about the crimes in a city. Find the general index number by fixed base method considering the year 2010 as base year.

Type of Crime \ Year	2007	2008	2009	2010
Murder	110	128	134	129
Violence and rape	30	45	40	48
Robbery	610	720	770	830
Theft of property	2450	2630	2910	2890

Fixed base index number $I = \frac{\text{Value of variable in current year (period)}}{\text{Value of variable in base year (period)}} \times 100$

Type of Crime \ Year	2007	2008	2009	2010
Murder	$\frac{110}{129} \times 100 = 85.27$	$\frac{128}{129} \times 100 = 99.22$	$\frac{134}{129} \times 100 = 103.88$	$\frac{129}{129} \times 100 = 100$
Violence and rape	$\frac{30}{48} \times 100 = 62.5$	$\frac{45}{48} \times 100 = 93.75$	$\frac{40}{48} \times 100 = 83.33$	$\frac{48}{48} \times 100 = 100$
Robbery	$\frac{610}{830} \times 100 = 73.49$	$\frac{720}{830} \times 100 = 86.75$	$\frac{770}{830} \times 100 = 92.77$	$\frac{830}{830} \times 100 = 100$
Theft of property	$\frac{2450}{2890} \times 100 = 84.78$	$\frac{2630}{2890} \times 100 = 91.00$	$\frac{2910}{2890} \times 100 = 100.69$	$\frac{2890}{2890} \times 100 = 100$
Total	306.04	370.72	380.67	400
General Index number of crime $= \frac{\text{Total}}{4}$	$\frac{306.04}{4}$ $= 76.51$	$\frac{370.72}{4}$ $= 92.68$	$\frac{380.67}{4}$ $= 95.17$	$\frac{400}{4}$ $= 100$

Downloaded from https:// www.studiestoday.com

EXERCISE 1.1

1. The data about average daily wage of a group of workers employed in a factory in a city during the year 2008 to 2015 are as follows. Find the index number by (1) Fixed base method (taking base year 2008) (2) Chain base method (3) Fixed base method by taking average of average daily wages of the years 2011 to 2013 as the wage for the base year.

Year	2008	2009	2010	2011	2012	2013	2014	2015
Average daily wage (₹)	275	284	289	293	297	313	328	345

2. From the following data about the retail prices of sugar in a city, find the index numbers of price of sugar by (1) Fixed base method with year 2008 as base year (2) Chain base method (3) taking the average price of sugar for the year 2009 and 2010 as the base year price.

Year	2008	2009	2010	2011	2012	2013	2014	2015
Price of Sugar per kilogram (₹)	28	28.50	29.50	30	31	32	34	36

3. The following data are obtained about the annual average prices of wheat, rice and sugar in the wholesale market of a city. Find the general index number for three items by fixed base method with base year 2011 and by chain base method.

Item \ **Year**	2011	2012	2013	2014	2015
Wheat	18	18.50	18.90	19	19.50
Rice	30	36	38	38	39
Sugar	30	31	32	34	36

4. The prices of five fuel related items in the years 2012 and 2014 are as follows. Calculate the general index number for five fuel items by taking the year 2012 as the base year and state the overall increase in the prices of fuel items.

Item	Electricity	Gas	Match Box	Kerosene	Wood
Unit	Unit	Cylinder	Box	Litre	Kilogram
Price in 2012 (₹)	3	345	1.00	15	12
Price in 2014 (₹)	3.5	370	1.50	20	15

*

1.5 Conversion from Fixed Base to Chain Base and from Chain Base to Fixed Base

Generally, whenever the fixed base or chain base index numbers only are available instead of the original information about the values of the variable, the conversion of base is necessary for the following reasons. If the need arises to find the short term changes in the values of the variable then it becomes difficult to find it from the fixed base index numbers. It is easier to find the short term variations after converting the given fixed base index numbers into chain base index numbers.

Sometimes, it is necessary to compare the value of the variable at a given period to the value of another period in a series of values of the variable. This is not possible if only chain base index numbers are available. The above comparison is possible in this situation if the chain base index numbers are converted to the fixed base index numbers. Thus, it is necessary to convert the chain base index numbers into the fixed base index numbers. Hence, the base conversion is carried out as follows :

Conversion of the fixed base index number to the chain base index numbers : The formula for the conversion of the fixed base index numbers into the chain base index numbers is as follows.

$$\text{Chain base index number} = \frac{\text{Fixed base index number of current year}}{\text{Fixed base index number of preceding year}} \times 100$$

Note : If the base year is not mentioned then we will take the chain base index number for the first year as 100. If the base year is mentioned then the fixed base index number of the first year will be taken as its chain base index number.

Illustration 8 : Convert the following index numbers obtained by fixed base method about the production of craft industry of a state into the chain base index numbers.

Year	2009	2010	2011	2012	2013	2014
Fixed base index numbers	120	132	96	144	138	108

Since the base year is not mentioned here, we will take 100 as the chain base index number for the first year.

$$\text{Chain base index number} = \frac{\text{Fixed base index number of current year}}{\text{Fixed base index number of preceding year}} \times 100$$

Year	**Index number**	**Chain base index number**
2009	120	= 100
2010	132	$\frac{132}{120} \times 100 = 110$
2011	96	$\frac{96}{132} \times 100 = 72.73$
2012	144	$\frac{144}{96} \times 100 = 150$
2013	138	$\frac{138}{144} \times 100 = 95.83$
2014	108	$\frac{108}{138} \times 100 = 78.26$

Illustration 9 : The wholesale price index numbers for commodities with the base year 2007-08 are as follows. Compute the chain base index numbers.

Year	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16
Wholesale price index number	126	130.8	143.3	156.1	167.6	177.6	181.2	177.2

The base year 2007-08 is mentioned here. Hence, we will take the given fixed base index number for the year 2008-09 as chain base index number. Thus, the chain base index number for the first year is 126.

$$\text{Chain base index number} = \frac{\text{Fixed base index number of the current year}}{\text{Fixed base index number of the preceding year}} \times 100$$

Year	**Wholesale price index number of commodities**	**Chain base index number**
2008-09	126	= 126
2009-10	130.8	$\frac{130.8}{126} \times 100 = 103.81$
2010-11	143.3	$\frac{143.3}{130.8} \times 100 = 109.56$
2011-12	156.1	$\frac{156.1}{143.3} \times 100 = 108.93$
2012-13	167.6	$\frac{167.6}{156.1} \times 100 = 107.37$
2013-14	177.6	$\frac{177.6}{167.6} \times 100 = 105.97$
2014-15	181.2	$\frac{181.2}{177.6} \times 100 = 102.03$
2015-16	177.2	$\frac{177.6}{181.2} \times 100 = 97.79$

Conversion of chain base index numbers to fixed base index number : If the year-wise chain base index numbers are given, the fixed base index number can be found accordingly. To obtain the fixed base index numbers, the chain base index number of that year is multiplied by the fixed base index number of the previous year and the product is divided by 100.

$$\text{Thus, Fixed base index number of current year} = \frac{\left(\begin{array}{c}\text{Chain base index number}\\ \text{of the current year}\end{array}\right) \times \left(\begin{array}{c}\text{Fixed base index number of the}\\ \text{preceding year to current year}\end{array}\right)}{100}$$

Let us understand this method with an illustration.

Illustration 10 : The chain base index numbers obtained for food items from the year 2008-09 to 2015-16 are as follows. Compute the fixed base index numbers. (Take 2007-08 as base year)

Year	2008 – 09	2009 – 10	2010 – 11	2011 – 12	2012 – 13	2013 – 14	2014 – 15	2015 – 16
Index number of food items	134.8	115.28	115.57	107.29	109.91	112.80	106.24	102.48

The year 2007-08 is to be taken as the base year here. Hence, the fixed base index number for the year 2008-09 will not change.

$$\text{Current year fixed base index number} = \frac{\left(\begin{array}{c}\text{Current year chain base}\\ \text{index number}\end{array}\right) \times \left(\begin{array}{c}\text{Fixed base index number of}\\ \text{preceding year to current year}\end{array}\right)}{100}$$

Year	Index number of food items	Fixed base index number
2008-09	134.8	= 134.8
2009-10	115.28	$\frac{115.28 \times 134.8}{100}$ = 155.40
2010-11	115.57	$\frac{115.57 \times 155.40}{100}$ = 179.60
2011-12	107.29	$\frac{107.29 \times 179.60}{100}$ = 192.69
2012-13	109.91	$\frac{109.91 \times 192.69}{100}$ = 211.79
2013-14	112.80	$\frac{112.80 \times 211.79}{100}$ = 238.9
2014-15	106.24	$\frac{106.24 \times 238.9}{100}$ = 253.81
2015-16	102.48	$\frac{102.48 \times 253.81}{100}$ = 260.10

EXERCISE 1.2

1. The chain base index numbers of agricultural production of a state from the year 2008 to 2014 are as follows. Compute the fixed base index numbers. (Take 2007 as base year.)

Year	2008	2009	2010	2011	2012	2013	2014
Index number of agricultural production	100	110	95	108	120	106	110

2. Obtain the chain base index number from the fixed base index numbers given below with the year 2007-08 as the base year for the wholesale prices of machines and equipments.

Year	2008 – 09	2009 – 10	2010 – 11	2011 – 12	2012 – 13	2013 – 14	2014 – 15
Index number of machines and equipments	117.4	118	121.3	125.1	128.4	131.6	134.6

3. The fixed base index numbers of food from the month of January to October in the year 2015 for the industrial workers of Ahmedabad are as given below. Compute the chain base index numbers.

Month	January	February	March	April	May	June	July	August	September	October
Index number of food	271	270	268	268	278	283	283	293	293	299

4. The chain base index numbers for sales of a certain type of scooter from the year 2010 to 2015 are as follows. Find fixed base index numbers.

Year	2010	2011	2012	2013	2014	2015
Index number of sale	110	112	109	108	105	111

*

1.6 Specific Formulae for Computing Index Number

We have seen that the index number is useful to study the changes in the values of variable for an item or the values of variables for items in a group. Simple average is used in the construction of an index number and every item is given equal weightage. But the importance of every item is generally not same in practice. For example, the importance given to grains is not same as the importance given to vegetables, pulses or edible oil. Thus, the index number of food will be more realistic and meaningful if each item is assigned weight according to its importance.

The weights of items included are decided in the construction of different types of index numbers. Generally, the weights given to the items for constructing the index number are determined on the basis of their quantity consumed. We shall study some specific formulae for computing index numbers by taking this fact into consideration where different methods of selecting weights are taken for the construction of index number.

Method of weighted average : Suppose I_i is the index number of the i th group among the groups of items (or items) wheat, rice and pulses with the corresponding weight W_i, then the general index number of these groups is obtained using the following formula.

General index number $I = \frac{\Sigma I_i W_i}{\Sigma W_i} = \frac{\Sigma IW}{\Sigma W}$

Note : We shall ignore the suffix 'i' for the simplicity of calculation.

For example, if the index numbers of these groups are 120, 150 and 300 respectively and their corresponding weights are 3, 2 and 1 then the general index number for the group of items is

$$I = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{120 \times 3 + 150 \times 2 + 300 \times 1}{3 + 2 + 1}$$

$$= \frac{360 + 300 + 300}{6}$$

$$= \frac{960}{6}$$

$$= 160$$

Laspeyre's Formula

This method of finding the index number is given by Laspeyere. It is one of the important methods of finding index number. In this method, base year price is denoted by p_0 and the quantity is denoted by q_0 whereas the prices of items in the current year are denoted by p_1. The expenditure $p_0 q_0$ is assigned as weight to the price relative $\frac{p_1}{p_0}$. The formula of weighted index number thus obtained is called the formula of **Laspeyre's index**, which is denoted by I_L. The Laspeyre's formula is as follows :

$$\textbf{Laspeyre's index number } I_L = \frac{\Sigma\left[\frac{p_1}{p_0}\right] \times p_0 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{\Sigma \frac{p_1}{p_0} \times p_0 q_0}{\Sigma p_0 q_0} \times 100$$

$$\therefore \quad I_L = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Paasche's Formula

This method is given by an economist named Paasche. If we denote p_0 as base year price, p_1 as current year price and q_1 as current year quantity then the expenditure $p_0 q_1$ is assigned as weight for the price relative $\frac{p_1}{p_0}$. The formula of weighted index number thus obtained is called the formula of **Paasche's index number**, which is denoted by I_P. The formula of Paasche's index number is as follows :

$$\textbf{Paasche's index number } I_P = \frac{\Sigma\left[\frac{p_1}{p_0}\right] \times p_0 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{\Sigma \frac{p_1}{p_0} \times p_0 q_1}{\Sigma p_0 q_1} \times 100$$

$$\therefore \quad I_P = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

Fisher's Formula

The base year and current year quantities are taken into condsideration for computing the weight in Laspeyre's and Paasche's method respectively. Prof. Irving Fisher has constructed an index number by considering quantities of both the years. The geometric mean of Laspeyre's and Paasche's index numbers is called **Fisher's index number**, which is denoted by I_F. The formula of Fisher's index number is as follows :

Fisher's index number $I_F = \sqrt{I_L \times I_P}$ or

Fisher's index number $I_F = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$

The Fisher's index number is called ideal index number due to the following reasons :

(1) The quantities of both the years, base year and current year, are taken in the computation for constructing this index number.

(2) This index number satisfies both the important fundamental tests, time reversal and factor reversal tests, of index numbers.

(3) The geometric mean is used to calculate this index number which is the best average for the construction of index number.

(4) This index number is free from bias as it balances the demerits of Laspeyre's and Paasche's index number.

Thus, Fisher's index number is an ideal index number.

Illustration 11 : Find the index number for the year 2016 with base year 2011 by weighted average method from the following data of price and weights of five different items.

Item	Weight	Price (₹)	
		Year 2011	Year 2016
A	40	160	200
B	25	400	600
C	5	50	70
D	20	10	18
E	10	2	3

The weights of different items are given here. We shall compute the general index number from the price relatives of the year 2016 based on the prices of the year 2011.

Item	Weight W	Price (₹) p_0	Price (₹) p_1	$I = \frac{p_1}{p_0} \times 100$	IW
A	40	160	200	$\frac{200}{160} \times 100 = 125$	5000
B	25	400	600	$\frac{600}{400} \times 100 = 150$	3750
C	5	50	70	$\frac{70}{50} \times 100 = 140$	700
D	20	10	18	$\frac{18}{10} \times 100 = 180$	3600
E	10	2	3	$\frac{3}{2} \times 100 = 150$	1500
Total	**100**				**14,550**

Index number of year 2016 $I = \frac{\Sigma IW}{\Sigma W}$

$$= \frac{14550}{100}$$

$$= 145.50$$

Thus, we say that there is an increase of (145.50 – 100) = 45.5 % in prices in the year 2016 as compared to the year 2011.

Illustration 12 : Find Laspeyre's, Paasche's and Fisher's index numbers for the year 2016 with base year 2015 from the data about price and consumption of food items given below.

Item	Unit	Year 2016		Year 2015	
		Price (₹)	Quantity	Price (₹)	Quantity
Rice	Kilogram	40	1.5 Kilogram	39	1 Kilogram
Milk	Litre	44	10 Litre	40	12 Litre
Bread	Kilogram	50	1.5 Kilogram	45	2 Kilogram
Banana	Dozen	36	1.5 Dozen	30	2 Dozen

We will take price p_0 and quantity q_0 for the base year, price p_1 and quantity q_1 for the current year.

Item	**Unit**	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
Rice	Kilogram	39	1	40	1.5	40	39	60	58.5
Milk	Litre	40	12	44	10	528	480	440	400
Bread	Kilogram	45	2	50	1.5	100	90	75	67.5
Banana	Dozen	30	2	36	1.5	72	60	54	45
Total						**740**	**669**	**629**	**571**

Laspeyre's index number $I_L = \dfrac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$

$= \dfrac{740}{669} \times 100$

$= 110.6128$

$\simeq 110.61$

Thus, there is a rise of $(110.61 - 100) = 10.61$ % in prices of the year 2016 as compared to the base year 2015.

Paasche's index number $I_P = \dfrac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100$

$= \dfrac{629}{571} \times 100$

$= 110.1576$

$\simeq 110.16$

Thus, there is a rise of $(110.16 - 100) = 10.16$ % in prices of the year 2016 as compared to the base year 2015.

Fisher's index number $I_F = \sqrt{I_L \times I_P}$

$= \sqrt{110.61 \times 110.16}$

$= 110.3847$

$\simeq 110.38$

Thus, there is a rise of $(110.38 - 100) = 10.38$ % in prices of the year 2016 as compared to the base year 2015.

Illustration 13 : Compute Laspeyre's, Paasche's and Fisher's index numbers for the year 2016 from the data given below by taking 2015 as the base year.

Item	Unit	Price (₹)		Quantity (Consumption)	
		Year 2015	Year 2016	Year 2015	Year 2016
A	20 Kilogram	300	440	5 Kilogram	8 Kilogram
B	Quintal	500	700	10 Kilogram	15 Kilogram
C	Kilogram	60	75	1200 Gram	2000 Gram
D	Meter	14.25	15	15 Meter	25 Meter
E	Litre	32	36	18 Litre	30 Litre
F	Dozen	30	36	8 Pieces	10 Pieces

The base year is 2015 and the current year is 2016. Hence, we will take price p_0 and quantity q_0 for the year 2015, price p_1 and quantity q_1 for the year 2016.

The price of item *A* is per 20 kg here whereas the unit for quantity is kg. The price of item *B* is per quintal but the unit for the quantity is kg. The price of item *C* is per kg whereas the unit for quantity is gram. The price for item *F* is per dozen whereas the unit for quantity is piece. The calculation of the price per item of these four items will be as follows :

The price of item *A* in the year 2015 is ₹ 300 per 20 kg. Hence, its price = $\frac{300}{20}$ = ₹ 15 per kg.

Similarly, the price of item *A* in 2016 is $\frac{440}{20}$ = ₹ 22 per kg.

It is convenient to express the price of item *B* per kg than quintal. Hence, the price for the year 2015 = $\frac{500}{100}$ = ₹ 5 per kg and the price for the year 2016 = $\frac{700}{100}$ = ₹ 7 per kg.

The price of item *C* is per kg. Hence, it is convenient to express its quantity in kg.

Thus, the quantity in the year 2015 = $\frac{1200}{1000}$ = 1.2 kg and the quantity for the year 2016 = $\frac{2000}{1000}$ = 2 kg.

The price of item *F* is per dozen which is convenient to express in per piece. Hence, the price of the year 2015 = $\frac{30}{12}$ = ₹ 2.5 per piece and the price for the year 2016 = $\frac{36}{12}$ = ₹ 3 per piece. Now, the index number will be calculated as follows :

Item	Unit	Year 2015		Year 2016		p_1q_0	p_0q_0	p_1q_1	p_0q_1
		p_0	q_0	p_1	q_1				
A	kg	15	5	22	8	110	75	176	120
B	kg	5	10	7	15	70	50	105	75
C	kg	60	1.2	75	2	90	72	150	120
D	Meter	14.25	15	15	25	225	213.75	375	356.25
E	Litre	32	18	36	30	648	576	1080	960
F	Piece	2.5	8	3	10	24	20	30	25
Total						**1167**	**1006.75**	**1916**	**1656.25**

Laspeyre's index number $I_L = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$

$= \frac{1167}{1006.75} \times 100$

$= 115.9175$

$\simeq 115.92$

Thus, we can say that there is a rise of (115.92 – 100) = 15.92 % in the prices in the year 2016 as compared to the year 2015.

Paasche's index number $I_P = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100$

$= \frac{1916}{1656.25} \times 100$

$= 115.6830$

$\simeq 115.68$

Thus, it can be said that there is (115.68 – 100) = 15.68 % rise in the prices in the year 2016 as compared to the year 2015.

Fisher's index number $I_F = \sqrt{I_L \times I_P}$

$= \sqrt{115.92 \times 115.68}$

$= 115.7999$

$\simeq 115.8$

Thus, it can be said that there is (115.8 – 100) = 15.8 % rise in the prices in the year 2016 as compared to the year 2015.

Illustration 14 : Find the ideal index number for the year 2015 from the following data.

Item	Base year 2014		Current year 2015	
	Price (₹)	Quantity	Price (₹)	Quantity
A	16	10	20	11
B	20	9	24	9
C	32	16	40	17

Fisher's index number is considered as an ideal index number. So, we will find Fisher's index number here. We will take price p_0 and quantity q_0 for base year, price p_1 and quantity q_1 for the current year.

Item	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	16	10	20	11	200	160	220	176
B	20	9	24	9	216	180	216	180
C	32	16	40	17	640	512	680	544
Total					**1056**	**852**	**1116**	**900**

$$\text{Fisher's index number } I_F = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

$$= \sqrt{\frac{1056}{852} \times \frac{1116}{900}} \times 100$$

$$= \sqrt{1 \cdot 5369} \times 100$$

$$= 1.2397 \times 100$$

$$I_F = 123.97$$

Thus, it can be said that there is (123.97 – 100) = 23.97 % rise in the prices in the year 2015 as compared to the year 2014.

Illustration 15 : Find Fisher's index number for the year 2015 by taking the year 2014 as the base year from the data given below about consumption and total expenditure of five different items.

Item	Base Year 2014		Current Year 2015	
	Consumption	Total expenditure	Consumption	Total expenditure
A	50 kg	2500	60 kg	4200
B	120 kg	600	140 kg	700
C	30 litre	330	20 litre	200
D	20 kg	360	15 kg	300
E	5 kg	40	5 kg	50

The consumption and total expenditure for the items are given here.

Total expenditure of item = (Price of item per unit) × (Consumption of item)

$\therefore$ Price of item per unit = $\dfrac{\text{Total expenditure of item}}{\text{Consumption of item}}$

We will obtain the price per unit of each item using the above formula.

Item	Base year 2014		Current year 2015		p_1q_0	p_0q_0	p_1q_1	p_0q_1
	Quantity	$p_0 = \dfrac{\text{Expenditure}}{q_0}$	Quantity	$p_1 = \dfrac{\text{Expenditure}}{q_1}$				
	q_0	p_0	q_1	p_1				
A	50	$\frac{2500}{50} = 50$	60	$\frac{4200}{60} = 70$	3500	2500	4200	3000
B	120	$\frac{600}{120} = 5$	140	$\frac{700}{140} = 5$	600	600	700	700
C	30	$\frac{330}{30} = 11$	20	$\frac{200}{20} = 10$	300	330	200	220
D	20	$\frac{360}{20} = 18$	15	$\frac{300}{15} = 20$	400	360	300	270
E	5	$\frac{40}{5} = 8$	5	$\frac{50}{5} = 10$	50	40	50	40
Total					**4850**	**3830**	**5450**	**4230**

Fisher's index number $I_F = \sqrt{\dfrac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \dfrac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$

$$= \sqrt{\frac{4850}{3830} \times \frac{5450}{4230}} \times 100$$

$$= \sqrt{1 \cdot 6315} \times 100$$

$$= 1.2773 \times 100$$

$$I_F \simeq 127.73$$

Thus, it can be said that there is (127.73 – 100) = 27.73 % rise in the prices in the year 2015 as compared to the year 2014.

Illustration 16 : The health department has implemented a certain policy for the industrial units in the year 2003 to control the possibility of cancer due to the chemical process which is hazardous to the health of workers employed in the industrial units of a certain industrial area who are residing in the same area. To evaluate this policy, a survey was conducted about deaths due to cancer of persons in the different age groups. The following data are obtained for the years 2003 and 2008. Find the index number of deaths due to cancer using weighted average method by taking the population of this industrial area in the year 2003 as weight and interpret it.

Age-group (years)	Population in year 2003 (thousand)	Deaths in year 2003	Deaths in year 2008
< 5	10	200	65
5-15	8	145	100
15-40	48	610	480
40-60	38	350	225
> 60	14	550	465

We shall obtain the general index number by finding the relative percentages of cancer deaths for the year 2008 and taking the population in different age-groups in 2003 as weights.

Age-group (years)	Population in year 2003 (thousand) W	Deaths in year 2003 p_0	Deaths in year 2008 p_1	$I = \frac{p_1}{p_0} \times 100$	IW
< 5	10	200	65	$\frac{65}{200} \times 100 = 32.5$	325
5-15	8	145	100	$\frac{100}{145} \times 100 = 68.97$	551.76
15-40	48	610	480	$\frac{480}{610} \times 100 = 78.69$	3777.12
40-60	38	350	225	$\frac{225}{350} \times 100 = 64.29$	2443.02
> 60	14	550	465	$\frac{465}{550} \times 100 = 84.55$	1183.7
Total	**118**				**8280.6**

Index number for year 2008 $I = \frac{\Sigma IW}{\Sigma W}$

$$= \frac{8280.6}{118}$$

$$= 70.1745$$

$$\simeq 70.17$$

Thus, it can be said that there is a decrease of (100 – 70.17) = 29.83 % in the deaths due to cancer in the year 2008 as compared to the year 2003.

EXERCISE 1.3

1. The information about six different items used in the production of an electronics item is as follows. Find the index number and interpret it.

Items	*A*	*B*	*C*	*D*	*E*	*F*
Weight	5	10	10	30	20	25
Percentage price relative	290	315	280	300	315	320

2. The information about six different items used in the furniture items is as follows. Find the index number for the year 2015 with the base year 2014 and interpret it.

Item	*A*	*B*	*C*	*D*	*E*	*F*
Weight	17	15	22	16	12	18
Price in year 2014 (₹)	30	20	50	32	40	16
Price in year 2015 (₹)	24	24	70	40	48	24

3. Find the Laspeyre's, Paasche's and Fisher's index numbers for the year 2015 with the base year 2014 using the following information.

Item		Wheat	Rice	Pulses	Oil	Cloth	Kerosene
	Unit	kg	kg	kg	kg	Meter	Litre
Year 2014	**Quantity**	20	10	10	6	15	18
	Price (₹)	15	20	26.50	24.80	21.25	21
Year 2015	**Quantity**	30	15	15	8	25	30
	Price (₹)	18	31.25	29.50	30	25	28.80

4. Find the Laspeyre's, Paasche's and Fisher's index numbers for the year 2015 with the base year 2014 using the following information.

Item	**Unit**	**Price (₹)**		**Quantity (Consumption)**	
		Year 2014	**Year 2015**	**Year 2014**	**Year 2015**
A	20 kg	80	120	5 kg	7 kg
B	kg	20	24	2400 gm	4000 gm
C	Quintal	2000	2800	10 kg	15 kg
D	Dozen	48	72	30 pieces	35 pieces

5. Find the ideal index number from the following data for the year 2015.

Item	Unit	Base year 2014		Base year 2015	
		Price (₹)	Quantity	Price (₹)	Quantity
A	20 kg	120	10 kg	280	15 kg
B	5 Dozen	120	3 Dozen	140	48 pieces
C	kg	4	5000 gm	8	4 kg
D	5 Litre	52	15 Litre	58	20 Litre

6. Find the Paasche's and Fisher's index numbers for the year 2015 with the base year 2014 using the data given below.

Item		*A*	*B*	*C*	*D*	*E*
Year 2014	Price (₹)	100	100	150	180	250
	Total expenditure	400	500	600	1080	1000
Year 2015	Price (₹)	120	120	160	200	300
	Total expenditure	720	600	800	1000	1200

*

1.7 Cost of Living Index Number

The cost of living index number is constructed to measure and study the changes in the cost of living of people from different sections of the society due to the fluctuations in prices. Thus, "The number showing the percentage of relative changes in the cost of living of the people of a certain section of the society in the current year (period) as compared to the base year (period) is called the **cost of living index number**."

The cost of living index number is prepared separately for the people of different sections of the society and regions.

For example, a family spends ₹ 15,000 per month for their living in the year 2012 and the same family spends ₹ 18,000 per month for their living in the year 2014 for the same lifestyle. Their cost of living index number can be obtained as follows :

$$\text{Cost of living index number} = \frac{\text{Current year (period) monthly expenditure}}{\text{Base year (period) monthly expenditure}} \times 100$$

$$= \frac{18000}{15000} \times 100$$

$$= \frac{600}{5}$$

$$= 120$$

Thus, it can be said that there is a rise of (120 – 100) = 20 % in the monthly expense in the year 2014 as compared to the year 2012.

1.7.1 Construction of Cost of Living Index Number

The following points should be considered while constructing the cost of living index number :

(1) Purpose : The purpose of every index number should be explained before constructing it. We should ascertain the class of people in the society for whom the cost of living index number is to be constructed. The requirements of the people of worker class and rich class are different. For example, the rise in price of grains does not affect much to the cost of living of the people from rich class whereas it affects a lot to the cost of living of the people from worker class. Thus, it is necessary to clarify the purpose of the construction of the cost of living index number.

(2) Family Budget Inquiry : A sample of some families is randomly selected from the families of that class of people for whom the cost of living index is to be prepared. The budget of the families selected in the sample is studied. The information is obtained about the list of different items consumed by them, their consumption, list of retail prices, the expenses incurred on them and place of purchase, etc. This type of inquiry is called sample family budget inquiry.

The data obtained from the inquiry of families included in the sample are generally divided into five sections : (a) food (b) clothing (c) house rent (d) fuel and electricity and (e) miscellaneous.

The importance of different items in the expenditure for living can be known from the sample inquiry of family budget. Hence, the importance of each item selected in the construction of the index number in its group and the importance of each group in the total expenditure can be determined.

(3) Availability of Prices of Items : The retail prices of items are collected from the areas of residence of the people from the class of families for whom the cost of living index number is to be obtained. As far as possible, these prices should be collected from the standard or government approved shops. The average price of items should be taken into consideration when the prices obtained from various shops at different times are different.

(4) Base Year : A normal year is selected as a base year. The price relatives are found as follows for each item by taking the retail prices of the normal year as the base year prices :

Price relative $I = \frac{p_1}{p_0} \times 100$

Where, p_1 = retail price of item in current year

p_0 = retail price of item in base year

(5) Average : It is necessary to find a general price relative from the price relatives of different items. A proper average should be used for this purpose. Theoretically, the geometric mean is the ideal average for the construction of index number. But due to the difficulty of its computation, it is common to use the weighted mean for the construction of index number.

(6) Weight : The importance of different items selected in the construction of index number is not same. The number associated with items in propotion to their importance is called weight. These weights can be of two types : (i) Implicit Weight and (ii) Explicit Weight.

(i) Implict Weight : This is an indirect method of assigning weights. According to this method, the weights are determined as per the number of varieties of different items selected in the construction of index number. This method is called implicit method as the weights can not be accurately quantified.

(ii) Explicit Weight : This is a direct method of assigning weights. The weights of items are expressed numerically in proportion to their importance. In this method, the weights of items are determined according to the consumption, sale, production or the expenditure for that item. Thus, the weights given in accordance with the importance of the items are called explicit weights.

The two methods of assigning explicit weight are as follows :

(1) Method of total expenditure (2) Method of family budget

(1) Method of Total Expenditure : In this method, the expenditure for every item in the base year and current year is found using the consumption of these items and further the total expenditure of all the items is obtained for both the years. The percentage ratio of the total expenditure of the current year with the total expenditure of the base year is called the index number by the method of total expenditure.

Suppose p_0 = price of base year, q_0 = quantity of base year

p_1 = price of current year, q_1 = quantity of current year

If the quantity of the base year is used for finding the total expenditure of the current and base year,

$\Sigma p_1 q_0$ = total expenditure of current year and $\Sigma p_0 q_0$ = total expenditure of base year.

Index number I = $\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$. This formula is the formula of Laspeyre's index number.

If the quantity of the current year is used for finding the total expenditure of the current and base year,

$\Sigma p_1 q_1$ = total expenditure of current year and $\Sigma p_0 q_1$ = total expenditure of base year

Index number $I = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$. This formula is the formula of Paasche's index number.

(1) Method of Family Budget : In this method, the percentage price relative I is found first for every item. Here, $I = \frac{p_1}{p_0} \times 100$ where, p_1 = price of current year and p_0 = price of base year. Then, the expenditure of every item $p_0 q_0$ in the base year is found and it is taken as the weight W for the percentage price relative I. The formula for index number by the method of family budget using the weighted average with weight $W = p_0 q_0$ is as follows :

$$\text{Index number } I = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{\Sigma\left[\frac{p_1}{p_0} \times 100 \times p_0 q_0\right]}{\Sigma p_0 q_0}$$

$$= \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Thus, the index number obtained by the method of family budget is the Laspeyre's index number.

Illustration 17 : Find the cost of living index number by the family budget method from the following information about index numbers of different groups of items for living and their weights.

Group	Food items	Clothing	Electricity-fuel	House rent	Miscellaneous
Index number	281	177	178	210	242
Weight	46	10	7	12	25

The index numbers of different groups and their weights are given here. Hence, we will use family budget method which is a method of weighted average.

Group	Index number I	Weight W	IW
Food Items	281	46	12,926
Clothing	177	10	1770
Electricity-fuel	178	7	1246
House Rent	210	12	2520
Miscellaneous	242	25	6050
Total		**100**	**24,512**

$$\text{Index number } I = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{24512}{100}$$

$$= 245.12$$

Thus, it can be said that there is a rise of (245.12 – 100) = 145.12 % in the total expenditure in the current year as compared to the base year.

Illustration 18 : Calculate the cost of living index number by the total expenditure method and the family budget method for the year 2015 with the base year 2014 using the following data.

Item	Wheat	Rice	Tuver Dal	Oil	Cloth	Kerosene
Unit	Quintal	kg	kg	litre	meter	litre
Quantity of year 2014	35 kg	25 kg	20 kg	10 litre	20 meter	15 litre
Price of year 2014 (₹)	1600	40	60	80	30	28
Price of year 2015 (₹)	1800	45	120	90	45	35

The base year is 2014. We will take p_0 = price of 2014, q_0 = quantity of 2014 and p_1 = price of 2015. We shall make uniform units for the price and quanity of each item.

Method of Total Expenditure

Item	Unit	Year 2014		Year 2015	p_1q_0	p_0q_0
		q_0	p_0	p_1		
Wheat	kg	35	16	18	630	560
Rice	kg	25	40	45	1125	1000
Tuver Dal	kg	20	60	120	2400	1200
Oil	litre	10	80	90	900	800
Cloth	meter	20	30	45	900	600
Kerosene	litre	15	28	35	525	420
Total					**6480**	**4580**

Index number by total expenditure method $= \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$

$= \frac{6480}{4580} \times 100$

$= 141.4847$

$\simeq 141.48$

Thus, there is a rise of (141.48 – 100) = 41.48 % in the total expenditure in the year 2015 as compared to the base year 2014.

Method of Family Budget

Item	Unit	Year 2014		Year 2015	$I = \frac{p_1}{p_0} \times 100$	$W = p_0q_0$	IW
		q_0	p_0	p_1			
Wheat	kg	35	16	18	$\frac{18}{16} \times 100 = 112.5$	560	63,000
Rice	kg	25	40	45	$\frac{45}{40} \times 100 = 112.5$	1000	1,12,500
Tuver dal	kg	20	60	120	$\frac{120}{60} \times 100 = 200$	1200	2,40,000
Oil	litre	10	80	90	$\frac{90}{80} \times 100 = 112.5$	800	90,000
Cloth	meter	20	30	45	$\frac{45}{30} \times 100 = 150$	600	90,000
Kerosene	litre	15	28	35	$\frac{35}{28} \times 100 = 125$	420	52,500
Total						**4580**	**6,48,000**

Index number by family budget method $= \frac{\Sigma IW}{\Sigma W}$

$= \frac{648000}{4580}$

$= 141.4847$

$\simeq 141.48$

Note : We can see here that the index numbers obtained by total expenditure method and family budget method are same.

Illustration 19 : The data referring to worker class of a city are as follows. Find the general index numbers for the years 2014 and 2015. If the wages of these workers in 2014 are increased by 5 % in the year 2015, is this rise in wages sufficient to maintain their standard of living ?

Group	Food	Clothing	Fuel and Electricity	House Rent	Miscellaneous
Weight	48	18	8	12	14
Group index number of 2014	210	220	210	200	210
Group index number of 2015	230	225	220	200	235

Group	**Weight W**	**Group index number of year 2014 I_1**	**Group index number of year 2015 I_2**	I_1W	I_2W
Food	48	210	230	10,080	11,040
Clothing	18	220	225	3960	4050
Fuel and Electricity	8	210	220	1680	1760
House Rent	12	200	200	2400	2400
Miscellaneous	14	210	235	2940	3290
Total	**100**			**21,060**	**22,540**

Index number for year 2014 $= \frac{\Sigma I_1W}{\Sigma W} = \frac{21060}{100} = 210.60$

Index number for year 2015 $= \frac{\Sigma I_2W}{\Sigma W} = \frac{22540}{100} = 225.40$

There is a rise of (225.4 – 210.6) = 14.8 % in the cost of living of workers in the year 2015 than in the year 2014 with reference to the base year.

Thus, the percentage increase in the cost of living index number in the year 2015 is $\frac{14.8}{210.6} \times 100 = 7.03$ as compared to the year 2014. Hence, the rise of 5% in the wages of the year 2014 is not sufficient to maintain the same standard of living of the workers in the year 2015.

Illustration 20 : The following data are obtained from the budget inquiry of middle class families. State the change in the cost of living in the current year 2015 with respect to the base year 2014 by finding the index number. If the average monthly disposable income of a family is ₹ 30,000 during the year 2014 and their average monthly disposable income during the year 2015 is ₹ 35,000 then according to family budget index number, what should be the rise in the average monthly disposable income of the family to maintain the same standard of living of the base year ?

Group	Food	Clothing	Rent	Fuel	Miscellaneous
Weight	45	20	15	10	10
Percentage price relative of the group in year 2015	130	150	120	160	120

The weights W and the percentage price relatives I of the groups for the year 2015 are given here. Hence, we shall calculate the index number by family budget method.

Group	Food	Clothing	Rent	Fuel	Miscellaneous	**Total**
Percentage price relative I	130	150	120	160	120	
Weight W	45	20	15	10	10	**100**
IW	5850	3000	1800	1600	1200	**13,450**

Index number by family budget method $I = \frac{\Sigma IW}{\Sigma W}$

$$= \frac{13450}{100}$$

$$= 134.5$$

Thus, it can be said that there is a rise of (134.5 – 100) = 34.5 % in the cost of living in the year 2015 as compared to the year 2014.

According to the family budget method index number of the year 2015, the average monthly disposable income to maintain the same standard of living as the base year

$$= \frac{\text{Index number of current year}}{\text{Index number of base year}} \times \text{income of base year}$$

$$= \frac{134.50}{100} \times 30{,}000$$

$$= ₹\ 40{,}350$$

The required increase in the average monthly disposable income to maintain the standard of living of the family = ₹ 40,350 – ₹ 35,000 = ₹ 5350

1.7.2 Uses and limitations of cost of living index number

This index number is prepared by studying the changes in the cost of living of the people of different classes. Thus, the cost of living index number is used for different objectives as follows :

(1) The changes in the purchasing power of money of a class of people can be measured using the corresponding cost of living index number. If the increase in the price of item is higher than the increase in the income, there is a decrease in the real income of the earning members and subsequently their purchasing power decreases. Thus, the cost of living index number is useful to find the actual purchasing power of money and the real income. The purchasing power of money and real income can be obtained by the following formulae.

(i) Purchasing power of money $= \frac{1}{\text{Cost of living index number}} \times 100$

(ii) Real income $= \frac{\text{Income}}{\text{Cost of living index number}} \times 100$

(2) The cost of living index number for each class shows the real economic condition of the respective class. Hence, this measure is used to suggest the changes in the wage, dearness allowance, bonus, etc. paid to the people of that class.

(3) This index number measures the effect of retail prices on the cost of living of people. Hence, this index number guides the government regarding the items to be controlled under special acts and the items to be kept for free trade.

(4) The cost of living index number helps the government as an indicator to frame the tax policies, policies regarding issues like price-regulation and fare-regulation. Moreover, it is possible to know how the living of people of different classes gets affected by imposing tax on certain items and the tax policy can be planned accordingly.

(5) The government agencies and public institutions use it as a base to determine the necessity of special facilities to elevate the standard of living of people of different classes.

Limitations of cost of living index number :

(1) It is not possible to construct one common cost of living index number for all sections of the society.

(2) The cost of living index number obtained for a certain class of people in a certain region cannot be used for some other region even for the same class of people.

(3) The cost of living index number shows the average percentage changes in the cost of living of a certain class. Thus, it is not possible to measure the changes in the cost of living of an individual.

(4) It is necessary to construct separate index numbers for different classes of people as well as different regions.

(5) The expenditure of the people of any class depends upon the size of the family, life style, liking, habits, etc. There is no uniformity in the expenditure of all the families of the same class.

(6) Its calculation is based on the assumption that the life style of the families does not change in the current year as compared to the base year. In reality, the liking, habits and choices of people change with time. Thus, it is necessary to conduct family budget inquiry at regular intervals of time and change the items and their weights.

EXERCISE 1.4

1. The following data are obtained from the family budget inquiry of middle class people. State the change in the cost of living in the year 2015 with respect to the year 2013 by finding the index number. If the average monthly disposable income of a family in the year 2013 is ₹ 15,000 then obtain the estimate of the necessary average monthly disposable income in the year 2015.

Group	Food	Fuel-Electricity	Rent	Clothing	Miscellaneous
Weight	45	15	10	20	10
Expenditure in 2013 (₹)	3000	1450	1500	600	1600
Expenditure in 2015 (₹)	3900	1850	2400	900	1920

2. Find the index number for the year 2014 by the method of family budget from the following data about prices and consumption of food items and interpret it.

Item	**Year 2010**		**Year 2014**
	Quantity	**Price (₹)**	**Price (₹)**
Wheat	60	15	18
Rice	40	32	40
Bajri	15	12	14
Tuver Dal	25	50	70

3. Compute the cost of living index number by the method of total expenditure from the following data.

Item	*A*	*B*	*C*	*D*	*E*
Unit	Quintal	20 kg	10 litre	dozen	meter
Quantity of year 2014	50 kg	18 kg	12 litre	20 pieces	14 meter
Price of year 2014 (₹)	1200	340	30	15	12
Price of year 2015 (₹)	1700	380	40	24	16

4. Compute the general index number for the production using the following data.

Item	Cotton Cloth	Grains	Sugar	Steel	Copper	Cement
Weight	15	23	15	25	10	12
Index number of production	220	225	190	215	198	220

5. The details of expenditure on clothing for the worker class of a region are as follows. Find the index number for clothing by the total expenditure and family budget method.

Item	Saree	Dhoti	Shirting	Other
Unit	Piece	Piece	Meter	Meter
Quantity in year 2010	5	8	20	15
Price in year 2010 (₹)	300	70	32.40	20.90
Price in year 2014 (₹)	400	100	38	23.80

*

Typical examples :

Illustration 21 : The prices of three items *A*, *B* and *C* among five items have increased in the year 2015 by 90 %, 120 % and 70 % respectively with respect to the year 2010, whereas the prices of two items *D* and *E* have decreased by 2 % and 5 % respectively. Item *A* is four times important than item *B* and item *C* is six times important than item *A*. The importance of items *D* and *E* is two and half times the importance of item *B*. Compute the general price index number of the year 2015 for all the five items.

The percentage increase and decrease in the prices of items is given here. Similarly, the weight W of the items are the numbers showing their relative importance.

Suppose the relative importance of item B is 1.

Then the importance of item A will be 4, importance of C will be 24 and that of D and E will be 2.5 each.

The general price index number will be calculated as follows :

Item	**Percentage increase (+) decrease (–)**	**Index number I = (100 + increase) = (100 – decrease)**	**Weight W**	**IW**
A	+ 90	100 + 90 = 190	4	760
B	+ 120	100 + 120 = 220	1	220
C	+ 70	100 + 70 = 170	24	4080
D	– 2	100 – 2 = 98	2.5	245
E	– 5	100 – 5 = 95	2.5	237.5
Total			**34**	**5542.5**

General price index number $= \frac{\Sigma IW}{\Sigma W}$

$= \frac{5542.5}{34}$

$= 163.0147$

$\simeq 163.01$

Thus, it can be said that there is a rise of (163.01 – 100) = 63.01 % in the prices in the current year 2015 as compared to the base year 2014.

Illustration 22 : The details of expenses on fuel for a group of workers of a region are as follows.

Item	Base year 2012		Year 2014
	Quantity	Price per unit (₹)	Price per unit (₹)
Coal	5 Kilogram	25	30
Kerosene	20 Litre	40	45
Wood	5 Kilogram	22	25
Match-box	10 Boxes	0.90	1

Prepare the index number of the group of fuel expenditure from these data. If the expenditure for food, clothing, house rent and miscellaneous groups in the year 2015 are 3, 2.5, 4.5 and 3.25 times respectively that of the year 2012, and if the expenditures on these groups are 42 %, 15 %, 10 % and 12 % respectively of the total expenditure then prepare the cost of living index number for the workers.

First of all, we shall prepare the group index number for fuel-expenditure from its details. We will take the base year 2012 and obtain the index number by total expenditure method.

Note : The method of family budget can also be used here for calculation.

Item	Year 2012		Year 2014	p_1q_0	p_0q_0
	q_0	p_0	p_1		
Coal	5	25	30	150	125
Kerosene	20	40	45	900	800
Wood	5	22	25	125	110
Match box	10	0.90	1	10	9
Total				**1185**	**1044**

Index number for fuel expenditure $= \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$

$= \frac{1185}{1044} \times 100$

$= 113.5057$

$\simeq 113.51$

The expenditure for the groups of food, clothing, house rent and miscellaneous are 3, 2.5, 4.5 and 3.25 times respectively than the base year. Hence, the index numbers of these four groups are (3 × 100) = 300; (2.5 × 100) = 250; (4.5 × 100) = 450 and (3.25 × 100) = 325 respectively. The index number of fuel category is obtained as 113.51. We will take the percentage expenditure for all the five groups as the weights for their corresponding index numbers to find the cost of living index number.

The expenditures for the groups of food, clothing, house rent and miscellaneous are given here as 42 %, 15 %, 10 % and 12 % respectively. These will be taken as their respective weights *W*. Total expenditure is 100 %. Hence, the weight for fuel expenditure index number will be 100 – (42 + 15 + 10 + 12) = 21 %.

The calculation of cost of living index number is as follows.

Group	Food	Clothing	House rent	Miscellaneous	Fuel	Total
Index number *I*	300	250	450	325	113.5	
Weight *W*	42	15	10	12	21	**100**
IW	12,600	3750	4500	3900	2383.71	**27,133.71**

$$\text{Cost of living index number} = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{27133.71}{100}$$

$$= 271.3371$$

$$\simeq 271.34$$

Thus, it can be said that there is a rise of (271.34 – 100) = 171.34 % in the cost of living in the current year as compared to the base year.

Illustration 23 : Find the real wages for the worker class of a city from the following data about their average monthly wage and the cost of living index number (base year 2001). Find the purchasing power of money in the year 2015 by taking the base year 2001 and state the importance of this answer.

Year	2010	2011	2012	2013	2014	2015
Average monthly wage (₹)	15,000	15,600	16,200	17,000	18, 000	20,000
Cost of living index number	192	203	228	268	270	287

The calculation of real wage using the wages and cost of living index numbers will be as follows.

$$\text{Real wage} = \frac{\text{Average monthly wage}}{\text{Cost of living index number}} \times 100$$

Year	Average monthly wage (₹)	Cost of living index number	Real wage (₹)
2010	15,000	192	$\frac{15000}{192} \times 100 = 7812.5$
2011	15,600	203	$\frac{15600}{203} \times 100 = 7684.73$
2012	16,200	228	$\frac{16200}{228} \times 100 = 7105.26$
2013	17,000	268	$\frac{17000}{268} \times 100 = 6343.28$
2014	18,000	270	$\frac{18000}{270} \times 100 = 6666.67$
2015	20,000	287	$\frac{20000}{287} \times 100 = 6968.64$

The purchasing power of money is the reciprocal of the cost of living index number of the current year with the respective base year.

$\therefore$ We can say that the purchasing power of money in the year 2015 with the base year 2001 $= \frac{100}{287} = 0.3484 \simeq 0.35$. Hence, it can be said that if the unit of money is rupee then the value of rupee in the year 2005 is 35 paise with respect to the base year 2001.

Thus, although the actual average monthly wage of the workers of this class in the year 2015 is more than the base year 2001, the real disposable wage in the year 2015 is only ₹ 6968.64 with respect to the base year.

Illustration 24 : Answer the following questions :

(1) The price of wheat was ₹ 1600 per quintal in the year 2014 and it was ₹ 1800 per quintal in the year 2015. Find the price index number of wheat for the year 2015 with the base year 2014 and interpret it.

Price index number of wheat for year 2015 $I = \frac{p_1}{p_0} \times 100$

$$= \frac{1800}{1600} \times 100$$

$$= 112.5$$

Thus, it can be said that there is a rise of $(112.5 - 100) = 12.5$ % in the price of wheat per quintal in the year 2015 as compared to the year 2014.

(2) The Laspeyre's index number is $\frac{8}{9}$ times the Fisher's index number. If the Fisher's index number is 180, find the Paasche's index number.

The Laspeyre's index number is $\frac{8}{9}$ times the Fisher's index number.

$$\therefore \ I_L = \frac{8}{9} \times I_F$$

$\therefore \quad I_L = \frac{8}{9} \times 180$

$I_L = 160$

Now, $I_F = \sqrt{I_L \times I_P}$

$180 = \sqrt{160 \times I_P}$

$(180)^2 = 160 \times I_P$

$\therefore \quad I_P = \frac{180 \times 180}{160} = 202.5$

(3) The production of an item in the year 2015 has increased by 3 times the production in the base year. Find the index number of production for the year 2015.

Consider the index number of the base year as 100. The production has increased by 3 times in the year 2015.

Production index number of year 2015 = index number of base year + increase in index number in current year

$= 100 + (3 \times 100)$

$= 100 + 300 = 400$

(4) If $\Sigma p_1q_0 : \Sigma p_0q_0 = 3:2$ and $\Sigma p_1q_1 : \Sigma p_0q_1 = 5:3$, find I_L, I_P and I_F.

$\frac{\Sigma p_1q_0}{\Sigma p_0q_0} = \frac{3}{2}$

$I_L = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$

$= \frac{3}{2} \times 100 = 150$

$\frac{\Sigma p_1q_1}{\Sigma p_0q_1} = \frac{5}{3}$

$I_P = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100$

$= \frac{5}{3} \times 100 = 166.67$

$I_F = \sqrt{I_L \times I_P} = \sqrt{150 \times 166.67} = \sqrt{25000.5} = 158.12$

(5) If the average monthly disposable income of middle class families in the year 2014 is ₹ 14,400 and if the cost of living index number for the year 2015 with the base year 2014 is 115 then estimate the average monthly disposable income of these families in the year 2015.

The cost of living index number of the middle class families for the year 2015 is 115 with the base year 2014. Hence, the index number has increased by $(115 - 100) = 15$ % as compared to the base year. Thus, there should be a 15 % rise in the average disposable income of the middle class families.

$\therefore$ Average monthly disposable income of the families $= 14400 + (14400 \times \frac{15}{100})$

$= 14400 + 2160 = 16560$

Hence, the average monthly disposable income of these families in the year 2015 should be ₹ 16,560.

(6) If the cost of living index number of the current year has increased to 180 from the base year index number 100 and if the average income of workers has increased from ₹ 6000 to ₹ 9000, is there an increase or decrease in the purchasing power of the workers ? How much is it ?

The index number has increased to 180 from 100 here which means that there is an increase of 80 %. Hence, the income should also increase by 80 %.

Average income $= 6000 + (6000 \times \frac{80}{100})$

$= 6000 + 4800 = ₹\ 10,800$

Hence, the average income of workers should be ₹ 10,800. But the average income of workers has increased to ₹ 9000. Thus, there is a decrease of (10800 – 9000) = ₹ 1800 in the average income with reference to the index number. Hence, it can be said that there is a decrease in their purchasing power.

(7) The wholesale price index numbers of the year 2015 and 2016 are found to be 150.2 and 165.7 respectively. Find the rate of inflation using index numbers of both the years.

The index number of the year 2015 is 150.2 and the index number of current year 2016 is 165.7.

We will use the following formula of the rate of inflation.

$$\text{Rate of inflation} = \frac{\left(\begin{array}{c}\text{Wholesale price index}\\ \text{number of current year}\end{array}\right) - \left(\begin{array}{c}\text{Wholesale price index}\\ \text{number of previous year}\end{array}\right)}{\text{Wholesale price index number of previous year}} \times 100$$

$$= \frac{165.7 - 150.2}{150.2} \times 100$$

$$= \frac{15.5}{150.2} \times 100$$

$= 10.3196$

$\simeq 10.32$

Thus, rate of inflation is 10.32 %.

(8) If the increase in the price relatives of three items are 250 %, 265 % and 300 % respectively and if the ratio of the importance of these items is 8 : 7 : 5, find the general price index number.

The percentage increase in the index numbers (price relatives) I and the relative importance W are given here.

We will calculate the index number.

Item	**Index Number (I)** (**Index Number of base year + increase**)	**Weight W**	**IW**
A	100 + 250 = 350	8	2800
B	100 + 265 = 365	7	2555
C	100 + 300 = 400	5	2000
Total		**20**	**7355**

General index number $= \frac{\Sigma IW}{\Sigma W} = \frac{7355}{20} = 367.75$

General index number = 367.75

Thus, there is a rise of (367.75 – 100) = 267.75 % in the prices in the current year as compared to the base year.

Summary

- The price, production, demand, supply, quantity, etc. of an item are called the variable for that item.
- The changes taking place in the values of the variable at two different time periods are compared by two methods : (1) Method of absolute measure (difference) and (2) Method of relative measure (ratio)
- The ratio of changes in the values of the variable at two different time periods is called relative change.
- The measure showing the percentage relative change in the prices of an item at different time periods is called price index number.
- The measure showing the percentage relative change in the quantities of an item at different time periods is called quantity index number.
- The average of the percentage change in the values of a variable associated with one or more items for the given period compared to its value in the fixed (base) period is called general index number for the group.
- When the price of an item is compared with the price of the same item in some specific (fixed) year of the past, then that specific year is called the base year.
- The year for which the price of an item is to be compared with the price of the base year is called the current year.
- Two methods of selecting base year : (1) Fixed base method (2) Chain base method.
- The expenditure p_0q_0 is assigned as weight to the price relative $\frac{p_1}{p_0}$ of the items. The formula of weighted average obtained by this method is called the formula of Laspeyre's index number.
- The expenditure p_0q_1 is assigned as weight to the price relative $\frac{p_1}{p_0}$ of the items. The formula of weighted average obtained by this method is called the formula of Paasche's index number.
- The geometric mean of Laspeyre's and Paasche's index numbers is called the Fisher's index number.
- The number showing the percentage of relative changes in the cost of living of the people of a certain section of the society in the current year (period) as compared to the base year (period) is called the cost of living index number.
- The points for the construction of index numbers : purpose, family budget inquiry, availability of prices of items, choice of base year, choice of average and choice of weight.
- In the construction of index number, the number associated with the selected items in proportion to their importance is called weight of that item.
- There are two types of weights : (i) Implicit weights (ii) Explicit weights
- Implicit weights : The weights are included in the selection of items and they cannot be expressed numerically. This indirect method of assigning weight is called implicit weight.
- Explicit weights : The weight to be assigned are determined in proportion to the importance of the item and can be expressed numerically. Such a weight is called explicit weight.
- There are two popular methods of assigning explicit weight : (i) Method of total expenditure (ii) Method of family budget.

List of Formulae

(1) Price relative $= \dfrac{\text{Price of current year (period)}}{\text{Price of base year (period)}}$

$$= \frac{p_1}{p_0}$$

(2) Quantity relative $= \dfrac{\text{Quantity of current year (period)}}{\text{Quantity of base year (period)}}$

$$= \frac{q_1}{q_0}$$

(3) Index number $I = \dfrac{\text{Value of variable in current year (period)}}{\text{Value of variable in base year (period)}} \times 100$

$$I = \frac{p_1}{p_0} \times 100$$

(4) Index number based on price relatives of n items $= \dfrac{\Sigma\left[\frac{p_1}{p_0}\right]}{n} \times 100$

(5) Fixed base index number $= \dfrac{\text{Value of variable in current year (period)}}{\text{Value of variable in base year (period)}} \times 100$

$$I = \frac{p_1}{p_0} \times 100$$

(6) Chain base index number $= \dfrac{\text{Value of variable in current year (period)}}{\text{Value of variable in preceding year (period)}} \times 100$

$$I = \frac{p_1}{p_0} \times 100$$

(7) Conversion of fixed base index number into chain base index number :

Chain base index number $= \dfrac{\text{Fixed base index number of current year}}{\text{Fixed base index number of preceding year}} \times 100$

(8) Conversion of chain base index number into fixed base index number :

Fixed base index number $= \dfrac{\text{(Chain base index number of current year)} \times \text{(Fixed base index number of preceding year)}}{100}$

Table 3.2. Electronic configuration of elements of second transition series in ground state and their oxidation states

Element	Atomic Number	Electronic configuration	Oxidation state
Y	39	$[Kr]4d^1 5s^2$	(+3)
Zr	40	$[Kr]4d^2 5s^2$	+2, +3, (+4)
Nb	41	$[Kr]4d^4 5s^1$	+3, (+5)
Mo	42	$[Kr]4d^5 5s^1$	+1, +3, +4, +5, (+6)
Tc	43	$[Kr]4d^5 5s^2$	(+4), +5, (+6)
Ru	44	$[Kr]4d^7 5s^1$	+2, (+3), +4, +6
Rh	45	$[Kr]4d^8 5s^1$	+2, (+3), +4
Pd	46	$[Kr]4d^{10} 5s^0$	(+2), +4
Ag	47	$[Kr]4d^{10} 5s^1$	(+1), +2, +3
Cd	48	$[Kr]4d^{10} 5s^2$	(+2)

Note: Stable Oxidation state is shown in parenthesis.

On the basis of the table 3.2, it can be said that in this transition series the electronic configuration of Pd, Ag and Cd in ground state have $4d^{10}$ that is 4d-orbital is completely filled with electrons; but in +2 of Pd and +2 of Ag, oxidation states, the electronic configurations, 4d-orbital is incompletely filled and so they are considered as transition elements but in the electronic configuration of Cd^{2+}, 4d-orbital is completely filled ($4d^{10}$). Hence **Cd is not considered as a transition element.** The electronic configuration of elements of third transition series in ground state and their oxidation states are shown in table 3.3.

Table 3.3. Electronic configuration of third transition elements in ground state and their oxidation states

Element	Atomic Number	Electronic configuration	Oxidation state
La	57	$[Xe]5d^1 6s^2$	(+3)
Hf	72	$[Xe]4f^{14} 5d^2 6s^2$	+2, +3, (+4)
Ta	73	$[Xe]4f^{14} 5d^3 6s^2$	+3, (+5)
W	74	$[Xe]4f^{14} 5d^4 6s^2$	+1, +4, +5, (+6)
Re	75	$[Xe]4f^{14} 5d^5 6s^2$	(+3), (+4), +5, (+6)
Os	76	$[Xe]4f^{14} 5d^6 6s^2$	(+4), +5, +6
Ir	77	$[Xe]4f^{14} 5d^7 6s^2$	(+3), (+4)
Pt	78	$[Xe]4f^{14} 5d^9 6s^1$	(+2), (+4)
Au	79	$[Xe]4f^{14} 5d^{10} 6s^1$	+1, (+3), +5
Hg	80	$[Xe]4f^{14} 5d^{10} 6s^2$	+1, (+2)

Note: Stable oxidation state is shown in parenthesis.

On the basis of table 3.3, it can be said that in the electronic configuration of Au and Hg in ground state there is $5d^{10}$ that is 5d-orbital is completely filled but in the electronic configuration of Au^{3+}, 5d-orbital is incompletely filled ($5d^8$) and so Au is considered transition element while in electronic configuration of Hg, 5d orbital is completely filled ($5d^{10}$) and so **Hg is not considered transition element.**

3.2.3 Occurrence of Transition Elements :

Definite transition metal is obtained from its definite mineral and a definite method is used for obtaining pure metal from the mineral. You have studied in unit 4 of semester III about the methods for the extraction of Cu, Fe and Zn from their minerals.

3.2.4 General Characteristics of Transition Elements :

- All the transition elements are metallic elements.
- These elements are hard and strong.
- Their melting points are high.
- These elements can form alloys with each other.
- Most of these elements dissolve in acid, but acid has no effect on certain noble elements.
- These elements possess various valencies.
- They possess property of malleability and ductility.
- They are good conductors of electricity and heat.
- Some of their ions possess paramagnetic property.

3.3 Periodic Trends in Properties of Elements of First Transition Series

(1) Metallic property : All the elements of first transition series possess metallic property. This point can be understood from the study of general characteristics of transition elements (see point 3.2.4)

(2) Atomic radii and Ionic radii : Generally in the periodic table, moving from left to right, there is decrease in atomic radii. This trend is observed in the transition elements, but this decrease in atomic radii is less. The tendency of ionic radii in transition elements is found to be similar to atomic radii. The atomic radii and ionic radii of first transition series elements are shown in tables 3.4 and 3.5 respectively.

Table 3.4 Atomic radii of first transition series elements

Element	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
Atomic Radius (pm)	144	132	122	117	117	117	116	115	117	125

Table 3.5 Ionic radii of first transition series elements

Element	Sc^{2+}	Ti^{2+}	V^{2+}	Cr^{2+}	Mn^{2+}	Fe^{2+}	Co^{2+}	Ni^{2+}	Cu^{2+}	Zn^{2+}
Ionic Radius (pm)	-	90	79	82	82	77	74	70	73	75

As shown in table 3.4, the atomic radius decreases from Sc to V while the atomic radii from Cr to Cu is almost similar. In addition to this, instead of decreasing the atomic radius of Zn is found to be increasing. The reason for this is that, as we move from left side to right side in period, the positive electric charge of nucleus increases and the entering electron is added to 3d-orbital. This electron, increases the shielding effect for attraction of electron present in 4s-orbital due to increase in positive electric charge in the nucleus. As a result, the electrons present in 4s-orbital are not attracted more towards the nucleus (Relative to neighbouring transition element). Thus, the orbit does not contract hence the atomic radii remain same. 3d-orbital of Zn atom is completely filled. Hence, it decreases the attraction towards electron of 4s-orbital of positive electric charge of nucleus due to its shielding effect. Also repulsion between electron-electron in 3d orbital increases more than the value of attraction of electrons of 4s-orbital and nucleus. Hence, there is expansion of orbit; so the value of atomic radius of Zn is seen to be increasing.

(3) Ionisation enthalpy : Moving from left side to right side in first transition series elements, the nuclear electric charge increases and so the value of ionization enthalpy increases but the increase in this value, is not equal to the increase in enthalpy in the elements of the main group. Thus, there is not much change in the first ionization enthalpies of two neighbouring transition elements (table 3.6). The values of first, second and third ionization enthalpies are shown in table 3.6.

Table 3.6 First, second and third ionization enthalpies (kJ mol^{-1}) of first transition series elements

Element	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
$\Delta_i H_1$	631	656	650	653	717	762	758	736	745	906
$\Delta_i H_2$	1235	1309	1414	1592	1509	1561	1644	1752	1958	1734
$\Delta_i H_3$	2393	2657	2833	2990	3260	2962	3243	3402	3556	3829

As shown in table 3.6, there is not much difference seen in the values of second ionization enthalpies like the changes in the first ionization enthalpies of first transition series elements. But chromium and copper are found to be exceptions. The values of second ionization enthalpy of these two elements are more than those of their neighbouring elements, because both these elements attain the electronic configuration Cr^+ : $[Ar]3d^5$ and Cu^+ : $[Ar]3d^{10}$ after loss of one-one electron from these two elements. From Cr and Cu, the second electron is removed from half filled or completely fill 3d orbitals respectively having more stability. It is natural that more energy will be required for removal of such electrons.

(4) Electrode potential : Thermodynamic stability of transition metals depends on their magnitude of ionization enthalpy. The compounds of metals having less ionization enthalpy are more stable, while the stability of compounds in solution is more dependent on relative electrode potentials.

The value of electrode potential is determined on the basis of the sum of the enthalpy changes (ΔH_T) of different reactions. The stabilities of different oxidation states of transition metal ions are determined on the basis of the electrode potentials. More negative the values of standard reduction potientials more will be the stabillity of ions in aqueous medium.

3.4 Characteristic Properties of Elements of First Transition Series

(1) Colour : Most of the ionic and covalent compounds of transition elements are coloured. It is due to the presence of incompletely filled d-orbitals. When visible light falls on transition metal ions, they absorb light of definite wavelength and emit the remaining light. Our eye catches the colour of this emitted light as the colour of those ions. Hence, ions appear coloured. For example, when visible light passes through the aqueous solution of $[Ni(H_2O)_6]^{2+}$, green colour is emitted and other colours are absorbed. Hence, the colour of this aqueous solution appears to be green. In the similar, when visible light passes through aqueous solution of $[Co(NH_3)_6]^{3+}$, then red and green colours are emitted and other colours are absorbed. Because of the mixing of the emitted red and green colour, the colour of this solution appears to be yellowish orange. During absorption of light the electrons in d-orbital of ions receive energy and go to the d-orbital having higher energy. This transition of electrons is called **d-d transition.** It is necessary to note here, that generally the energy of all five d-orbitals is equal but according to crystal field theory, there is splitting of d-orbitals in different energy levels. The splitting of d-orbitals of different energy levels in transition compounds having tetrahedral and octahedral geometry are shown in figure 3.1.

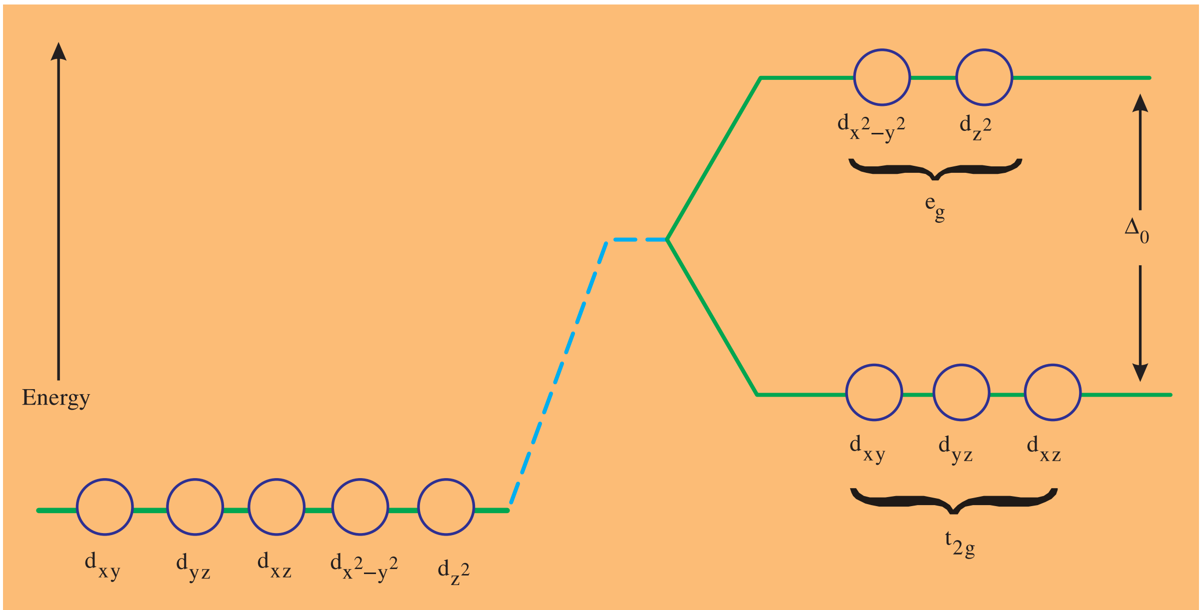


Fig.3.1(a) Splitting of d-orbitals in octahedral complex compounds

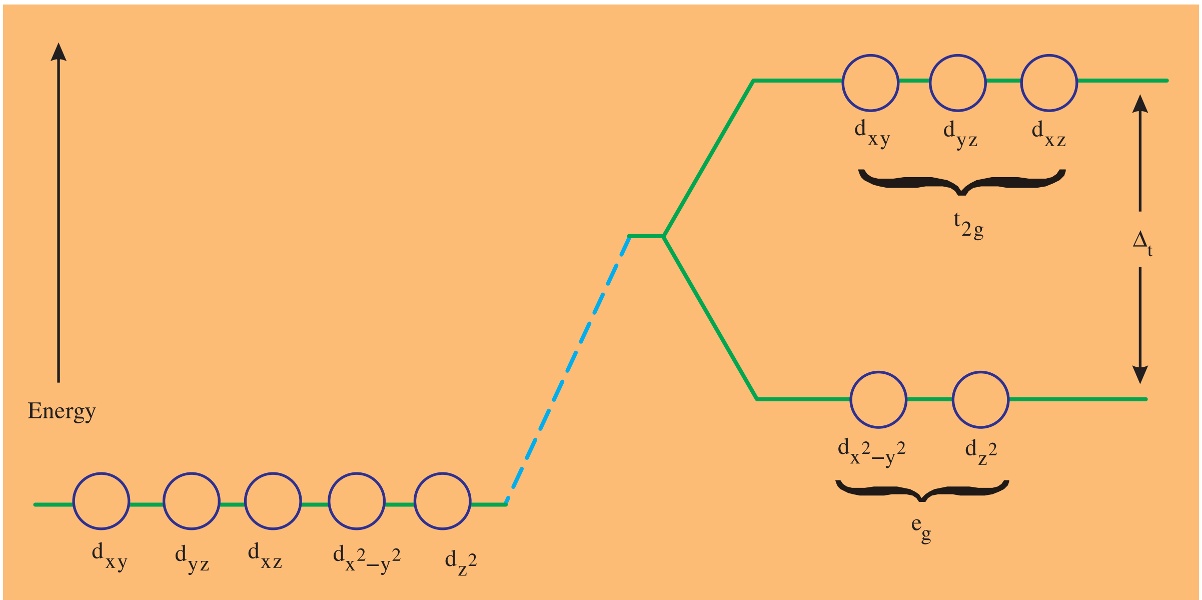


Fig.3.1(b) Splitting of d-orbitals in tetrahedral complex compounds

The colours of some metal ions are mentioned in table 3.7

Table 3.7 Colours of some of the hydrated metal ions of first transition series

Metal ions	Colour
Cu^{2+}, Cr^{2+}, Co^{3+}, V^{4+}	Blue
Ni^{2+}, Fe^{2+}, V^{3+}	Green
V^{2+}, Cr^{3+}, Mn^{3+}	Violet
Co^{2+}, Mn^{2+}	Pink
Fe^{3+}	Yellow
Ti^{3+}	Purple
Zn^{2+}, Sc^{3+}, Ti^{4+}, Cu^{+}	Colourless

Student friends, the verfication of colour of certain metal ions mentioned in table 3.7 can be done in chemistry laboratory by observing easily the colours from the chemicals available- $CuSO_4 \cdot 5H_2O$, $NiCl_2 \cdot 6H_2O$, $FeSO_4 \cdot 7H_2O$, $Co(NO_3)_2 \cdot 6H_2O$, $MnCl_2 \cdot 4H_2O$, $FeCl_3$ and $ZnCl_2$.

(2) Catalytic property : Transition metals and their certain compounds increase the rate of chemical reaction. Hence, they are useful catalysts. These substances used as catalyst are in solid state. The random and pointed peak points on the surface at the ends of the edges of their particles are responsible centres for catalysis work. The molecule-atom in the depth in the matter of solid substance are surrounded by similar other atoms. Hence, the magnetic field of their unpaired electrons is destroyed by the effect on each other, while atoms of the ends of the edges or the peak points, get affected by the magnetic field which are active centres for catalysis. The molecules of reactant are attracted by the magnetic field and so the required activation energy of the reaction decreases. As a result, the reaction rate increases viz. Ni in hydrogenation of vegetable oil, Fe in industrial production of NH_3 by Haber process, V_2O_5 catalyst is used in obtaining SO_3 from SO_2 in contact process for manufacturing of H_2SO_4. You have studied in detail the point on catalysis in unit 2 : Surface Chemistry.

(3) Magnetic property : When a substance is placed in the magnetic field, then they possess two types of magnetic properties (i) Paramagnetism and (ii) Diamagnetism.

The molecule, atom or ion of a substances in which there are unpaired electrons, such substances possess paramagnetism, so they are called **paramagnetic substances;** while molecule, atom or ion in which all the electrons are paired, possesses diamagnetism. Hence, they are called **diamagnetic substances.** In the electronic configuration of transition elements, the (n–1)d-orbitals are incompletely filled; hence they possess unpaired electrons and so the atoms of these elements are paramagnetic. Because of paramagnetism they possess magnetic moment. Magnetic moment is produced due to rotation of unpaired electron on its axis and orbital rotation. In transition metal ions the unpaired electrons are present in the outermost orbit. Hence, in such cases axial rotation contribution is much more important than the orbit contribution. The value of magnetic moment based on only rotation on axis, can be calculated with the help of following formula :

$$\mu = \sqrt{n\,(n+2)}\ \text{BM}$$

where μ = Magnetic moment, n = Number of unpaired electrons, BM = Bohr Magneton (unit)

The value of magnetic moment increases with increase in number of unpaired electrons. Thus, through the measurement of magnetic moment we come to know about the number of unpaired electrons in atom, molecule or ion. Maximum five unpaired electrons can be in d-orbitals, so the theoretical values of magnetic moments are shown in table 3.8.

Table 3.8 Theoretical Value of Magnetic Moment

Number of unpaired electrons (n)	Magnetic moment μ (BM)
1	1.73
2	2.83
3	3.87
4	4.90
5	5.92

The magnetic moments of ions of transion elements or their compounds, are based on the axial rotation; so sometimes their values obtained experimentally are slightly more or less than the theoretical values. The reason for this is the rotation-orbital combination. This type of combination being directional, the value obtained is slightly more or less i.e. it depends on how these directional combinations occur. Earlier you have studied about diamagnetic or paramagnetic properties in semester III in unit 1 on Solid State.

Example : 1 Calculate magnetic moment of Co^{2+}(aq)

Solution :

		3d					4s
Co (Z = 27) : [Ar]	↑↓	↑↓	↑	↑	↑		↑↓
Co^{2+} : [Ar]	↑↓	↑↓	↑	↑	↑		

Here number of unpaired electrons n=3

Now, magnetic moment $\mu = \sqrt{n(n+2)}$ BM

$= \sqrt{3(3+2)}$ BM

$= \sqrt{15}$ BM = 3.87 BM

Thus, the value of magnetic moment of Co^{2+}(aq) will be 3.87 BM

(4) Capacity of transition metal ions to form complex compounds : Transition metal ions combine with one or more anion or neutral molecule (ligand) and form complex species by co-ordinate covalent bond having definite characteristics which are called complex compounds. In depth, study of such **complex compounds** will be done in Unit 4 : complex compounds. The capacity of transition elements to form complex compounds is more than that with other elements; the reason of which is the below mentioned characteristics of transition metal ions.

- The size of transition metal ions is small.
- The nuclear electric charge and ionic electric charge of transition metal ions is comparatively more.
- The electronic configurations of transition metal ions is favourable for formation of complexes. In these metal ions, d-orbitals are vacant and so the electron pairs coming from the ligand can be accommodated.

- As very less difference is there between energy values of 3d, 4s, 4p or 4d orbitals, hybridization of different types can be possible with these orbitals. Thus, the different hybrid orbitals produced are helpful in formation of complex compounds.
- Due to formation of various types of hybridization and because of co-ordinate covalent bonds are directional, different types of geometrical shapes containing complex compound can be formed.
- Transition metal ions exhibit various oxidation states. Thus, varieties of complex compounds can be formed.

3.5 Interstitial Compounds

The atoms in solid state of transition metals are arranged in definite crystal structure. There are definite voids between atoms in such an arrangement. Hence, non-metallic atoms of smaller size viz. H, C, N and B can be easily arranged in the voids of crystal structure. The compounds formed in such a manner are called **interstitial compounds;** in which chemical bond is not formed between non-metal elements of smaller size arranged in the voids and the metal atom, hence the proportion of components in such compound is not definite. So interstitial compounds are not detinite. In fact interstitial compounds are nonstoichiometric or non-proportionate compounds, e.g., $TiH_{1.7}$, $VH_{0.56}$ etc. On the basis of the non-metal arranged in the interstitial void, they can be classified as hydrides, carbides, nitrides and borides, because elements like hydrogen, carbon, nitrogen and boron can be arranged in interstitial positions.

A notable change is observed in the characteristic properties of metals due to presence of non-metal elements like H,C,N and B in the voids of crystal structure. The small size of non-metal atoms present here in the void attract free electrons of metallic bond, hence free electrons of metallic bond are localised and so the strength of the bond increases. Hence, the properties of the metals like hardness, melting point resistance to wear, resistance to corrosion etc. are notably increased. Therefore, interstitial compounds are used in preparation of tools, machinery parts, vehicles etc. This type of compounds have not definite molecular formula. VN, Fe_3N, Fe_3C, Cr_3C_2, Mn_3C, TiC, VB, CrB_2 etc. are interstitial compounds.

3.6 Alloys

The characteristic properties like hardness, conductance, malleability, resistance to corrosion are essential for machineries, tools, vehicles and vessels of house-hold usages. There is no combination of all these properties and is not together found in any pure metal. Hence, instead of using pure metals in practice, alloys prepared from two or more metal-elements are used frequently..

Scientists **Hume** and **Rothery** presented rules as follows to obtain alloys of combination of useful properties :

(1) The atomic size of two metals forming the alloy must be the same. There must not be more than 15 % difference in their atomic radii.

(2) The chemical properties of the metals used for preparation of alloys must be same, that is, their electronic configurations of valence shell must be the same.

(3) The crystal structures of pure metallic elements used for alloys must be similar.

22 carat gold ornaments is the best example of alloy. It is the alloy of Au and Cu. The difference between atomic size of Au (atomic size = 134 pm) and Cu (atomic size = 117 pm) is about 14.5 %. Both of them possess cubic close pack structure. Both of them being members of group-11 the electronic configuration of their valence shells is same. Thus best alloy can be prepared according to laws

suggested by Hume and Rothery. The difference in atomic sizes between Cr, Mn, Fe, Co, Ni, Cu metals, is less than 2%. There is much less difference in the electronic configuration of valence shell of these elements. Hence, these elements form number of alloys having different proportions which are very useful in practice. Certain important alloys, their components, properties and uses are shown in table 3.9.

Table 3.9 Alloys, their components, properties and uses

Alloy	Components	Properties	Uses
Stainless steel	Fe (70%), Cr (20%) Ni (10%)	No effect of air, water and alkali and does not get rusted	In preparation of utensils, blades, surgical instruments.
Brass	Cu (70%), Zn (30%)	Ductile, hard, corrosion resistant and can be shaped easily	In preparation of cooking vessels, parts of machine and musical instruments.
Bronze	Cu (90%), Sn (10%)	Very strong and possessing more corrosion resistance	In preparation of statues, currency coins and medals.
Nitinol	Ti (45%), Ni (55%)	Light in weight and strong and resists corrosion. It has marvellous property of memory.	Rivetting and useful in space research.
Cupronickel	Cu (75-85%), Ni (15-25%)	Strong and corrosion resistant.	In preparation of currency coins.
	Cu (50-55%), Ni (45-50%)	Electrical resistance is more.	In preparation of electric resistant wires.
German-silver	Ni (40-50%), Zn (25-30%) Cu (25-30%)	Possesses shining as silver	In preparation of household vessels,art models and resistant wires.
Nichrome	Ni (60%), Cr (40%)	Electrical resistance is very high	As electric resistant wire in electric furnaces and electric heaters.

In addition, amalgam with mercury - alloy is also well known. In this alloy there are Hg (50%), Ag (35%), Sn (12%), Cu (3%) and Zn (0.2%). This alloy is, used in filling the cavity in the tooth. When this alloy is to be filled in cavity in the tooth, all these metals are mixed, some time before filling. This alloy is soft. By the time the dentist can fill in the cavity of tooth, it remains soft so that it becomes convenient for dentist to fill in the cavity. After filling in this alloy in cavity, it becomes hard and it does not expand.

3.7 Some Important Compounds of 3d-Transition Elements

Many compounds of 3d-transition elements are familiar. We shall study here, preparation, properties and uses of only potassium dichromate $(K_2Cr_2O_7)$ and potassium permanganate $(KMnO_4)$.

(1) Potassium dichromate $\left(K_2Cr_2O_7\right)$:

Perparation : Sodium chromate (Na_2CrO_4) and ferric oxide (Fe_2O_3) are formed by fusion of chromite mineral ($FeCr_2O_4$) with sodium carbonate and quick lime in presence of air. After the reaction, the roasted mass is extracted with water when Na_2CrO_4 is completely dissolved while Fe_2O_3 is left insoluble in presence of air.

$$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$

Filtering the yellow coloured solution of sodium chromate, it is acidified with H_2SO_4; so that sodium dichromate $\left(Na_2Cr_2O_7\right)$ is formed, which reacts with potassium chloride and forms potassium dichromate and sodium chloride. $K_2Cr_2O_7$ being much less soluble than NaCl it crystallizes out on cooling.

$$2Na_2CrO_4 + 2H^+ \rightarrow Na_2Cr_2O_7 + 2Na^+ + H_2O$$

$$Na_2Cr_2O_7 + 2KCl \rightarrow K_2Cr_2O_7 + 2NaCl$$

Properties :

- Potassium dichromate is orange coloured crystalline substance.
- It is soluble in water.
- It acts as a strong oxidizing agent in acidic medium.
- When an alkali is added to an orange coloured solution of potassium dicromate a yellow coloured solution results due to the formation of potassium chromate and on acidifying it, the colour again changes to orange due to the reformation of potassium dichromate.

Chromate and dichromate ions are interconvertible in aqueous solution; which depends on the pH of the aqueous solution; because the oxidation state of chromium in chromate and dichromate is the same.

$$\underset{\text{Yellow colour}}{2CrO_4^{2-}} + 2H^+ \rightarrow \underset{\text{Orange colour}}{Cr_2O_7^{2-}} + H_2O$$

$$\underset{\text{Orange colour}}{Cr_2O_7^{2-}} + 2OH^- \rightarrow \underset{\text{Yellow colour}}{2CrO_4^{2-}} + H_2O$$

Uses :

- Potassium dichromate is mainly useful in leather industry and formation of azo compounds.
- As oxidizing agent used in synthesis of organic compounds, used as a reagent in chemical oxygen demand (COD) measurement in polluted water. In addition, the mixture of potassium dichromate with concentrated sulphuric acid which is known as chromic acid, is used for cleaning of glasswares in laboratory. It is corrosion inhibitant.
- It is used as titrant in redox titrations to determine the amount of metal ions like iron (II).

(2) Potassium permanganate $\left(KMnO_4\right)$:

Preparation : Dark green coloured potassium manganate (K_2MnO_4) is formed by fusion of manganese dioxide (MnO_2) with KOH in presence of air or oxidizing agent like KNO_3. Potassium permanganate is formed when sulphuric acid is added to this solution and made acidic.

$$2MnO_2 + 4KOH + O_2 \rightarrow \underset{\text{(Green colour)}}{2K_2MnO_4} + 2H_2O$$

$$3K_2MnO_4 + 2H_2SO_4 \rightarrow 2KMnO_4 + 2K_2SO_4 + MnO_2 + 2H_2O$$

Properties :

- Potassium permanganate is dark purple coloured crystalline substance.
- It is soluble in water.
- It acts as oxidising agent in acidic, basic and neutral medium.

Uses :

- Potassium permanganate is used as strong oxdising agent in synthesis of organic compounds.
- It is used as bleaching agent for cotton cloth, silk and wood and in textile industries.
- An aqueous solution of potassium permanganate is used for gargling to keep mouth germfree as it is antiseptic.
- It is useful as titrant in redox titrations to know the proportion of metal ions like iron (II) and organic compounds like oxalic acid.

3.8 Applications of d-Block Elements

- The alloys-stainless steel, brass, bronze, nitinol, cupronickel, german silver and nichrome prepared from elements of d-block elements are used for preparation of household utensils, currency coins, statues, and machinery.
- Some elements and compounds of these elements act as catalyst in chemical reaction e.g., Ni in hydrogenation of vegetable oil, Fe in Haber process of industrial production of NH_3, V_2O_5 catalyst for obtaining SO_3 from SO_2 in contact process for production of H_2SO_4.
- Alloy like mercury-amalgam is useful for filling in tooth cavity.
- Alloy of gold and copper is more appropriate for preparation of ornaments.
- Mercury is used in thermometer.
- The use of inert metal like platinum is made as electrodes in the experiments of electro-chemistry.
- Some compounds like $KMnO_4$ and $K_2Cr_2O_7$, of d-block elements act as strong oxidizing agents in synthesis of organic compounds. They are used as titrant in redox titrations.
- MnO_2 is used in dry cell.
- To prevent water pipes and the roof on the house from corrosion they are changed in to galvanized form with the help of zinc metal.

3.9 Innertransition Elements (Elements of f-Block)

The f-block elements are distributed into two series : (A) Lanthanide series and (B) Actinide series. The series of fourteen elements immediately after lanthanum- Ce(Z = 58) to Lu (Z = 71) is called **lanthanide series.** The elements of this series are known as **lanthanoids.** Lanthanoids are expressed by general symbol Ln. As lanthanum has more similarity with lanthanoids, it is included in the lanthanide

Downloaded from https:// www.studiestoday.com

series, during discussion. In the periodic table, series of fourteen elements immediately after actinium-Th (Z = 90) to Lr (Z = 103) is called **actinide series.** The elements of this series are known as **actinoids.** As actinium possesses more similarity with actinoids, it is included in the discussion of actinide series. **The general electronic structure of outermost electrons of elements of f-block is $(n-2)f^{0-14}(n-1)d^{0-1}ns^2$.**

3.10 Lanthanide Series

To understand the lanthanide series, we shall study here, its electronic configuration, oxidation state, atomic size, chemical reactivity and lanthanide contraction.

3.10.1 Electronic Configuration and Oxidation State :

In the electronic configuration of elements of lanthanide series, $6s^2$ is common in all the elements but the electrons in 4f-orbital keep changing. **All the lanthanoids and lanthanum element possess stable oxidation state + 3.** The electronic configuration and oxidation states of lanthanum and lanthanoids are shown in table 3.10.

Table 3.10 Electronic configuration and oxidation states of lanthanum and lanthanoids

Element	Atomic Number	Electronic configuration	Oxidation state
La	57	$[Xe]5d^16s^2$	(+3)
Ce	58	$[Xe]4f^15d^16s^2$	(+3), +4
Pr	59	$[Xe]4f^36s^2$	(+3)
Nd	60	$[Xe]4f^46s^2$	+2, (+3)
Pm	61	$[Xe]4f^56s^2$	(+3)
Sm	62	$[Xe]4f^66s^2$	+2, (+3)
Eu	63	$[Xe]4f^76s^2$	+2, (+3)
Gd	64	$[Xe]4f^75d^16s^2$	(+3)
Tb	65	$[Xe]4f^96s^2$	(+3), +4
Dy	66	$[Xe]4f^{10}6s^2$	(+3), +4
Ho	67	$[Xe]4f^{11}6s^2$	(+3)
Er	68	$[Xe]4f^{12}6s^2$	(+3)
Tm	69	$[Xe]4f^{13}6s^2$	+2, (+3)
Yb	70	$[Xe]4f^{14}6s^2$	+2, (+3)
Lu	71	$[Xe]4f^{14}5d^16s^2$	(+3)

Note : Stable oxidation state is shown in parenthesis

It is apparent from the study of electronic configuration shown in table 3.10, that only in the lanthanoids like Ce, Gd and Lu electrons are filled in 5d orbital. In Gd, because of half filled orbital like $4f^7$, stability may be obtained and so newly added electron enters into 5d orbital, while in Lu, 4f orbital being completely filled, the added new electron enters into the 5d orbital. The filling of electron of Ce

Downloaded from https:// www.studiestoday.com

in 5d-orbital is accepted at present as an exception. Thus, **the general electronic configuration of lanthanoids is [Xe]$4f^{1-14}5d^{0-1}6s^2$.** Amongst lanthanoids, promethium (Pm) is a radioactive element.

3.10.2 Atomic Size and Lanthanide Contraction :

Like the elements in any period of periodic table, in elements of lanthanide series going from cerium(Ce) to lutetium (Lu), the atomic radii and ionic radii go on decreasing, In the elements of this series, with the increase in atomic number the addition of new electron is in inner orbit 4f(n = 4) instead of last orbit (n = 6). Hence, with the increase in positive electric charge, the electrons in 4f-orbital possess more attraction towards it. Hence, there is contraction of atom. i.e. atomic radius decreases. This contraction occurring in lanthanide elements, is called **lanthanide contraction.** The effect of this lanthanide contraction is observed on the atomic radii of the elements after lanthanide series. The atomic radii of some of the elements of third transition series after lanthanide series are equal to the atomic radii of some elements of second transition series earlier to this, which can be understood from table 3.11.

Table 3.11 Atomic radii (pm) of elements of second and third transition series

Second Transition Series	Y 162	Zr 145	Nb 134	Mo 129	Tc -	Ru 124	Rh 125	Pd 128
Third Transition Series	Lanthanoids	Hf 144	Ta 134	W 130	Re 128	Os 126	Ir 126	Pt 129

3.10.3 Chemical Reactivity :

As lanthanoids possess (+3) oxidation state they form hydroxides of the type $Ln(OH)_3$. These hydroxides are less basic than $Ca(OH)_2$ but more basic than $Al(OH)_3$. The carbonates and nitrates of these elements decompose faster into their oxide by heating. The size of ion decreases on going from Ce^{3+} to Lu^{3+}, hence their basicity decreases, that is $Ce(OH)_3$ is most basic and $Lu(OH)_3$ is least basic. These elements possess similarity in physical and chemical properties so their separation is carried out on the basis of the difference in their basicity. Ln_2O_3 type oxides of these elements are ionic and basic. The property of basicity decreases with ionic size. Some ions of these elements possess paramagnetic property because of unpaired electrons in f-orbital. Their certain ions are coloured and give coloured solution. The general chemical reactions of lanthanoids are shown below :

$$Ln \xrightarrow{H_2O} Ln(OH)_3 + H_2$$

$$Ln \xrightarrow{C / 2773\ K} LnC_2$$

$$Ln \xrightarrow[\Delta]{N} LnN$$

$$Ln \xrightarrow[\Delta]{S} Ln_2S_3$$

$$Ln \xrightarrow{\text{Combusting in } O_2} Ln_2O_3$$

$$Ln \xrightarrow{\text{Acid}} Ln^{3+} + H_2$$

$$Ln \xrightarrow{X_2} LnX_3$$

3.11 Actinide Series

We shall study electronic configuration and oxidation states to understand the actinide series.

3.11.1 Electronic Configuration and Oxidation State :

In the electronic configuration of elements of actinide series, $7s^2$ is common in all the elements but electrons keep on changing in 5f orbital. Irregularity is found more in the electronic configuration of actinoids. Actinoids possess more than one oxidation states. The electronic configuration and the oxidation states of actinium and actinoids are shown in table 3.12.

Table 3.12 Electronic configuration and oxidation states of actinium and actinoids

Element	Atomic Number	Electronic configuration	Oxidation state
Ac	89	$[Rn]6d^1 7s^2$	(+3)
Th	90	$[Rn]6d^2 7s^2$	(+4)
Pa	91	$[Rn]5f^2 6d^1 7s^2$	+3, +4, (+5)
U	92	$[Rn]5f^3 6d^1 7s^2$	+3, +4, +5, (+6)
Np	93	$[Rn]5f^4 6d^1 7s^2$	+3, +4, (+5), +6, +7
Pu	94	$[Rn]5f^6 7s^2$	+3, (+4), +5, +6, +7
Am	95	$[Rn]5f^7 7s^2$	(+3), +4, +5, +6
Cm	96	$[Rn]5f^7 6d^1 7s^2$	(+3), +4
Bk	97	$[Rn]5f^9 7s^2$	(+3), +4
Cf	98	$[Rn]5f^{10} 7s^2$	(+3)
Es	99	$[Rn]5f^{11} 7s^2$	(+3)
Fm	100	$[Rn]5f^{12} 7s^2$	(+3)
Md	101	$[Rn]5f^{13} 7s^2$	(+3)
No	102	$[Rn]5f^{14} 7s^2$	(+2), +3
Lr	103	$[Rn]5f^{14} 6d^1 7s^2$	(+3)

Note : Stable oxidation state is shown in parenthesis

In the electronic configuration of elements Th to Np as shown in table 3.12, irregularity is found to be more in comparison to other elements which is at present accepted as an exception, while in Cm and Lr, the half filled 5f-orbital and completely filled 5f-orbitals respectively, to attain stability, the new added electron is filled in 6d-orbital.

Thus **general electronic configuration is $[Rn]5f^{0-14}6d^{0-2}7s^2$.** All the actinoids are radioactive.

3.12 Comparison of Actinoids with Lanthanoids

- Actinoids are like silver in appearance. More irregularity is observed in the metallic radii in actinoids than lanthanoids. Hence, diversity is found in the structures of actinoids.
- As the outermost orbit in actinoids is far from the nucleus in comparison to lanthanoids, their electron can be easily removed. Hence, the values of ionization enthalpy of actinoids are less than the values of ionisation enthalpies of lanthanoids.

- The stable oxidation state of all the lathanoids is (+3). In actinoids, oxidation states (+2) to (+6) are seen.
- In lanthanoids only promethium is radioactive but all the actinoids are radioactive.

3.13 Applications of f-Block Elements

- Pyrophoric Misch metal (50 % Ce +40 % La + 7 % Fe + 3 % other metals) is used as reducing agent and as stones in gas lighters.
- CeO_2 is useful in pigments.
- Ceric compounds are used as oxidizing agent in volumetric analysis.
- Oxides of lanthanoids are useful in preparation of optical glass of camera having high refractive index.
- Gadolinium sulphate is used to produce very low temperature by magnetic effect.
- Metals like, uranium, plutonium, and thorium are useful in production of atomic energy. Electrical energy can be obtained from nuclear energy.

SUMMARY

Position in periodic table	Block
Groups 1 to 2	s-Block
Groups 13 to 18	p-Block
Groups 3 to 12	d-Block
Two horizontal rows at the bottom of the periodic table	f-Block

- d-block elements are in periods 4 to 7.
- f-block elements are in periods 6 and 7.

Elements of d-block (Transition metal elements)

- The elements which in their ground state or any one of their oxidation states, d-orbital is incompletely filled are called transition elements.
- Amongst the d-block elements Zn, Cd and Hg do not act as transition elements.
- All the transition elements are metallic elements.
- In the first transition series, atomic radii decrease from Sc to V, while atomic radius remains same in elements Cr to Cu and the atomic radius of Zn is found increasing instead of decreasing.
- Not much difference is observed in first and second ionization enthalpies of two neighbouring transition elements but the value of second ionization enthalpy of Cr and Cu are more than those of their neighbouring elements.
- Most of the ionic and covalent compounds of transition elements are coloured.

Downloaded from https:// www.studiestoday.com

- Compounds of transition elements act as catalysts in certain chemical reactions.
- The magnetic moment of transition element compounds,

  $\mu = \sqrt{n(n+2)}$ where μ = magnetic moment. n = number of unpaired element

  The unit of magnetic moment is BM (Bohr Magneton)
- The capacity to form complex compounds is much more than other elements because of definite characteristics of transition elements.
- In the formation of crystals of transition metals, the voids are there in which non-metal elements (H, C, N, B) arrange and form interstitial compounds.
- Scientists Hume and Rothery suggested the rules for the alloys and accordingly alloys having useful properties are obtained from transition metal elements.
- The compounds of transition element- $KMnO_4$ and $K_2Cr_2O_7$ are very useful in laboratory and in synthesis of organic compounds.

f-Block elements (Innertransition elements)

- f-Block elements are divided in to (1) Lanthanide series and (2) Actinide series.

Lanthanide series

- Lanthanide series : In period-6 Ce (Z = 58) to Lu (Z = 71)
- Elements of lanthanide series are called lanthanoids which are shown by symbol Ln.
- All the lanthanoids possess stable oxidation state (+3).
- The general electronic configuration of lanthanoids : $[Xe]4f^{1-14}5d^{0-1}6s^2$
- Amongst lanthanoids, promethium (Pm) is radioactive.

Actinide series

- Actinide series : In period-7 Th (Z = 90) to Lr (Z = 103).
- Elements of actinide series are called actinoids.
- The stable oxidation state in actioids is found to be from (+2) to (+6).
- The general electronic configuration of actinoids is : $[Rn]5f^{0-14}6d^{0-2}7s^2$
- All the actinoids are radioactive.

EXERCISE

1. Select the proper choice from the given multiple choices :

(1) Elements of which groups are called d-block elements ?

(A) Groups 1 to 2 (B) Groups 3 to 12

(C) Groups 13 to 18 (D) Groups 13 to 17

(2) Which of the following elements is of d-block even then it is not transition element ?

(A) Cu (b) Ca (C) Fe (D) Hg

(3) The aqueous solution of which of the following ions is green coloured ?

(A) Co^{2+} (B) Zn^{2+} (C) Ni^{2+} (D) Cr^{2+}

(4) What will be the order of energy of d-orbitals during their division in tetrahedral complex compounds ?

(A) $d_{xy} \cong d_{yz} \cong d_{xz} < d_{x^2-y^2} \cong d_{z^2}$ (B) $d_{x^2-y^2} \cong d_{z^2} < d_{xy} \cong d_{yz} \cong d_{xz}$

(C) $d_{xy} \cong d_{z^2} < d_{yz} \cong d_{xz} \cong d_{x^2-y^2}$ (D) $d_{x^2-y^2} \cong d_{xz} < d_{xy} \cong d_{yz} \cong d_{z^2}$

(5) Which of the following ions in its aqueous solution possesses the value of magnetic moment as 3.87 ?

(A) Cu^{2+} (B) Cr^{3+} (C) Co^{3+} (D) Fe^{3+}

(6) Which of the following is the alloy of iron ?

(A) Stainless steel (B) Brass (C) Bronze (D) Nichrome

(7) Which of the following compounds of transition elements is used in dry cell ?

(A) V_2O_5 (B) $KMnO_4$ (C) $K_2Cr_2O_7$ (D) MnO_2

(8) Which of the following elements is radioactive ?

(A) Pr (B) Pm (c) Gd (D) Tm

(9) What is the general electronic configuration of actinide series ?

(A) $[Xe]\ 4f^{0-14}\ 5d^{0-1}\ 6s^2$ (B) $[Xe]\ 4f^{0-14}\ 5d^{0-10}\ 6s^2$

(C) $[Rn]\ 5f^{0-14}\ 5d^{0-2}\ 6s^2$ (D) $[Rn]\ 5f^{0-14}\ 6d^{0-2}\ 7s^2$

(10) Which of the following statements is incorrect ?

(A) Atoms of all the transition elements are paramagnetic.

(B) All the transition elements are metal elements.

(C) All the elements of d-block are transition elements.

(D) The position of d-block is between s and p-block element in the periodic table.

2. Answer the following questions in brief :

(1) Innertransition elements are the members of which period ?

(2) Elements of which block are called innertransition elements ?

(3) Which of the elements of the first transition series (Sc to Zn) do not act as transition element ?

(4) Which are the elements of first transition series elements having d^5 and d^{10} electronic configurations ?

(5) Aqueous solutions of which ions of first transition elements are blue coloured ?

(6) Write unit of magnetic moment.

(7) Which non-metal elements form interstitial compounds with elements of first transition series ?

(8) Which centres are responsible for catalytic function of catalyst ?

(9) Which alloy is filled in tooth cavity ?

(10) Which alloy is used for preparation of electric resistant wire ?

(11) Cupronickel is the alloy of which metals ?

(12) Which compound of transition elements acts as strong oxidizing agent in acidic and basic media ?

(13) By which common symbol, the elements of lanthanide series are shown ?

(14) Write general electronic configuration of lanthanide series elements.

(15) Which compound of f-block elements is useful in pigment ?

(16) Give definitions :

(i) Transition elements (ii) d-d transition (iii) Interstitial compounds (iv) Alloy (v) Lanthanide series (vi) Actinide series

3. Write answers of following questions :

(1) Write electronic configuration of Cr and Cu.

(2) Write formula to determine magnetic moment and give identification of symbols in it.

(3) Write two properties and two uses of $KMnO_4$.

(4) Write two properties and two uses of $K_2Cr_2O_7$.

(5) Write four applications of f-block elements.

(6) Write three laws presented by scientists Hume and Rothery for preparation of alloys.

(7) Explain giving reason :

(i) The values of second ionization enthalpy of Cr and Cu are found to be more than their neighbouring elements.

(ii) The atomic radius of Zn in first transition series increases instead of decreasing.

(iii) The atomic radii of elements Cr to Cu in first transition series are almost same.

(iv) In electronic configuration of Pd, Ag and Cd, d^{10} electrons are present even then Pd and Ag are considered transition elements, while Cd is not considered transition element.

(8) Calculate magnetic moments : Sc^{3+}, Ti^{4+}, V^{4+}, Cr^{3+}, Cr^{6+}, Mn^{2+}, Fe^{2+}, Co^{3+}, Ni^{2+}, Cu^{2+}, Zn^{2+}

4. Answer the following questions in detail :

(1) Mention the general properties of transition elements.

(2) Discuss catalytic and magnetic properties of transition elements.

(3) Describe the characteristic capacity of transition metal ions to form complex compounds.

(4) Mention the preparation of potassium dichromate and potassium permanganate.

(5) Describe applications of d-block elements.

(6) Compare actinide elements with lanthanide elements.

(7) Write short notes :

(i) Interstitial compounds (ii) Alloys

(iii) Chemical reactivity of lanthanide elements (iv) Lanthanide contraction

•

Unit

4

Complex Salts or Co-ordination Compounds

4.1 Introduction

The salt that is obtained according to the laws of chemical combination by combination of two or more salts (compounds) having independent existence and maintains the properties of the original salts is called double salt, e.g., if saturated aqueous solution of K_2SO_4 and $Al_2(SO_4)_3$ are mixed and crystallized, then crystals of alum- $K_2SO_4 \cdot Al_2(SO_4)_3 \cdot 24H_2O$ are obtained, which is called double salt. On qualitative analysis of solution of this alum, yellow precipitate with picric acid are obtained. This indicates presence of K^+, with NaOH gum-like gelatinous precipitates are obtained which indicates presence of Al^{3+} and with $BaCl_2$ gives white precipitates which indicates presence of SO_4^{2-} and thus K_2SO_4 and $Al_2(SO_4)_3$ maintain their properties. Hence it is a double salt In addition, $FeSO_4$, $(NH_4)_2SO_4 \cdot 6H_2O$ ferrous ammonium sulphate and carnalite are also double salts whose equations are shown below :

$$K_2SO_4 + Al_2(SO_4)_3 + 24H_2O \rightarrow \underset{\text{Potassium alum}}{K_2SO_4 \cdot Al_2\ (SO_4)_3 \cdot 24H_2O}$$

$$FeSO_4 + (NH_4)_2SO_4 + 6H_2O \rightarrow \underset{\text{Ferrous ammonium sulphate}}{FeSO_4 \cdot (NH_4)_2SO_4 \cdot 6H_2O}$$

$$KCl + MgCl_2 + 6H_2O \rightarrow \underset{\text{Carnalite}}{KCl \cdot MgCl_2 \cdot 6H_2O}$$

Similarly, by combination of two or more compounds of independent existence, according to the laws of chemical combination, compounds obtained with new properties are called complex salts or co-ordination compounds, viz. when KCN is added to ferrous cyanide $Fe(CN)_2$, and ferric cyanide $Fe(CN)_3$ are formed and then both of them become soluble. The crystals having formula

$Fe(CN)_2 \cdot 4KCN$ and $Fe(CN)_3 \cdot 3KCN$ are obtanined. They are known as potassium ferrocyanide $K_4[Fe(CN)_6]$ and potassium ferricyanide $K_3[Fe(CN)_6]$ respectively. In these two, the qualitative analysis of Fe^{2+}, Fe^{3+} or CN^- ions cannot be carried out but new ions $[Fe(CN)_6]^{4-}$ or $[Fe(CN)_6]^{3-}$ are obtained. In electrolysis; they move towards anode. Equations for preparing some complex salts are given below :

$$CuSO_4 + 4NH_3 \rightarrow [Cu(NH_3)_4]SO_4$$

$$Fe(CN)_3 + 3KCN \rightarrow K_3[Fe(CN)_6]$$

$$Fe(CN)_2 + 4KCN \rightarrow K_4[Fe(CN)_6]$$

$$CoCl_3 + 6NH_3 \rightarrow [Co(NH_3)_6]Cl_3$$

$$PtCl_4 + 2KCl \rightarrow K_2[PtCl_6]$$

From the equation of $CuSO_4$ and NH_3 as shown above, it appears that the valency of Cu^{2+} and SO_4^{2-} ions is satisfied even then $CuSO_4$ combines with four molecules of ammonia and gives complex salt.

The elements in the d-block of modern periodic table, are called transition elements. Among these transition elements, Scandium to Zinc (Z = 21 to 30), Yttrium to Cadmium (Z = 39 to 48) and Lanthanum to Mercury (Z = 57 to 80). That is the elements of fourth, fifth and sixth periods are very important in modern inorganic chemistry. Some of the properties of these elements are same. There is change in their valency and the aqueous solution of their salts are coloured. In the electronic configuration of these transition elements, electrons are arranged successively in d-orbitals. **When (n–1)d, ns and np or ns, np and nd orbitals of the atoms or ions of transition elements are vacant, these transition elements accept electron pairs from negative ions or neutral molecules, which are called complex (co-ordination) compounds.** The bond that is formed between metal ions of metallic elements in this type of compound and the negative ions or neutral molecules is called co-ordinate covalent bond. In these compounds, around the metal ion in the centre of the molecules are arranged negative ions or neutral molecules by co-ordinate covalent bond. Mostly the transition elements have more tendency to form complex compounds.

4.2 Werner's Theory

The question that why the other compounds combine with stable compounds after satisfaction of valency; had become the subject of discussion in the beginning of the nineteenth century. In 1905, Swiss scientist Alfred Werner prepared many new compounds by mixtures of cobalt chloride and ammonia, studied them thoroughly and gave a new theory, which is known as Werner's co-ordination theory. By the Werner's theory many clarifications were made in this field. **Werner obtained different compounds of $CoCl_3$ and NH_3 viz. $[Co(NH_3)_6]Cl_3$ (yellow), $[Co(NH_3)_5Cl]Cl_2$ (purple, $[Co(NH_3)_4Cl_2]Cl$ (violet / green) and $[Co(NH_3)_3Cl_3]$, (bluish green).** Werner received nobel prize for this work in 1913.

The theory given by Werner for the formulas and structures of complex salt is "Certain metals have secondary valency in addition to primary valency. With this it combines strongly with the metal ions, neutral molecules or negative ions in its first attraction sphere [ ]."

According to Werner, metal ion possesses two types of valency. Primary valency or ionizing valency and Secondary valency or non-ionisable valency.

(i) The primary valency is similar to positive electric charge of positive ion of metal or oxidation number of metal. It forms ionic bond. The other ion combined by this valency becomes free by ionization of compound.

(ii) Secondary valency is non-ionized. The negative ions or neutral molecules contained with secondary valency are not ionized.

(iii) Primary valency is satisfied by only negative ions, while secondary valency is satisfied by negative ions or neutral molecules (ligand).

(iv) The secondary valency of metal ion or metal elements form co-ordinate covalent bond, so the secondary valency is called co-ordination number of metal.

(v) The co-ordination number of metal ion is definite which is independent of primary valency.

(vi) Secondary valency being directional, different types of geometrical structures are obtained.

Note : (1) In some complex compounds, the positively charged ligand like ^{+}NO, $^{+}NO_2$ are also seen.

(2) Now it is proved that the some metal ions of transition elements possess more than one co-ordination number.

(3) To understand these geometrical structures, it is important to know the magnetic properties of complex compounds.

The bond between metal ion combined through secondary valency, and ligand is called co-ordinate covalent bond. This is indicated by $\rightarrow$ which indicates the co-ordination site. The electron pair present between metal ion and ligand are given by negative ion or neutral molecule.

The complex ion in complex compounds is shown by [ ] bracket (first attraction sphere) whereas outside the bracket, on the left side, the positive ion combined by ionic bond is shown viz. In $K_3[Fe(CN)_6]$ complex compound $[Fe(CN)_6]^{3-}$ complex negative ion is combined by ionic bond with positive ion K^+. Similarly in $[Cr(NH_3)_6]Cl_3$ complex compound $[Cr(NH_3)_6]^{3+}$ complex positive ion is combined with negative ion Cl^- by ionic bond.

4.3 Clasification of Ligands

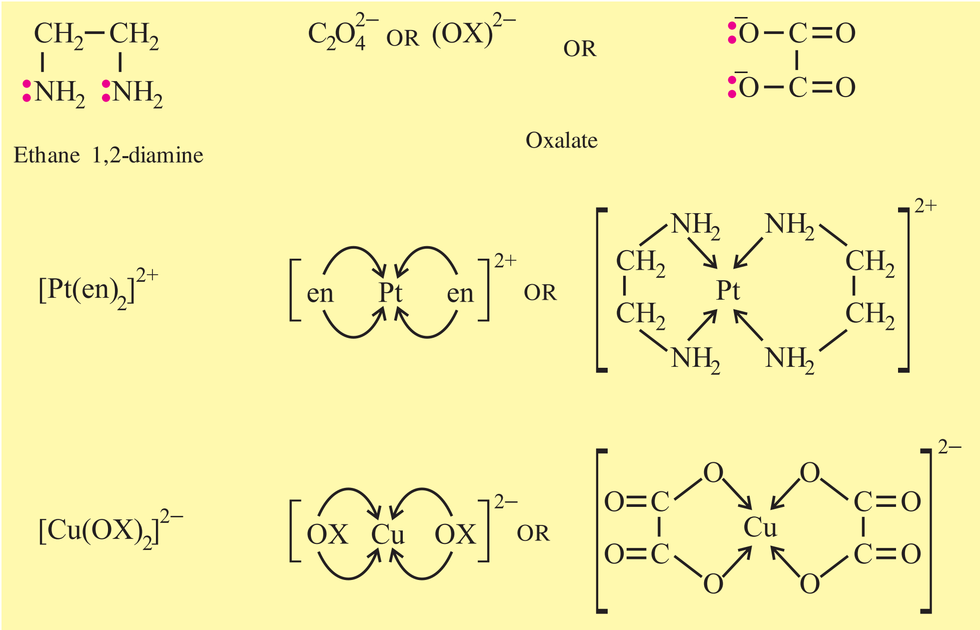
Generally the ligand has a negative electric change or is a neutral molecule. The classification of ligands is made on the basis of the number of electron pair donating atoms.

(I) Unidentate ligand : If only one atom of ligand of negative ion or neutral molecule donates one atom to metal ion by giving one electron pair and form one co-ordinate covalent bond, then it is called unidentate ligand. Neutral molecules like H_2O, NH_3, CO and negative ions like Cl^-, CN^-, F^- which combine with metal ion by giving one electron pair; viz. In $[Cr(NH_3)_6]Cl_3$ complex compound, nitrogen atom of each ammonia combine by giving one electron pair to chromium metal ion, it is called unidentate ligand. In $[Cr(H_2O)_6]Cl_3$ complex compounds each molecule of water co-ordinates with chromium metal ion by giving one electron pair, so it is unidentate ligand. In these unidentate ligands, nitrogen atom in ammonia molecule and oxygen atom in water molecule donate one electron pair to metal ion. Hence, the **atom present in negative ion or neutral molecule gives electron pair to metal-ion which is called the co-ordinate site of ligand** and it is shown as $M \leftarrow L$, where M is metal ion and L is ligand.

(II) Didentate ligand : The ligand which can donate two electron pairs and forms two co-ordinate covalent bonds, is called didentate ligand. In this type of ligand two atoms combine with metal ion by giving two electron pairs and form two co-ordinate covalent bonds. Thus, only one ligand satisfies two secondary valencies of metal ion viz. In ethane 1-2 diamine (ethylene diamine–en), its one molecule forms two co-ordinate covalent bonds by giving two electron pairs on two nitrogen atoms, to metal ion.

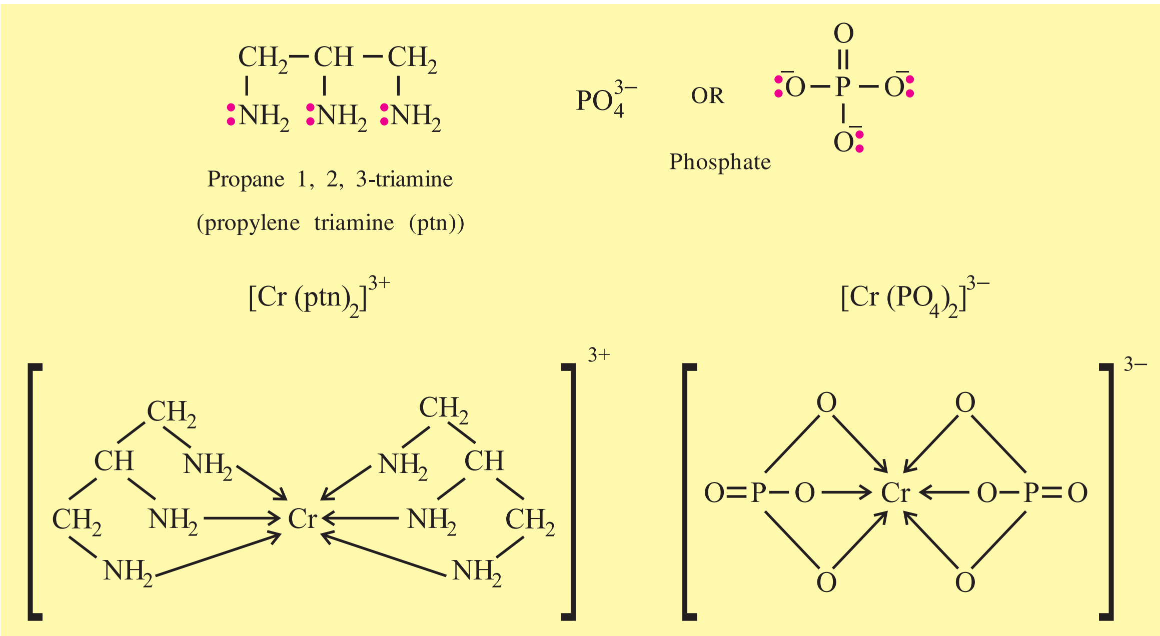
In oxalate $(OX)^{2-}$ negative ion, two electron pairs on two oxygen atoms are given to metal-ion and two co-ordinate covalent bonds are formed.

e.g.,



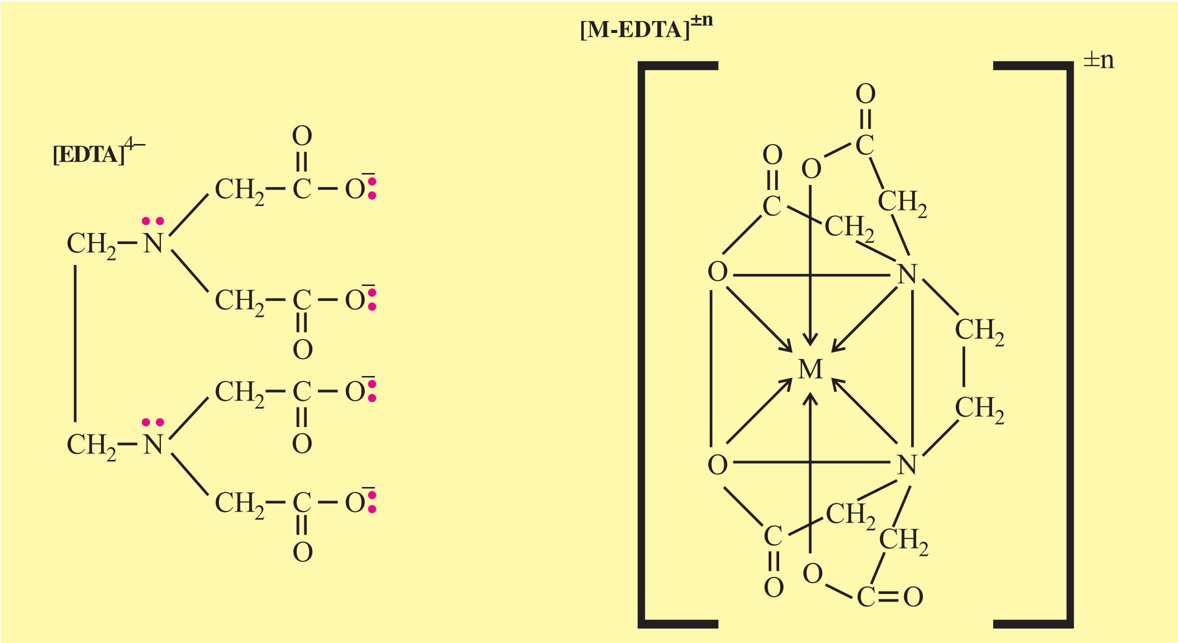
(III) Tridentate ligand : The ligand in which there are three co-ordination sites is called tridentate ligand. In this type of ligand, three atoms form three co-ordinate covalent bonds by giving three electron pairs. Thus only one ligand satisfies three secondary valencies of metal ion viz. Propane 1-2-3 triamine (Propylene triamine-ptn) one neutral molecule form three co-ordinate covalent bonds by giving three electron pairs on its three nitrogen atoms. Similary, PO_4^{3-} is a negative ion tridentale ligand.

e.g.



(IV) Hexadentate ligand : The ligand in which there are six co-ordination sites is called hexadentate ligand. Six atoms of this type of ligand give six electron pairs to metal ion and form six co-ordinate covalent bonds. Thus, only one ligand satisfies six secondary valencies of metal ion viz. In ethylene diamine tetraacetate $(EDTA)^{4-}$, four oxygen atoms and two nitrogen atoms give four electron pairs and two electron pairs respectively to metal ion and form six co-ordinate covalent bonds.

e.g. $[M\ EDTA]^{\pm n}$

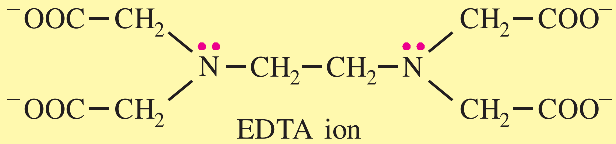


Generally, the ligand in which two or more co-ordination sites are there or the ligand in which two or more atoms donate electron pairs to metal ion and form co-ordinate covalent bond, is called multidentate ligand.

Some ligands and their types are shown in table 4.1

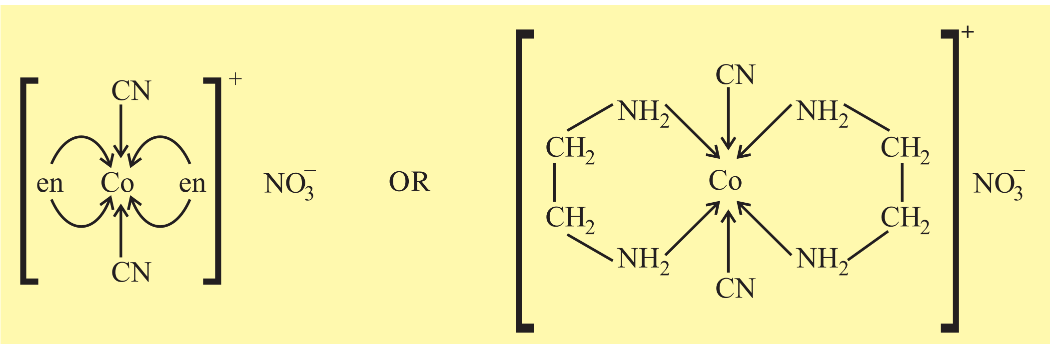
Table 4.1 Some ligands and their types

Type	Ligand	Electric charge
Unidentate Neutral ligand	$H_2\ddot{O}:$, $:NH_3$, $:CO$, $:NO$, $CH_3\ddot{N}H_2$, $C_5H_5\ddot{N}$ (py)	0
Unidentate Negative ion ligand	^-OH, F^-, Cl^-, Br^-, I^-, ^-CN, $^-NH_2$, NO_3^-, NO_2^-, $CH_3COO^-(AcO^-)$, O^{2-}, S^{2-}, N^{3-}	– 1
Unidentate Positive ion ligand	^+NO, $^+NO_2$,	+ 1
Didentate Neutral ligand	$H_2\ddot{N}-CH_2-CH_2-\ddot{N}H_2$ (en), $CH_2(:NH_2)-CH_2-CH_2(:NH_2)$ (pn)	0
Didentate Negative ion ligand	CO_3^{2-}, SO_4^{2-}, COO^-–COO^- $(OX)^{2-}$	– 2
Tridentate Neutral ligand	$CH_2(:NH_2)-CH(:NH_2)-CH_2(:NH_2)$ (ptn)	0

Tridantate Negative ion ligand	PO_4^{3-}, AsO_4^{3-}	– 3
Hexadentate Negative ion ligand		– 4

When multidentate ligands like en, OX^{2-}, pn, ptn, $(EDTA)^{4-}$ form co-ordinate covalent bond with metal ion, it results into cyclic structure involving central metal ion. **Thus, the complex compound formed by ligand and metal ion having cyclic structure is called chelate.** The stability of such chelate compounds is more than the stability of simple complex compounds (formed by monodentate ligand).

e.g. $[Co(CN)_2(en)_2]^+ NO_3^-$



4.4 Requirements for Formation of Complex Compounds

Some basic requirements are necessary for formation of complex compound. The capacity of transition element metal ions of formation of complex compounds is more than that of other elements.

Basic requirements :

(i) The ligand must have electron pair which can be easily donated.

(ii) There must be vacant orbitals in the metal ion to receive electron pairs.

(iii) The metal ions should have the symmetry of vacant orbitals same as the symmetry of orbitals of ligand having electron pairs. Hence, co-ordinate covalent bond can be formed by overlapping of orbitals having electron pair in ligand.

These basic requirements are satisfied by metal ions and so complex salts are easily formed. Even then, it is not necessary that each transition element forms complex compounds with equal ease. There is change in capacity of formation of complex compounds according to different oxidation states of metal ions. In addition, there is difference in the stability of the compounds that are formed.

4.5 Stability of Complex Compounds and Strength of Ligand

As the strength of different ligands to form co-ordinate covalent bond varies, stronger ligand possesses more attraction with metal ion. As a result, complexes having stronger ligands possess more stability and the complexes having weak ligand possess less stability. Hence, it can be said that the strength of the complexes is determined on the basis of strength of ligand. The basis of strength of ligand can be determined by same metal ion combined with different ligands. The order of strength of some complex compounds is as follows : **$Cl^- < F^- < OH^- < H_2O < NH_3 < CN^-$**

$[NiCl_4]^{2-} < [NiF_4]^{2-} < [Ni(OH)_4]^{2-} < [Ni(H_2O)_4]^{2-} < [Ni(NH_3)_4]^{2-} < [Ni(CN)_4]^{2-}$

$[CuCl_4]^{2-} < [CuF_4]^{2-} < [Cu(OH)_4]^{2-} < [Cu(H_2O]^{2+} < [Cu(NH_3)_4]^{2+} < [Cu(CN)_4]^{2-}$

Hence, change in the capacity is seen in the formation of complex ions of metal ions of transition elements according to different conditions.

It is not necessary that any metal ion form only one type of complex salt with ligand. Sometimes different types of ligands form complex compounds with metal ion. This type of complex compounds are called mixed ligand complex compounds. viz. In the complex $[Cr(NH_3)_4(CN)_2]NO_3$, four ammonia molecules and two cyanide ions, total six unidentate ligands are combined. If ligand is of only one type, the complex salts are called simple complex compounds.

As in mixed ligand complex compounds two or more types of ligands combine and form complex compound, similary if only one metal ion is present in any complex compound, then it is called unicentred complex compound, e.g. $K[MnO_4]$. **If more than one metal ions are present in the complex compound then it is called polycentred complex compound.** e.g. $K_2[Cr_2O_7]$.

In unicentred or polycentred complex compounds, the three dimensional arrangement of ligand combined to metal ion are directional, so different geometrical structures formed accordingly are called co-ordination polyhedral. Mostly, the shape of this geometrical structure is tetrahedral, square planar, octahedral, square pyramidal and trigonal bipyramidal.

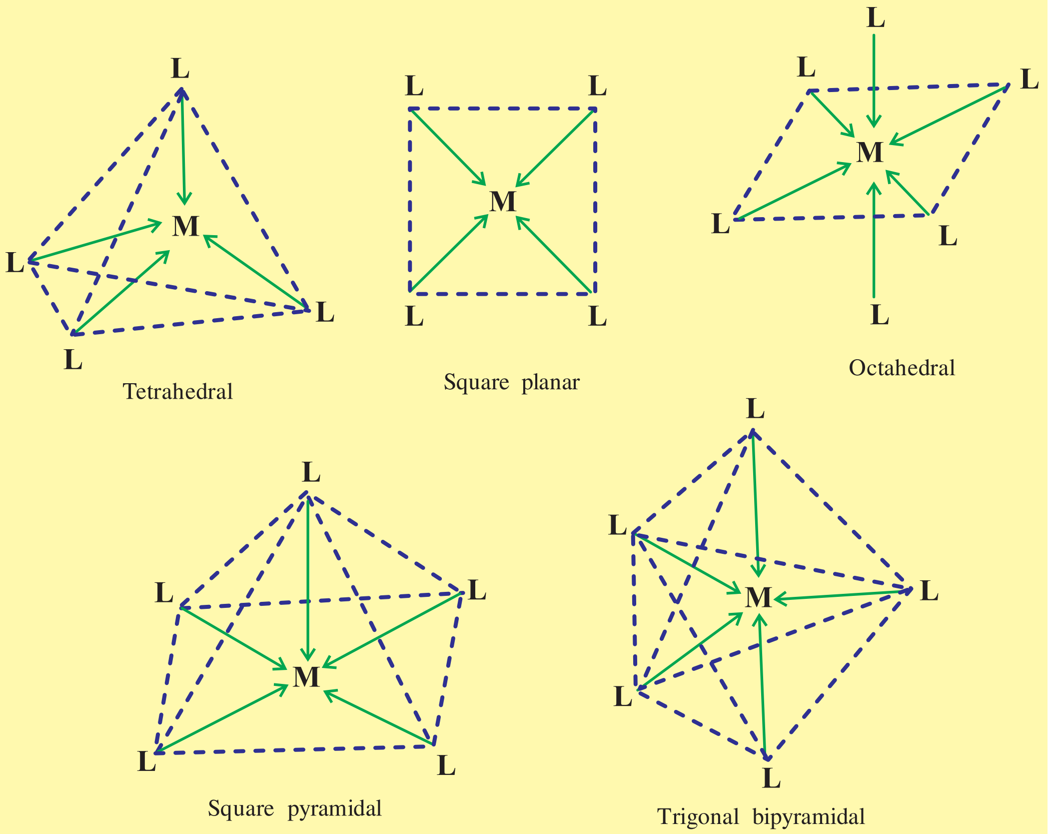


Fig. 4.1

4.6 Geometry of Complex Ions

The co-ordination number of metal ion and the geometrical structures of ions can be known from the studies of magnetic properties and crystal field theory. It is not necessary that metal ion should possess same co-ordination number and one definite type of geometrical structure containing complex ions. Sometimes, it may happen that the metal ion possesses more than one co-ordination number and even if the co-ordination number remains the same, it can form complex ions having different types of geometrical structures. In most of the metal complex ions, the co-ordination number 4 and 6 are very common and the complex ions showing these co-ordination numbers are found comparatively in larger number.

Co-ordination number 4 : The metal ions possessing co-ordination number 4 form complexes having two types of geometrical structures :

(i) Tetrahedral structure containing complex ions and (ii) square planar structure containing complex ions.

(i) Tetrahedral complex ions : Metal ions form tetrahedral complex ions and they are stable in special conditions. Most of the tetrahedral complex ions are obtained in negative ion form viz. MnO_4^- . While, tetrahedral complex like $[Ni(CO)_4]$ is obtained in form of neutral molecule.

(ii) Square planar complex ions : Metal ions of only some of the transition elements form square planar complexes. Square planar complexes are formed by combination of Ni^{2+} metal ion with negative ion and / or neutral molecule ligand viz. $K_2[Ni(CN)_4]$ and $[Ni(NH_3)_2Cl_2]$.

In table 4.2 some examples of tetrahedral and square planar complexes possessing co-ordination number 4 of some transition elements in the fourth period are shown.

Table 4.2 Some complexes of transition elements possessing co-ordination number-4

Transition Element	Oxidation state	Electrons of 3d-orbital	Complex compound	Geometrical structure
Mn	+7	$3d^0$	$K[MnO_4]$	Tetrahedral
Co	+2	$3d^7$	$K_2[CoCl_4]$	Tetrahedral
Ni	0	$3d^{10}$	$K_4[Ni(CN)_4]$	Tetrahedral
	+2	$3d^8$	$K_2[NiCl_4]$	Tetrahedral
	+2	$3d^8$	$K_2[Ni(CN)_4]$	Square planar
	+2	$3d^8$	$[Ni(NH_3)_2Cl_2]$	Square planar

Co-ordination number 6 : The complex compounds of metal ions possessing co-ordination number 6 are easily available. The geometrical structures of these complexes are octahedral due to which different ligands deformations are also found viz. $[CrCl_2(en)_2]NO_3$ and $[Co(NH_3)_4CO_3]Cl$.

In table 4.3, some examples of octahedral complexes possessing co-ordination number 6 of elements of some transition elements of fourth period are given.

Table 4.3 Some complexes possessing co-ordination number 6 of transition elements

Transition Element	Oxidation state	Electrons of 3d-orbital	Complex compound (Octahedral structure)
Cr	0	$3d^6$	$[Cr(CO)_6]$
	+1	$3d^5$	$K_4[Cr(CN)_5(NO)]$
	+3	$3d^3$	$K[Cr(NH_3)_2(CO_3)_2]$
	+4	$3d^2$	$K_2[CrF_6]$
Mn	+2	$3d^5$	$[Mn(H_2O)_6]Cl_2$
	+3	$3d^4$	$[Mn(en)_3]Cl_3$
Fe	+2	$3d^6$	$K_4[Fe(CN)_6]$
	+3	$3d^5$	$Na_3[Fe(OX)_3]$
Co	+2	$3d^7$	$Na_4[Co(NO_2)_6]$
	+3	$3d^6$	$[Co(NH_3)_6]Cl_3$
	+4	$3d^5$	$K_2[CoF_6]$
Ni	+2	$3d^8$	$[Ni(H_2O)_6]Cl_2$
	+3	$3d^7$	$K_3[Ni(CN)_6]$
	+4	$3d^6$	$K_2[NiF_6]$

4.7 Hybridization of Orbitals of Metal Ions of Complex Compounds and Magnetic Properties

Most of the metal ions have geometrical structures- tetrahedral, square planar or octahedral. To understand this geometrical structures, hybridization of metal-ion orbitals and magnetic property are very useful.

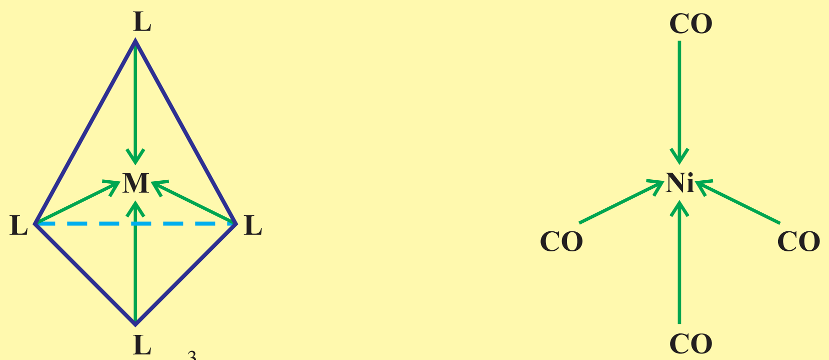
sp^3 hybridization : When one 4s-orbital and three $4p_x$, $4p_y$, $4p_z$ orbitals overlap with one another and get hybridised, then new four hybrid orbitals are produced. These four orbitals are called sp^3 hybrid orhitals. The energy value of these four hybrid orbitals are same and are extended towards the four corners from the centre of the tetrahedral. Also, the angle between any two sp^3 hybrid orbitals is $109^0 28'$. Hence, if the metal ion orbitals in complex ion hybridization of sp^3 type, then the geometrical structure of complex compounds is tetrahedral.

dsp^2 hybridization : When one 3d-orbital, one 4s-orbital and two $4p_x$, $4p_y$ orbitals of metal ion overlap with one another and get hybridized, new four hybrid orbitals are produced. These four hybrid orbitals are called dsp^2 hybrid orbitals. The value of energy of these four hybrid orbitals is same and are extended towards four corners from the centre of the plane square. Also, the angle between any two nearby dsp^2 orbital is 90^0. Hence, if the hybridization of metal ion orbitals in compound, is of dsp^2 type then, the geometrical structure of metal in the compound becomes of dsp^2 type; then the geometrical structure of complex compounds is square planar.

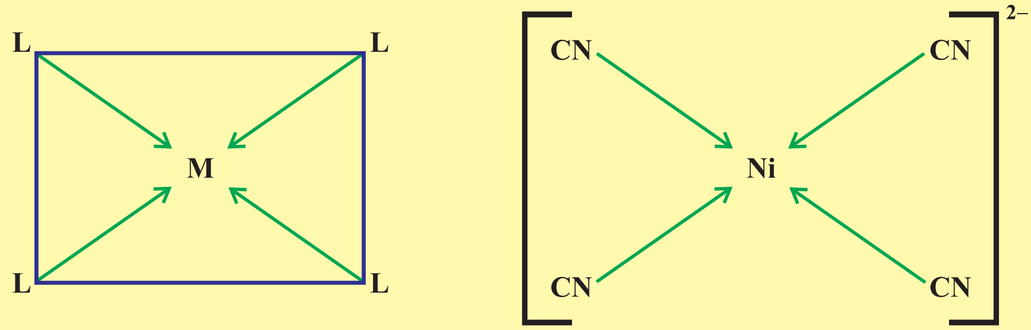
d^2sp^3 hybridization : When two 3d-orbitals, one 4s-orbital and three $4p_x$, $4p_y$, $4p_z$ orbitals overlap with one another and get hybridized, then new six hybrid orbitals are produced. These hybrid orbitals are called d^2sp^3 hybrid orbitals. The value of energy of these six hybrid orbitals is same and it is extended to six corners of octahedral from the centre of that octagon. Also, all these hybrid orbitals are on X-axis, Y-axis and Z-axis to one another. So the angle between any two nearby d^2sp^3 orbitals is 90^0. Hence, the complex compounds in which the hybridization of metal ion in complex compound is of d^2sp^3 type, the geometrical structure of this complex compound is octahedral.

sp^3d^2 hybridization : Sometimes, sp^3d^2 hybridisation can occur to produce octahedral structure when 3d-orbitals of metal ion are not available, then one 4s-orbital, three $4p_x$, $4p_y$, and $4p_z$ orbitals and two 4d-orbitals overlap with one another and sp^3d^2 hybridization occurs. Here also, in sp^3d^2 hybridization the geometrical structure is octahedral.

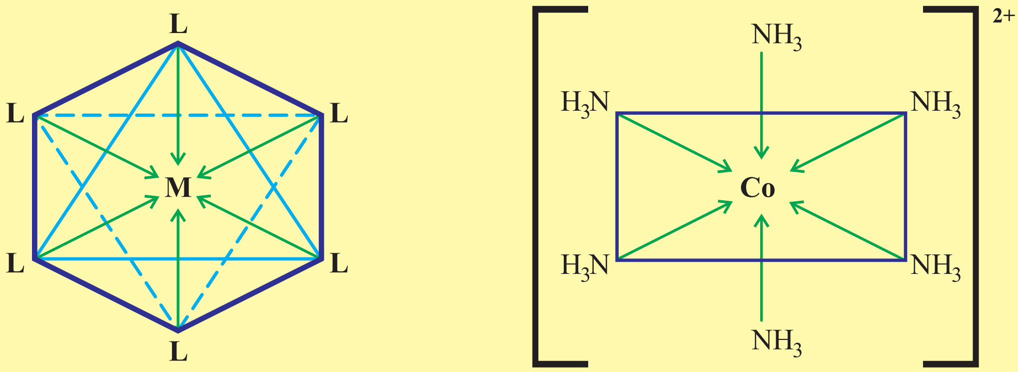
Hence, it can be said that in octahedral structure d^2sp^3 or sp^3d^2 hybridization occurs. To determine which type of hybridization is possessed by study of magnetic properties becomes very essential. In fig. 4.2 the geometrical structures of complex molecules or ions on the basis of sp^3, dsp^2, d^2sp^3 or sp^3d^2 hybridization are shown.



sp^3 hybridization-tetrahedral structure



dsp^2 hybridization-square planar structure



d^2sp^3 Or sp^3d^2 hybridization-octahedral structure

Fig. 4.2 Geometrical structures of complex ions

Downloaded from https:// www.studiestoday.com

Magnetic properties : If the electrons of 3d-orbital of the metal ion are paired in transition element in any complex then the complex is called diamagnetic. If there are unpaired electrons then the complex is called paramagnetic. The theoretical value of magnetic moment can be found out by the equation $\mu = \sqrt{n(n+2)}$ where n = number of unpaired electrons. The unit of this value is BM (Bohr magneton).

The magnetic moment of complex compounds of metal ions of transition elements is dependent on geometrical structure, type of ligand etc. In the detailed study of complex compounds, the calculation of magnetic moment is very helpful. In table 4.1, the theoretical and experimental values of magnetic moment of Sc^{2+} to Zn^{2+} (d^1 to d^{10}) ions are given.

Table 4.4 Unpaired electrons and magnetic moment of M^{2+} ions of transition elements

Ion	$3d^n$	Unpaired Electrons	Magnetic moment (μ) BM	
			Theoretical value	Experimental value
Sc^{2+}	d^1	1	1.73	1.73 – 1.74
Ti^{2+}	d^2	2	2.83	2.76
V^{2+}	d^3	3	3.87	3.86
Cr^{2+}	d^4	4	4.90	4.80
Mn^{2+}	d^5	5	5.92	5.96
Fe^{2+}	d^6	4	4.90	5.00 – 5.50
Co^{2+}	d^7	3	3.87	4.40 – 5.20
Ni^{2+}	d^8	2	2.83	2.90 – 3.40
Cu^{2+}	d^9	1	1.73	1.80 – 2.20
Zn^{2+}	d^{10}	0	0	0

4.8 IUPAC Nomenclature of Complex Compounds

There are different types of ligand for variety of complex compounds with various metal elements. Also many complex compounds possessing mixed ligands are prepared. In the earlier times, the names of the complex compounds were given on the basis of metal ion and their colours. Werner had named the complexes obtained from cobalt chloride and ammonia on the basis of their colours viz. greenio cobaltic ammonium chloride, violetio cobaltic ammonium chloride, purpleo cobaltic ammonium chloride etc. As many complexes were being formed, the accurate names of complex compounds could not be given. Hence to have accuracy in the names of the complex compounds, IUPAC method was used. The IUPAC nomenclature of monocentric complex compounds is done according to the definite rules as shown below :

(i) In showing name of ionic complex compound, the positive ion is written first and then the name of negative ion, is mentioned.

(ii) In complex compound co-ordination sphere [ ] in naming is done according to the rules.

(a) In nomenclature in co-ordination sphere the name of ligand according to English alphabates is first shown and then the name of the metal is written.

(b) For the ligands having negative electric charge, the suffix 'O' is joined with the name of the ligand viz OH^- hydroxo, CN^- cyano, NH_2^- amido, NO_2^- nitro, ONO^- nitrito, NO_3^- nitratato, SCN^- thiocyanato, CNO^- cyanato, CO_3^{2-} carbonato, O^{2-} oxo, OX^{2-} oxalato, PO_4^{3-} phosphato, AsO_4^{3-} arsenato etc. As per IUPAC rules, 2004, Cl^- is written as chlorido instead of chloro and Br^- as bromido instead of bromo.

(c) The neutral ligand is shown according to its original name viz. $CH_3 \cdot NH_2$ methyl amine, $H_2N–CH_2–CH_2–NH_2$ ethane 1-2 diamine, $NH_2–CH_2–CH(NH_2)–CH_2–NH_2$ propane-1,2,3-triamine $H_2N–CH_2–CH_2–CH_2–NH_2$ propane 1,2-diamine but as exception H_2O aqua, NH_3 ammine, CO carbonyl, NO nitrosyl etc. are written.

(d) If the number of ligand of only one type is more than 1, then according to the number of ligands 2, 3, 4, 5, 6, the words di, tri, tetra, penta, hexa etc are used as prefix respectively. If organic ligand is there, and the prefix is a number, then ligand is placed in bracket and for numbers 2, 3, 4 etc. corresponding bis, tris, tetrakis prefix are added.

(e) If there is monocentred complex ion, in which only one metal ion, then there is definite order of showing its name.

(i) If the complex is negative ion, then write names of ligand successively. The suffix 'ate' is joined with metal ion and then its oxidation state is shown in ( ) bracket in Roman numbers viz. ferrate (II), ferrate (III), chromate (III), manganate (VII), cobaltate (III), nickelate (II), molybdate (VI).

(ii) If the complex is positive ion or neutral molecule, then first writing the name of ligand successively, the name of metal is added at the end and its oxidation state is shown in ( ) bracket, in Roman numbers. The whole part of complex positive ion, negative ion and neutral molecule is placed in the square bracket [ ].

The formulae and IUPAC names of some complex compounds on the basis of these rules are given below :

Formula	IUPAC
$K[MnO_4]$	Potassium tetraoxomanganate(VII)
$Na_2[CoCl_4]$	Sodium tetrachloridocobaltate(II)
$[Ni(CO)_4]$	Tetracarbonylnickel(0)
$K_2[NiCl_4]$	Potassium tetrachloridenickelate(II)
$[Ni(CN)_4]^{2-}$	Tetracyanonickelate(II)ion
$[Cu(NH_3)_4]Cl_2$	Tetraamminecopper(II)chloride
$[Cr(CO)_6]$	Hexacarbonylchromium(0)

$[Ni(H_2O)_2(NH_3)_4]SO_4$	Tetraamminediaquanickel(II)sulphate
$[Cr(OX)_3]^{3-}$	Trioxalatochromate(III)ion.
$[Co(en)_2(CN)_2]Cl$	Dicyanobis(ethane-1,2-diamine)cobalt(III)chloride
$K_4[Fe(CN)_6]$	Potassium hexacyanoferrate(II)
$(NH_4)_2[MoO_4]$	Ammonium tetraoxomolybdate(VI)
$K_2[CrF_6]$	Potassium hexafluoridoochromate(IV)
$Na_2[Fe(NO)(CN)_5]$ (Sodium introprusside)	Sodium pentacyanonitrosoniumferrate(II)
$Na_4[Co(NO_2)_6]$	Sodium hexanitrocobaltate(II)
$NH_4[Co(NH_3)_2(OX)_2]$	Ammonium diamminedioxalatocobaltate(III)
$[Pt(Pn)_2CO_3]SO_4$	Carbonatobis(propane-1,3-diammine)platinum(II)sulphate
$[Ag(NH_3)_2][Ag(CN)_2]$	Diammineargentinun(I) dicyanoargentate(I)
$[Cr(en)_3][Cr(OX)_3]$	Tris(ethane-1,2-diamine)chromium(III)trioxalatochromate(III)
$[Pt(NH_3)_2ClNO_2]$	Diamminechloridonitroplatinum(II)
$[Co(NH_3)_4CO_3]Cl$	Tetramminecarbonatocobalt(III)chloride.
$[Cr(NH_3)_4(ONO)Cl]NO_3$	Tetraamminechloridonitritochromium(III)nitrate
$Na_2[Ni(EDTA)]$	Sodium ethylenediamminetetraacetatonickelate(II)
$[Pt(Py)_4][PtBr_4]$	Tetrapyridineplatinum(II)tetrabromidoplatinate(II)
$[CuCl_2(CH_3NH_2)_2]$	Dichloridodi(methanamine)copper(II)
$[Cr(NH_3)_6][Cr(SCN)_6]$	Hexaamminechromium(III)hexathiocyanatochromate(III)

Formula from the IUPAC name of the complex

(a) Tetrammineaquachloridocobalt(III)chloride

$[Co(NH_3)_4(H_2O)Cl]Cl_2$

(b) Potassium tetrahydroxozincate(II)

$K_2[Zn(OH)_4]$

(c) Sodium trioxalatoaluminate(III)

$Na_3[Al(C_2O_4)_3]$

(d) Dichloridobis(ethane-1-2-diamine)cobalt(III)ion.

$[Co(en)_2Cl_2]^+$

(e) Pentacarbonyliron(0)

$[Fe(CO)_5]$

4.9 Geometrical Structures and The Magnetic Properties of Complex Compounds

$[Ni(CN)_4]^{2-}$ complex ion : In tetracyano nickelate (II) four CN^- strong ligands are combined with Ni^{2+} metal ion. In this compound, four ligands are joined and so possesses square planar or tetrahedral structure. Here the electronic configuration of Ni metal and Ni^{2+} ion are $[Ar]3d^8 4s^2$ and $[Ar]3d^8$ respectively. The arrangement of eight electrons in 3d-orbitals is shown in the following figure. The orbitals are shown by square blocks and the electrons by vertical arrow sign ↑ or ↓.

	3d					4s	4p		
Ni [Ar]	↑↓	↑↓	↑↓	↑	↑	↑↓			
Ni^{2+} [Ar]	↑↓	↑↓	↑↓	↑	↑				
$[Ni(CN)_4]^{2-}$ [Ar]	↑↓	↑↓	↑↓	↑↓	××	××	××	××	

Unpaired electrons n = 0 — dsp^2 hybridization.

Hence μ = 0 BM

In one square block, the vertical arrows in opposite directions show electron pairs in which the rotation of both the electrons is in opposite directions. Here Ni^{2+} metal ion combines with four CN^- strong ligand and forms complex ion. Each CN^- strong ligand gives one electron pair to Ni^{2+} metal ion Hence Ni^{2+} metal ion receives four electron pairs from four ligands.

If the structure of $[Ni(CN)_4]^{2-}$ is square planar, then dsp^2 type hybridization occurs in Ni^{2+} metal ion. For this, one 3d-orbital, one 4s-orbital and two 4p-orbitals take part in hybridization of dsp^2, orbitals having equal energy are produced. For this, the arrangement of eight electrons in 3d-orbitals of Ni^{2+} metal ion, the charge is necessary because of the strong ligand like CN^-. In this two unpaired electrons in 3d-orbital, one unpaired electron of 3d-orbital enters into the other 3d-orbital having unpaired electron and forms a pair. As a result one 3d-orbital becomes vacant.

[Note : Because of strong ligands like CN^-, NH_3 and CO, the unpaired electrons get paired in rearrangement of electrons.]

This vacant one 3d-orbital, one 4s-orbital and two 4p-orbitals overlap and form dsp^2 four hybrid orbitals and arrange at 90^0 angle in square planar form. In the four dsp^2 orbitals produced, this four electrons pairs coming from four strong ligand CN^-, get arranged, which are shown by [××] sign. **Here $[Ni(CN)_4]^{2-}$ complex possesses dsp^2 hybridization and its all 3d-orbitals have paired electrons and so it becomes diamagnetic and its geometric structure is square planar.**

$[NiCl_4]^{2-}$ complex ion : In tetrachlorido nickelate (II) complex ion, Ni^{2+} metal ion is combined with four weak Cl^- ligands. As CN^- strong ligand comes nearer to the metal ion because it possesses more attraction, while weak Cl^- ligand does not come near to metal ion because it possesses less attraction. Hence, the 3d-orbital of Ni^{2+} ion is not capable of forming co-ordinate covalent bond with four Cl^- weak ligand. Hence, the rearrangement of electrons in 3d-orbitals if Ni^{2+} metal ion is not required. Here, one 4s, and three 4p-orbitals take part in hybridization and form sp^3 hybrid orbitals having same energy, which is arranged at the angle of $109^0 28'$ tetrahedrally in which the four electron pairs coming from four weak Cl^- ligand are arranged.

	3d					4s	4p		
Ni [Ar]	↑↓	↑↓	↑↓	↑	↑	↑↓			
Ni^{2+} [Ar]	↑↓	↑↓	↑↓	↑	↑				
$[NiCl_4]^{2-}$ [Ar]	↑↓	↑↓	↑↓	↑	↑	××	××	××	××

Unpaired electrons n = 2 sp^3 hybridization

Hence, μ = 2.83 BM

Because of this $[Ni(Cl)_4]^{2-}$ complex ion has sp^3 type of hybridization and its geometrical structure becomes tetrahedral. As there are two unpaired electrons in two 3d-orbitals, the theoretical value of its magnetic moment becomes 2.83 BM and the value of experimental magnetic moment is 2.90 BM. So it becomes paramagnetic.

$[Fe(CN)_6]^{4-}$ complex ion (Ferrocyanide ion) : In hexacyano ferrate (II) complex ion Fe^{2+} metal ion has combined with six CN^- strong ligand and so it possesses octahedral structure. Here, the electronic configuration of Fe metal and Fe^{2+} ion are [Ar] $3d^6 4s^2$ and [Ar]$3d^6$ respectively. The arrangement of six electrons in 3d-orbitals is shown below :

	3d					4s	4p		
Fe [Ar]	↑↓	↑	↑	↑	↑	↑↓			
Fe^{2+} [Ar]	↑↓	↑	↑	↑	↑				
$[Fe(CN)_6]^{4-}$ [Ar]	↑↓	↑↓	↑↓	××	××	××	××	××	××

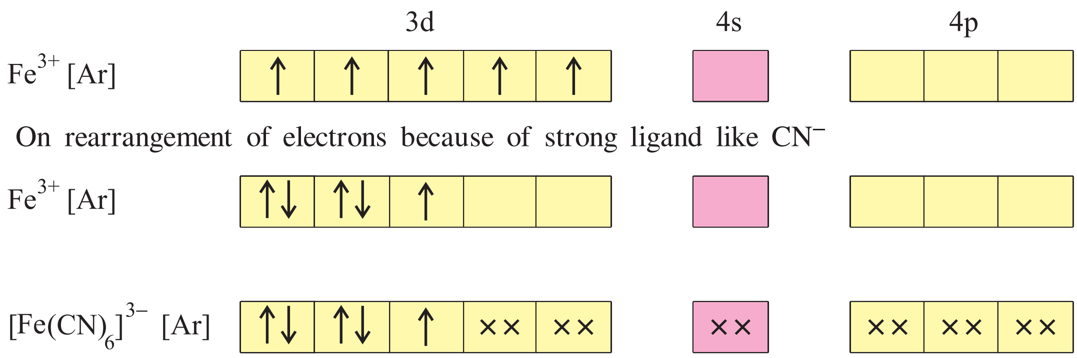
Number of unpaired electrons n = 0 d^2sp^3 hybridization

Hence, μ = 0 BM

In Fe^{2+} ion d^2sp^3 type hybridization occurs. Octahedral structure is obtained in this hybridization. In d^2sp^3 hybridization, the inner 3d-orbital takes part.

For d^2sp^3 hybridisation there must be two 3d-orbitals vacant in Fe^{2+} metal ion. For this, the rearrangement of six electrons in 3d-orbitals is necessary i.e. as the CN^- ion is a strong ligand, total six electrons form three electron pairs and get arranged in 3d-orbitals. As a result the vacated two 3d-orbitals, one 4s-orbital and three 4p-orbitals overlap and form d^2sp^3 hybridization; six d^2sp^3 hybrid orbitals having same energy, produced in the above are arranged octahedrally. In d^2sp^3 hybrid orbitals six electron pairs coming from six CN^- strong ligand are arranged. **Here, $[Fe(CN)_6]^{4-}$ complex ion possesses d^2sp^3 type hybrid orbitals and its geometrical structure is octahedral. In the 3d-orbitals of this complex only paired electrons are there and so it becomes diamagnetic.**

$[Fe(CN)_6]^{3-}$ Hexacyanoferrate (III) ion (Ferricyanide ion) : In hexacyanoferrate (III) complex ion, the oxidation state of Fe is +3 and its coordination number is 6. The electronic configuration of Fe^{3+} metal ion in this complex compound is [Ar] $3d^5$. As CN^- is a strong ligand, rearrangement of electrons in 3d-orbitals is necessary. On rearrangement of five electrons in 3d-orbitals, two orbitals become paired and one electron remains unpaired in 3d-orbital. Hence, two 3d-orbitals, one 4s-orbitals and three 4p orbitals overlap and form d^2sp^3 type hybridization. Six electron pairs coming from strong ligand CN^- are arranged in six d^2sp^3 hybrid orbitals having same energy. Here, $[Fe(CN)_6]^{3-}$ complex ion possesses d^2sp^3 hybridization.

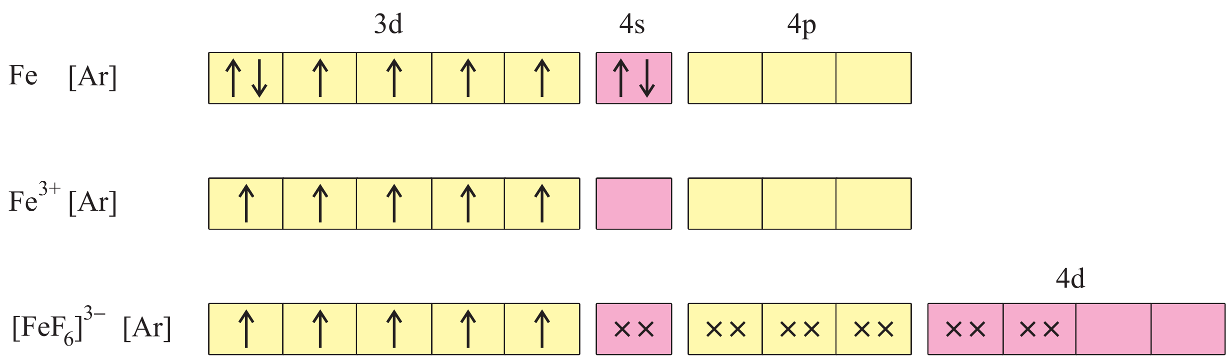


Unpaired electron n = 1 $\qquad$ d^2sp^3 hybridization

Hence μ = 1.73 BM

$[Fe(CN)_6]^{3-}$ complex ion possesses d^2sp^3 hybridization whose geometrical structure is octahedral. As there is one unpaired electron in its 3d-orbit it becomes paramagnetic and the theoretical value of magnetic moment becomes 1.73 B.M. and the experimental value is about 1.8 BM.

$[FeF_6]^{3-}$ complex ion : In hexafluorido ferrate (III) ion, six weak F^- ligands combine with Fe^{3+} metal ion. As six F^- weak ligands are combined with Fe^{3+} metal ion, this complex compound possesses sp^3d^2 hybridisation and octahedral structure. In this complex ion in 3d-orbitals, five unpaired electrons are there and so it becomes paramagnetic. Its theoretical value of magnetic moment is 5.92 B.M.

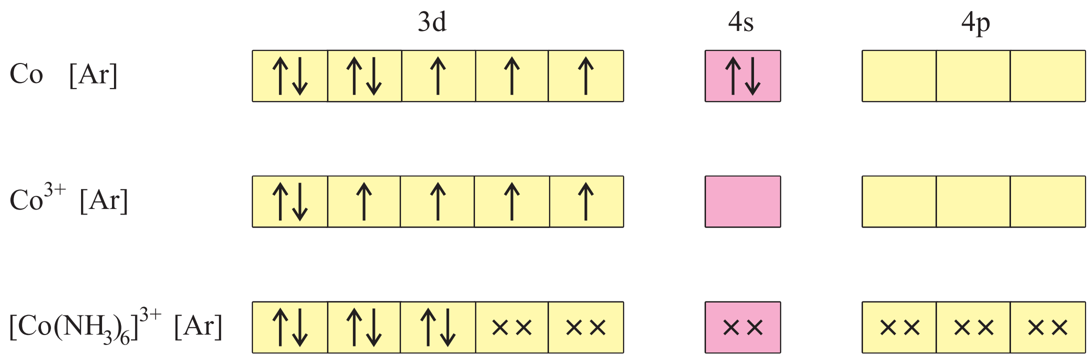


Unpaired electrons n = 5 $\qquad$ sp^3d^2 hybridization

Hence μ = 5.92 BM

If ligand had been strong, the electrons present in 3d-orbitals would have rearranged but as F^- is a weak ligand, the rearrangement of five electrons in 3d-orbital does not take place. So one 4s-orbital, three 4p-orbitals and two 4d-orbitals, aggregating six orbitals overlap and sp^3d^2 hybridization occurs. In this six sp^3d^2 hybrid orbitals having same energy, six electron pairs coming from six weak F^- ligand are arranged. **Here, $[FeF_6]^{3-}$ complex ion possesses sp^3d^2 hybridization. So, its geometrical structure is octahedral and complex is paramagnetic.** Generally, the basis of d^2sp^3 hybridisation or sp^3d^2 hybridisation is dependent on the strength of ligand.

$[Co(NH_3)_6]^{3+}$ complex ion : In hexaamminecobalt (III) complex ion the oxidation state of cobalt is +3. The electronic configuration of Co metal and Co^{3+} metal ion are $[Ar]3d^74s^2$ and $[Ar]3d^6$ respectively. Here, in Co^{3+} metal ion, the six electrons- one in 3d-orbital are paired and in four other 3d-orbitals unpaired electrons are present. When the strong ligand NH_3 gives six electron pairs to Co^{3+} metal ion, it forms six co-ordinate covalent bonds, the hybridization of Co^{3+} ion, is of d^2sp^3 and six electron pairs of six NH_3 strong ligand are arranged in six hybrid orbitals. Here, to have d^2sp^3 hybridization, two 3d-orbitals must be vacated. NH_3 being a strong ligand, there is rearrangement of six electrons in 3d-orbitals and become paired and two 3d-orbitals remain vacant. Two 3d-orbitals, one 4s-orbital and three 4p-orbitals overlap and d^2sp^3 hybridization occurs so that six hybrid orbitals having same energy get arranged octahedrally.

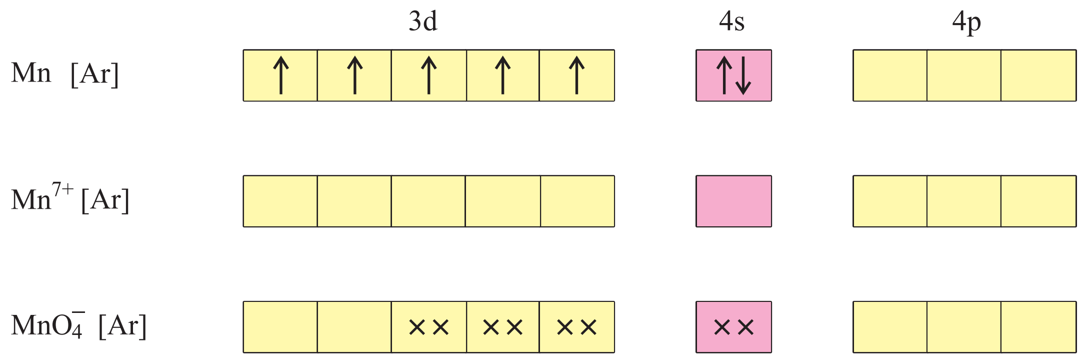


Unpaired electrons n = 0 $\qquad$ d^2sp^3 hybridization

Hence diamagnetic μ = 0 BM

Here, $[Co(NH_3)_6]^{3+}$ complex possesses d^2sp^3 hybridization and the complex becomes octahedral. As there is no unpaired electron in 3d-orbitals, the complex becomes diamagnetic.

$[MnO_4]^-$ complex ion : In tetraoxomanganate (VII) complex, four O^{2-} weak ligands are combined with Mn^{7+} metal ion, so it possesses tetrahedral structure. The electronic configuration of Mn metal and Mn^{7+} metal ion are $[Ar]3d^54s^2$ and $[Ar]3d^0$ respectively. Here 3d and 4s-orbitals are vacant. In Mn^{7+} metal ion, one 4s and three 3d-orbitals overlap and d^3s type hybridization occurs. In d^3s hybridisation, four hybrid orbitals having same energy are arranged on the four corners of tetrahedral. Four electron pairs of oxygen ions from co-ordination covalent bonds in hybrid orbitals.



d^3s hybridization.

In $[MnO_4]^-$ complex ion there is d^3s hybridization. The co-ordinate bonds formed by O^{2-} weak ligand electrons undergo d-d transition in d-orbital, so that eventhough there is no unpaired electron in Mn^{7+} of MnO_4^-, it gives coloured ion.

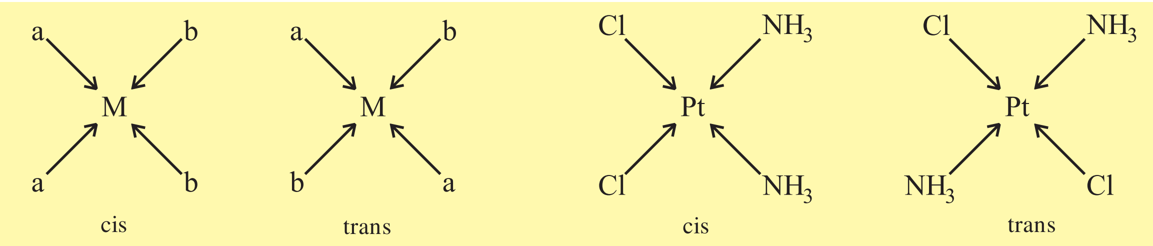
4.10 Isomerism in Complex Compounds

The complex compounds whose molecular formulae are same but the structural formulae are different, are called isomers of each other. Complex compounds possess geometrical isomerism, optical isomerism and structural isomerism.

Geometrical isomerism : Geometrical isomerism is generally observed in square planar and octahedral complex ions. If the two same ligands combined to metal ion are in nearby position to each other, then that isomer is called cis isomer and if they are in opposite positions then the isomer is called trans isomer.

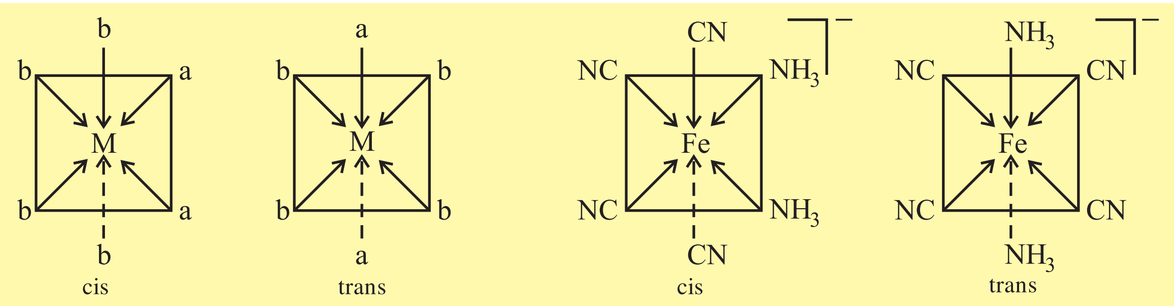
In the square planar structure if there are M metal ion and two different types of two ligands a and b combine then two geometrical isomerism cis and trans are produced

e.g. $[Pt(NH_3)_2Cl_2]$



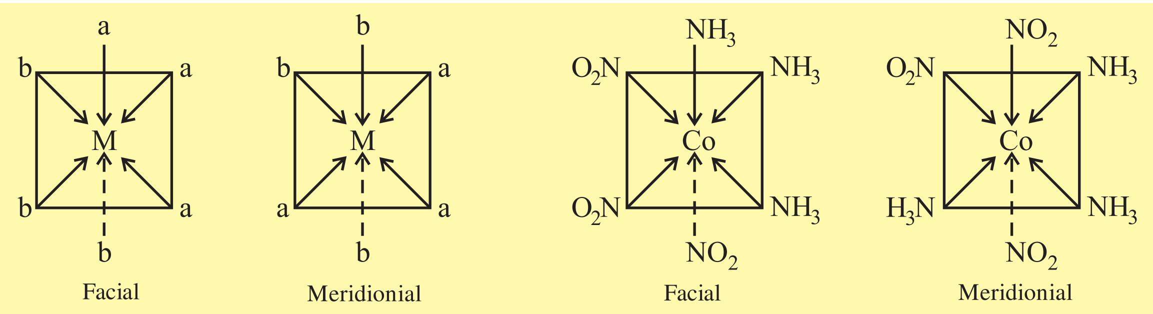
In octahedral structure, if M is metal ion and two ligands of type a and four ligands of type b are there, then two types of geometrical isomerism cis and trans are produced

e.g. $[Fe(NH_3)_2(CN)_4]^-$



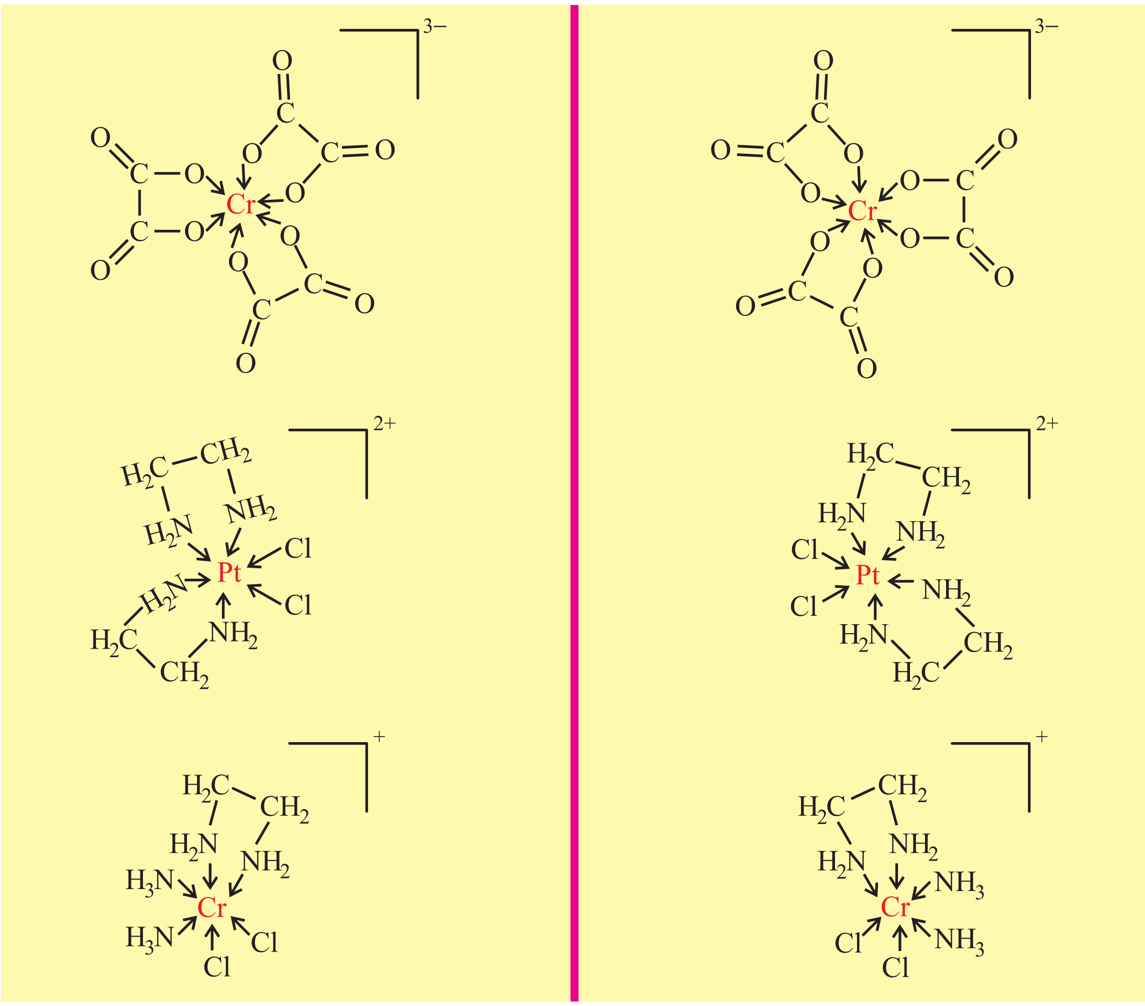
In octahedral structure, if M is metal ion and three ligands of two types a and b are present then two types geometrical isomerism-facial and meridional- are produced

e.g. $[Co(NH_3)_3(NO_2)_3]$



Optical isomerism : Optical isomerism is generally observed in octahedral complex-chelate ions. There are certain complex compounds in which molecular formula and structural formulas are same, but the two isomers produced, because of directional arrangement of ligands in it. They are mirror images to each other and the superimposition of two isomers on each other is not possible. This type of complex ions possess property of chirality and they are called optical isomers. **The main difference between these two isomers is that both the isomers rotate the plane polarized light to left or right in directions opposite to each other.** If it rotates angle of rotation on left side, then that isomer is called *l* (levo) or (–) and one which rotates angle of rotation on right side, than that isomer is called *d* (dextro) or (+). The equal proportion mixture of dextro and levo isomers is called racemic, *dl* or (±) mixture.

e.g. $[Cr(C_2O_4)_3]^{3-}$, cis $[PtCl_2(en)_2]^{2+}$, $[CrCl_2(NH_3)_2(en)]^+$



Structural isomerism : Various types of isomerism as compared to organic chemistry are observed because of different geometrical arrangements and different types of bonds in structural isomerism co-ordination. (i) Linkage isomerism (ii) Co-ordination isomerism (iii) Ionic isomerism (iv) Hydration isomerism.

(i) Linkage isomerism : NO_2^- ion combines with metal ion with nitrogen atom through co-ordination and acts as nitro ($-NO_2$) ligand or through any one of the oxygen by co-ordination and act as nitrito (ONO^-) ligand. Thus one ligand forms isomers by bonding of different atoms. Some of the examples are given below :

Downloaded from https:// www.studiestoday.com

$$= 1 - \frac{54}{504}$$

$$= 1 - 0.1071$$

$$= 0.8929$$

$$\therefore \; r \simeq 0.89$$

The value of r is near to 1. So, it can be said that there is a high degree of positive correlation between the sales and profit.

Note :

(1) The sum of difference in the ranks R_x and R_y is always zero. i.e. $\Sigma d = \Sigma(R_x - R_y) = 0$

(2) If $R_x = R_y$ for each pairs of the observations of two variables x and y then all corresponding values of d will be zero and hence $\Sigma d^2 = 0$. In this case, the value of r will be 1.

(3) If the ranks R_x and R_y are in exact reverse order of each other (see illustration 18) then $r = -1$.

Activity

Collect the information regarding the marks obtained by any ten students of your class in the subjects of Statistics and Economics. Find the correlation coefficient between the marks of two subjects using Karl Pearson's and Spearman's method and compare them.

Illustration 24 : A transport company wants to know the relation between driving experience and the number of accidents by the drivers. The sum of squares of differences in the ranks given to driving experience and the number of accidents by eight drivers is found to be 126. Find the rank correlation coefficient.

Here, $n = 8$ and the sum of squares of difference in the ranks is 126, i.e. $\Sigma d^2 = 126$.

$$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(126)}{8(64-1)}$$

$$= 1 - \frac{756}{504}$$

$$= 1 - 1.5$$

$$r = -0.5$$

Illustration 25 : Ten students selected from various schools of a district were ranked on the basis of their proficiency in Sports and General knowledge. The rank correlation coefficient obtained from the data was found to be 0.2. Later on, it was noticed that the difference in the ranks of the two attributes for one of the students was taken as 3 instead of 2. Find the correct value of rank correlation coefficient.

Here, $n = 10$

Incorrect $d = 3$

Correct $d = 2$

Now, $r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$

$\therefore \quad 0.2 = 1 - \frac{6\Sigma d^2}{10(100-1)}$

$\therefore \quad 0.2 = 1 - \frac{6\Sigma d^2}{990}$

$\therefore \quad \frac{6\Sigma d^2}{990} = 1 - 0.2$

$\therefore \quad \frac{6\Sigma d^2}{990} = 0.8$

$\therefore \quad \Sigma d^2 = \frac{0.8 \times 990}{6}$

$\therefore \quad \Sigma d^2 = 132$

Since one difference 2 is wrongly taken as 3, the corrected value of Σd^2 is obtained as follows :

Corrected $\Sigma d^2 = 132 - (\text{Wrong } d)^2 + (\text{Correct } d)^2$

$= 132 - 3^2 + 2^2$

$= 132 - 9 + 4$

$= 127$

$\therefore$ Correct value of the rank correlation coefficient is obtained as follows :

$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$

$= 1 - \frac{6(127)}{10(100-1)}$

$= 1 - \frac{762}{990}$

$= 1 - 0.7697$

$= 0.2303$

$\therefore r \simeq 0.23$

Merits and Limitations of Spearman's Rank Correlation Method

Merits :

(1) This method is easy to understand.

(2) The calculation of rank correlation method is easier than that of Karl Pearson's correlation coefficient.

(3) It is the only method when the data is qualitative.

(4) When dispersion is more or when the extreme observations are present in the data, Spearman's formula is preferred over Karl Pearson's formula.

Limitations :

(1) Since the ranks are used instead of the actual observations, there is always a loss of some information. So, this method does not provide accurate value of the correlation coefficient as compared to Karl Pearson's method.

(2) Unless the ranks are given, it is tedious to assign ranks when the number of observations is large.

(3) This method can not be used for a bivariate frequeny distribution. (In such a case, Karl Pearson's method is used and you will learn it in your higher studies.)

Exercise 2.3

1. Six companies are ranked by the two market analysts on the basis of their growth in the recent past.

Company	A	B	C	D	E	F
Rank by Analyst 1	5	2	1	4	3	6
Rank by Analyst 2	6	4	3	2	1	5

Find the rank correlation coefficient between the evaluation given by two analysts.

2. An official has ranked nine villages of a sample on the basis of the work done in the area of 'Swachhata Abhiyan' and 'Beti Bachavo Abhiyan' by the villages. The ranks are given below.

Village	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Rank for Swachhata Abhiyan	4	8	7	1	9	5	6	2	3
Rank for Beti Bachavo Abhiyan	6	8	5	1	9	7	3	4	2

Find the rank correlation coefficient between the performances of the villages in two Abhiyans.

3. The following information is obtained by a survey conducted by a town planning committee of a state.

City	A	B	C	D	E
Population (lakh)	57	45	14	18	8
Rate of growth (per thousand)	13	20	10	15	5

Find the rank correlation coefficient between the population of the cities and the rate of growth of the population.

4. The following information is obtained by taking a sample of ten students from the students of a Science college.

Student	1	2	3	4	5	6	7	8	9	10
Marks in Mathematics	39	65	62	90	82	75	25	98	36	78
Marks in Statistics	47	53	58	86	62	68	60	91	51	84

Find the rank correlation coefficient between the ability of the students in the subjects of Mathematics and Statistics.

5. From the following information of heights of husband and wife, calculate the rank correlation coefficient between their heights.

Height of husband (cms)	156	153	185	157	163	191	162
Height of wife (cms)	154	148	162	157	162	170	154

6. Two interviewers gave the following scores to the candidates on the basis of their performance in the interview. Find the rank correlation coefficient between the evaluation of two interviewers.

Candidate	A	B	C	D	E	F	G	H
Marks by first interviewer	28	44	10	28	47	35	19	40
Marks by second interviewer	32	45	25	32	41	32	24	38

7. Ten contestants are ranked in a beauty contest by two judges and the sum of squares of differences in their ranks is found to be 214. Find the rank correlation coefficient.

8. The coefficient of rank correlation of the marks obtained by 10 students in two particular subjects was found to be 0.5. Later on, it was found that one of the differences of the ranks of a student was 7 but it was taken as 3. Find the corrected value of the correlation coefficient.

*

2.9 Precautions in the Interpretation of Correlation Coefficient

The coefficient of correlation measures the strength of linear relationship between two variables. An erroneous interpretation of r may lead us to a misunderstanding about the relationship between two variables. The following are some of the points to be kept in mind as a precaution :

(1) Correlation is only a measure of strength of linear relationship between two variables. It gives no indication about presence of cause and effect relationship between them and it does not give any idea about the information that out of the two, which variable is the dependent (effect) and the other as independent (cause). The interpretation of the correlation coefficient depends very much on experience. The investigator must have thorough knowledge about the variables under consideration and the various factors which affect these variables. Several examples can be cited indicating no meaningful correlation between two variables though the value of $|r|$ is very near to 1. Generally, it happens when r is calculated without prior knowledge about cause and effect relationship between the variables. For example, the two series of data relating to the number of persons died in road accidents in a city and the price of Tuver Dal during the same period may exhibit a high correlation (i.e. r may be near to 1). But there can not be meaningful relationship between them. Therefore, this kind of correlation is known as nonsense or spurious correlation.

(2) Sometimes, due to the presence of other factors, the value of $|r|$ between given two variables may be close to 1 though two variables are not correlated. For example, the data relating to the yield of rice and sugarcane show a fairly high degree of positive correlation though there is no connection between these two variables. This may be due to the favourable effect of external factors like weather conditions, irrigation system, fertilizers etc.

(3) When $r = 0$, we can merely say that there is no linear correlation. i.e. there is a lack of linear correlation. But there may be a non linear (quadratic or any other type) relationship between the variables. e.g. :

x	–4	–3	–2	–1	1	2	3	4
y	16	9	4	1	1	4	9	16

If we calculate the Karl Pearson's coefficient of correlation for the above example then the value of r will be 0. So, we may interpret that the two variables are uncorrelated but it is a wrong interpretation. If we observe the values of two variables X and Y then we can see that they have the relation $Y = X^2$. This relation is not linear but it is quadratic. So, though there is a perfect quadratic relationship between the two variables, we get $r = 0$. So, from this example we can understand that $r = 0$ suggests a lack of linear correlation only but there may be other kind of correlation.

(4) If the correlation coefficient computed from bivariate data which is related to a given region or class or given time duration then its interpretation should be limited to that region or class or time duration only. The interpretation of r computed from such data should not be extended or generalised outside the region or class or time duration without proper verification in order to avoid any kind of misunderstanding.

e.g. If a company starts manufacturing a new product and advertises it for its sale then initially by increasing the advertisement cost, sale of the product also increases when the quality of product is good. But after some time limit, sale of the product may not increase even if its advertisement cost increases. Normally there is high degree of positive correlation between the advertisement cost and sales. During initial production period. But after some time that may not be the case. So, the interpretation that there is a high degree of positive correlation between the advertisement cost and sales can not be applied for the data outside its time period.

Some Illustrations :

Illustration 26 : Determine the value of the correlation coefficient from the following results.

$Cov(x, y) : s_x^2 = 3 : 5$ **and** $s_x : s_y = 1 : 2$

Here, $Cov(x, y) : s_x^2 = 3 : 5 \qquad \therefore \frac{Cov(x, y)}{s_x^2} = \frac{3}{5}$

and $s_x : s_y = 1 : 2 \qquad \therefore \frac{s_x}{s_y} = \frac{1}{2}$

$$\text{Now, } r = \frac{Cov(x, y)}{s_x \, s_y} = \frac{Cov(x, y)}{s_x^2} \times \frac{s_x}{s_y}$$

$$= \frac{3}{5} \times \frac{1}{2}$$

$$= \frac{3}{10}$$

$$\therefore r = 0.3$$

Illustration 27 : The following results are obtained from a bivariate data.

$n = 10$, $\Sigma(x-\bar{x})(y-\bar{y}) = 72$, $s_x = 3$ and $\Sigma(y-\bar{y})^2 = 160$ Find the correlation coefficient.

From the available results, first we shall find s_y.

$$s_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}} = \sqrt{\frac{160}{10}} = \sqrt{16} = 4$$

Now, substituting the necessary values in the following formula,

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n s_x s_y}$$

$$= \frac{72}{10(3)(4)}$$

$$= \frac{72}{120}$$

$$\therefore \quad r = 0.6$$

Illustration 28 : An educationalist has conducted an experiment to know the relation between the usage of Social Media in mobile phone and the result of the examination. A group of 10 students is selected for this and the following results were obtained regarding, the time spent x (in hours) in last week on Social Media and the marks (y) obtained out of 50 in the examination, taken immediately after it.

$\Sigma x = 133$, $\Sigma y = 220$, $\Sigma x^2 = 2344$, $\Sigma y^2 = 6500$ and $\Sigma xy = 3500$

Later on, it was found that one of the pairs of observations of X and Y was taken as (13, 20) instead of (15, 25). Find the correct value of the correlation coefficient between X and Y.

Here, $n = 10$, $\Sigma x = 133$, $\Sigma y = 220$, $\Sigma x^2 = 2344$, $\Sigma y^2 = 6500$ and $\Sigma xy = 3500$

Incorrect Pair : (13, 20)

Correct Pair : (15, 25)

Now, we find corrected values of these measures as follows :

$\Sigma x = 133 - 13 + 15 = 135$

$\Sigma x = 220 - 20 + 25 = 225$

$\Sigma x^2 = 2344 - (13)^2 + (15)^2 = 2344 - 169 + 225 = 2400$

$\Sigma y^2 = 6500 - (20)^2 + (25)^2 = 6500 - 400 + 625 = 6725$

$\Sigma xy = 3500 - (13 \times 20) + (15 \times 25) = 3500 - 260 + 375 = 3615$

Substituting these corrected values in the following formula,

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{10(3615) - (135)(225)}{\sqrt{10(2400) - (135)^2} \cdot \sqrt{10(6725) - (225)^2}}$$

$$= \frac{36150 - 30375}{\sqrt{24000 - 18225} \cdot \sqrt{67250 - 50625}}$$

$$= \frac{5775}{\sqrt{5775} \cdot \sqrt{16625}}$$

$$= \frac{5775}{\sqrt{96009375}}$$

$$= \frac{5775}{9798.4374}$$

$$= 0.5894$$

$$\therefore \quad r \simeq 0.59$$

Illustration 29 : (1) If the correlation coefficient between two variables X and Y is 0.5, find the value of the following : (i) $r\ (x,\ -y)$ (ii) $r\ (-x,\ y)$ (iii) $r\ (-x,\ -y)$

Here, $r\ (x,\ y) = 0.5$

From the property no. 5 of correlation coefficient,

(i) $r(x, -y) = -r(x, y) = -0.5$

(ii) $r(-x, y) = -r(x, y) = -0.5$

(iii) $r(-x, -y) = r(x, y) = 0.5$

(2) If $r(x, y) = 0.8$ then find $r(u, v)$ for the following.

(i) $u = x - 10$ and $v = y + 10$

(ii) $u = \frac{x-5}{3}$ and $v = 2y + 7$

(iii) $u = \frac{2x-3}{10}$ and $v = \frac{10-y}{100}$

(iv) $u = \frac{5-x}{2}$ and $v = \frac{5+y}{2}$

(v) $u = \frac{20-x}{3}$ and $v = \frac{20-y}{7}$

While defining u and v from the properties (no. 4 and no. 5), the value of $r(u, v)$ will be dependent on the signs of the coefficients of X and Y.

i.e. $r(u, v) = r(x, y)$ or $-r(x, y)$

(i) $r\ (x - 10, y + 10) = r(u, v) = 0.8$

(ii) $r\left(\frac{x-5}{3}, 2y + 7\right) = r(u, v) = 0.8$

(iii) $r\left(\frac{2x-3}{10}, \frac{10-y}{100}\right) = r(u, v) = -0.8$

(iv) $r\left(\frac{5-x}{2}, \frac{5+y}{2}\right) = r(u, v) = -0.8$

(v) $r\left(\frac{20-x}{3}, \frac{20-y}{7}\right) = r(u, v) = 0.8$

Illustration 30 : A project is conducted by the group of the students of an MBA Institute to know the relation between the results of the final year of school and final year of graduation for the students. The following information is obtained from a sample of 10 students regarding the percentage of marks in standard 12 (*x*) and the percentage of marks in the final year of graduation (*y*).

$$n = 10, \Sigma(x-65) = -2, \ \Sigma(y-60) = 2, \ \Sigma(x-65)^2 = 176, \ \Sigma(y-60)^2 = 140, \Sigma(x-65)(y-60) = 141$$

Find the correlation coefficient between the percentages of marks in Standard 12 and the final year of graduation.

Here $\Sigma(x-65) = -2 \neq 0$ $\quad\therefore\ A = 65$

$\Sigma(y-60) = 2 \neq 0$ $\quad\therefore\ B = 60$

(Here, the sum of deviations are not zero, so $65 \neq \bar{x}$ and $60 \neq \bar{y}$)

Now, let us define $u = (x-65)$ and $v = (y-60)$.

So, $\Sigma(x-65) = \Sigma u = -2, \ \Sigma(y-60) = \Sigma v = 2$

$\Sigma(x-65)^2 = \Sigma u^2 = 176, \ \Sigma(y-60)^2 = \Sigma v^2 = 140$

$\Sigma(x-65)(y-60) = \Sigma uv = 141$

Substituting the above values in the following formula,

$$r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n\Sigma u^2 - (\Sigma u)^2} \cdot \sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

$$= \frac{10(141) - (-2)(2)}{\sqrt{10(176) - (-2)^2} \cdot \sqrt{10(140) - (2)^2}}$$

$$= \frac{1414}{\sqrt{1756} \cdot \sqrt{1396}}$$

$$= \frac{1414}{\sqrt{2451376}}$$

$$= \frac{1414}{1565.6871}$$

$$= 0.9031$$

$$\therefore \quad r \simeq 0.90$$

Illustration 31 : To study the relation between the age (*X* years) of teenage children and their daily requirement of protein (*y* grams), the following information is obtained from a sample of 10 children taken by the Health Department of State.

$$\Sigma x = 140, \ \Sigma y = 150, \ \Sigma(x-10)^2 = 180, \quad \Sigma(y-15)^2 = 215, \Sigma(x-10)(y-15) = 60$$

Find the correlation coefficient between *X* and *Y*.

Here, $\bar{x} = \frac{\Sigma x}{n} = \frac{140}{10} = 14, \quad \bar{y} = \frac{\Sigma y}{n} = \frac{150}{10} = 15$

We can see that the deviations are not taken from actual mean $(\bar{x} = 14)$ for the variable X. So, to solve the example, it will be convenient to define $u = (x - A) = (x - 10)$ and $v = (y - B) = (y - 15)$.

We are given the following information.

$\Sigma(x-10)^2 = \Sigma u^2 = 180,\ \Sigma(y-15)^2 = \Sigma v^2 = 215,\ \Sigma(x-10)(y-15) = \Sigma uv = 60$

Now, in order to use an appropriate formula of r, first we need Σu and Σv.

$\Sigma u = \Sigma(x-10) = \Sigma x - \Sigma 10 = \Sigma x - n(10) = 140 - 10(10) = 140 - 100 = 40$

$\Sigma v = \Sigma(y-15) = \Sigma y - \Sigma 15 = \Sigma y - n(15) = 150 - 10(15) = 150 - 150 = 0$

$\left\{\because\ \Sigma k = \underbrace{k + k + k + \dots\dots\dots + k}_{n \text{ times}} = nk \quad \text{where, } k = \text{constant}\right\}$

Substituting the above values in the following formula,

$$r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n\Sigma u^2 - (\Sigma u)^2} \cdot \sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

$$= \frac{10(60) - (40)(0)}{\sqrt{10(180) - (40)^2} \cdot \sqrt{10(215) - (0)^2}}$$

$$= \frac{600 - 0}{\sqrt{1800-1600} \cdot \sqrt{2150-0}}$$

$$= \frac{600}{\sqrt{200} \cdot \sqrt{2150}}$$

$$= \frac{600}{\sqrt{430000}}$$

$$= \frac{600}{655.7439}$$

$$= 0.9150$$

$$\therefore \quad r \simeq 0.92$$

Illustration 32 : To know the relation between the ability in two different subjects for the students, a sample of seven students is taken from a school. From the information of marks in two subjects for 7 students, it is known that the sum of the squares of differences in the ranks of these marks is 25.5. It is also known that two students got equal marks in one subject and all the remaining marks are different. Find the rank correlation coefficient.

Here, $n = 7$ and $\Sigma d^2 = 25.5$

Two students got equal marks in a subject $(\therefore\ m = 2)$. So, we can say that there is a tie in assigning the ranks. Therefore, we need to take the term $\left(\frac{m^3 - m}{12}\right)$ only once to obtain CF.

$$CF = \left(\frac{m^3 - m}{12}\right) = \left(\frac{2^3 - 2}{12}\right) = 0.5$$

$$r = 1 - \frac{6\left[\Sigma d^2 + CF\right]}{n(n^2 - 1)}$$

$$= 1 - \frac{6[25.5 + 0.5]}{7(49 - 1)}$$

$$= 1 - \frac{6(26)}{336}$$

$$= 1 - \frac{156}{336}$$

$$= 1 - 0.4643$$

$$= 0.5357$$

$$\therefore \quad r \simeq 0.54$$

Summary

- **Correlation :** Simultaneous change in the values of two variables and direct or indirect cause-effect relationship between them.
- **Linear Correlation :** There are almost constant proportional changes in the values of two variables i.e. the points corresponding to the values of two correlated variables are on or nearer to a line.
- **Positive Correlation :** The changes in the values of two correlated variables are in the same direction.
- **Negative Correlation :** The changes in the values of two correlated variables are in the opposite direction.
- **Correlation Coefficient :** The numerical measure showing the strength of linear correlation between two variables is a correlation coefficient.
- **Scatter diagram :** A simple method for identifying linear correlation and its type (positive or negative).
- **Karl Pearson's Method :** The best method of obtaining type and strength of linear correlation using all observations.
- **Spearman's Rank Correlation Method :** A method for obtaining the correlation coefficient for qualitative variables and also preferable when dispersion is more in quantitative variables.
- The cause and effect relation between two variables cannot be proved but under the assumption that it does exist, the concept of correlation is studied.
- $r = 0$ indicates the absence of linear correlation only but there may be other type of correlation.

Chapter at a glance

Correlation

- **Linear Correlation**
  - **Methods**
    - **Scatter Diagram Method** — Only type of correlation can be known
    - **Karl Pearson's Method** — Best method to find correlation coefficient
    - **Spearman's Method** — Method for finding correlation coefficient between qualitative variables
- **Curvilinear Correlation**

List of Formulae

Karl Pearson's Method :

Correlation coefficient $= r$

(1) $$r = \frac{\text{Covariance}}{(\text{S.D of } X)(\text{S.D of } Y)} = \frac{Cov(X,Y)}{s_x \cdot s_y}$$

Where, $$Cov(X,Y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n} = \frac{\Sigma xy - n\bar{x}\,\bar{y}}{n}$$

$$s_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \quad \text{and} \quad s_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$$

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

(3) $$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

(4) $$r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n\Sigma u^2 - (\Sigma u)^2} \cdot \sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$ Where, $u = x - A$ or $\frac{x-A}{c_x}$, $v = y - B$ or $\frac{y-B}{c_y}$

(5) $$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n \cdot s_x \cdot s_y}$$

(6) $$r = \frac{\Sigma xy - n\bar{x}\,\bar{y}}{n \cdot s_x \cdot s_y}$$

(5) and (6): Specially for short sums

Spearman's Rank Correlation Method

(7) $$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$ When the observations are not repeated

(8) $$r = 1 - \frac{6\left[\Sigma d^2 + CF\right]}{n(n^2-1)}$$ When some of the observations are repeated

Where, d = Rank of x – Rank of y = $R_x - R_y$

CF = Correction Factor $= \Sigma\left(\frac{m^3-m}{12}\right)$

m = Number of times an observation is repeated

Exercise 2

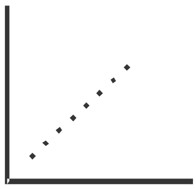
Section A

Find the correct option for the following multiple choice questions :

1. In context with correlation, what do you call the graph, if the points of paired observations (x,y) are shown in a graph ?

(a) Histogram (b) Circle diagram (c) Scatter diagram (d) Frequency curve

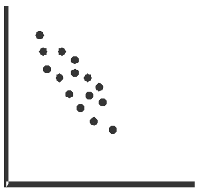
2. Which kind of the correlation exists if the following scatter diagram is of two variables X and Y ?



(a) Perfect Positive correlation (b) Partial Positive correlation

(c) Perfect Negative correlation (d) Partial Negative correlation

3. Which kind of the correlation exists if the following scatter diagram is of two variables X and Y ?



(a) Perfect Positive correlation (b) Partial Positive correlation

(c) Perfect Negative correlation (d) Partial Negative correlation

4. What is the value of r, if all the points plotted in a scatter diagram lie on a single line only ?

(a) 0 (b) 1 or –1 (c) 0.5 (d) –0.5

5. What is the range of the correlation coefficient r ?

(a) $-1 < r < 1$ (b) 0 to 1 (c) $-1 \le r \le 1$ (d) –1 to 0

6. The measurement unit of a variable 'Weight' is kg. and that of 'Height' is cm. What can you say about the measurement unit of the correlation coefficient between them ?

(a) kg (b) cm (c) km (d) does not have any unit

7. Which kind of the correlation can be obtained if the two variables are varying in opposite direction in constant proportion ?

(a) Partial Positive Correlation (b) Perfect Negative Correlation

(c) Perfect Positive Correlation (d) Partial Negative Correlation

8. What does the numerator indicate in the formula for calculating the correlation coefficient by Karl Perason's method ?

(a) Product of variance of X and Y (b) Covariance of X and Y

(c) Variance of X (d) Variance of Y

9. Which of the following values is not possible as a value of r ?

(a) 0.99 (b) –1.07 (c) –0.85 (d) 0

10. If $u = \frac{x-A}{c_x}$ and $v = \frac{y-B}{c_y}$, $c_x > 0$, $c_y > 0$ then which of the following statement is correct ?

(a) $r(x, y) \neq r(u, v)$ (b) $r(x, y) > r(u, v)$ (c) $r(x, y) = r(u, v)$ (d) $r(x, y) < r(u, v)$

11. If $r(x, y) = 0.7$ then what is the value of $r(x + 0.2, y + 0.2)$?

(a) 0.7 (b) 0.9 (c) 1.1 (d) –0.7

12. If $r(-x, y) = -0.5$ then what is the value of $r(x, -y)$?

(a) 0.5 (b) –0.5 (c) 1 (d) 0

13. What is the value of the rank correlation coefficient if $\Sigma d^2 = 0$?

(a) 0 (b) –1 (c) 1 (d) 0.5

14. In the method of rank correlation, in usual notations if $R_x = R_y$ for each pair of observations then what is the vlue of the r ?

(a) 0 (b) –1 (c) 1 (d) 0.1

15. In the method of rank correlation, what is the sum of differences of the ranks of two variables ?

(a) 0 (b) –1 (c) 1 (d) Any real number

16. In the method of rank correlation, if the ranks of two variables are exactly in reverse order then what is the value of r ?

(a) $r = 0$ (b) $r = -1$ (c) $r = 1$ (d) $r = 0.1$

17. In usual notations, which term is added in Σd^2 for each repeated observation in the rank correlation ?

(a) $\frac{m^2-1}{12}$ (b) $\frac{m^3-m}{12}$ (c) $\frac{6m^3-m}{12}$ (d) $n(n^2-1)$

18. Which kind of correlation will you get between the number of units sold and its revenue at constant price ?

(a) Perfect Positive (b) Partial Positive (c) Perfect Negative (d) Partial Negative

Section B

Answer the following questions in one sentence :

1. Define correlation.

2. Define correlation coefficient.

Identify, whether there is a positive correlation or negative correlation between the following pairs of variables (Question 3 to Question 6).

3. The age of an adult person and life insurance premium at the time of taking an insurance under a plan.

4. The sales and profit of last five years for a mostly accepted product of a company.
5. The rate of inflation and the purchase power of common man of a country when income of the common man is stable.
6. Altitude and amount of Oxygen in air.
7. What can be said about the correlation between the annual import of crude oil and the number of marriages during the same time period ?
8. The correlation coefficient between X and Y is 0.4. What will be the value of correlation coefficient if 5 is added in each observation of X and 10 is substracted from each observation of Y ?
9. What is the main limitation of scatter diagram method ?
10. If the value of $n(n^2-1)$ is six times the value of Σd^2 then what is the value of r ?
11. What will be the sign of r if the value of the covariance is negative ?

Section C

Answer the following questions :

1. Explain the meaning of positive correlation with an illustration.
2. Explain the meaning of negative correlation with an illustration.
3. Write the assumptions of Karl Pearson's method.
4. Define : Scatter Diagram.
5. What is spurious correlation ?
6. Explain the cause and effect relationship.
7. Explain : Perfect positive correlation
8. Explain : Perfect negative correlation
9. When is it necessary to use rank correlation ?
10. In which situation, the values of Karl Pearson's correlation coefficient and Spearman's rank correlation coefficient are equal ?
11. Find the value of r if $Cov(x, y)=120,\ s_x=12,\ s_y=15$.
12. Find the value of r if $\Sigma(x-\bar{x})(y-\bar{y})=-65,\ s_x=3,\ s_y=4$ and $n=10$.
13. For 10 pairs of observations, $\Sigma d^2=120$. Find the value of the rank correlation coefficient.

Section D

Answer the following questions :

1. Explain scatter diagram method.
2. Write merits and limitations of scatter diagram method.
3. Write the properties of correlation coefficient.
4. Write the merits and limitations of Karl Pearson's method.
5. Interpret $r=1$, $r=-1$ and $r=0$.
6. Explain Spearman's Rank correlation method.
7. Write merits and limitations of Spearman's rank correlation method.
8. How would you interpret partial correlation ?
9. State the necessary precautions to be taken while interpreting the value of correlation coefficient.
10. The following data is available for two variables rainfall in mm. (X) and yield of crop Qtl/ Hectare (Y).

    $n=10$, $\bar{x}=120$, $\bar{y}=150$, $s_x=30$, $s_y=40$ and $\Sigma xy=189000$. Find the correlation coefficient.
11. The following information is obtained for 9 pairs of observations.

    $\Sigma x=51$, $\Sigma y=72$, $\Sigma x^2=315$, $\Sigma y^2=582$, $\Sigma xy=408$. Find the correlation coefficient.
12. The information obtained on the basis of ranks given by two judges to eight contestants of a dance competition is given below.

    $\Sigma\left(R_x-R_y\right)^2=126$

    Where R_x and R_y are the ranks given to a contestant by the two judges respectively. Find Spearman's rank correlation coefficient.
13. The ranks given by two experts on the basis of interviews of five candidates for a job are (3, 5), (5, 4), (1, 2), (2, 3) and (4, 1). Find the rank correlation coefficient from this data.

Section E

Solve the following :

1. The following information is obtained to study the relation between the selling price of nose mask and its demand during an epidemic.

Price (₹)	38	45	40	42	35
Demand (units)	103	92	97	98	100

Find the correlation coefficient between the price and demand of mask by Karl Pearson's method.

2. In order to study the relationship between the abilities in the subjects of Human Resource Management and Personality Development for the students of a post graduate level course, a sample of 5 students is taken and the following information is obtained.

Student	1	2	3	4	5
Marks in HRM	45	25	40	20	45
Marks in PD	47	23	17	35	48

Calculate the Karl Pearson's correlation coefficient between the marks of both the subjects.

3. A vendor wants to display lipsticks of different brands according to their popularity. For that, he invites two experts Preyal and Nishi to rank the lipsticks of different brands.

Lipstick	A	B	C	D	E	F	G
Rank by Preyal	5	6	7	1	3	2	4
Rank by Nishi	5	7	6	2	1	4	3

Find the rank correlation coefficient to know the smilarity in the decision of both the experts.

4. A merchant wants to study the relation between prices of tea and coffee in Ahmedabad city. He obtains the following information about prices of tea and coffee of the last six months.

Price per kg for tea (₹)	340	370	450	320	300	360
Price per 100 grams for coffee (₹)	190	215	200	180	163	175

Calculate the rank correlation coefficient between the price of tea and coffee.

5. The demand of an imported fruit in a local market is very uncertain. To know the relation between the price of the fruit and its supply, a vendor collects the information about the average price and supply for last ten months.

Average price per unit (₹)	65	68	43	38	77	48	35	30	25	50
Supply (hundred units)	52	53	42	60	45	41	37	38	25	27

Find the rank correlation between the average price and the supply.

6. To know the relation between the results of the examinations taken in a span of short time, a teacher has conducted two examinations in last two weeks and the ranks obtained by seven students are as follows.

Student	A	B	C	D	E	F	G
Rank in Test 1	5	1	2	3.5	3.5	7	6
Rank in Test 2	7	1	4	6	5	3	2

Find the rank correlation coefficient to know the similarity between the results of two examinations.

Section F

Solve the following :

1. The information of fertilizer used (in tons) and productivity (in tons) of eight districts is given below.

Fertilizer (tons)	15	18	20	25	29	35	40	38
Productivity (tons)	85	93	95	105	115	130	140	145

Calculate the correlation coefficient by Karl Pearson's method.

2. Find the Karl Pearson's correlation coefficient from the following information of the average weekly hours spent on Video games and the grade points obtained in an examination by 6 children of a big city.

Weekly average hours spent for Video games	43	47	45	50	40	51
Grade points obtained in an examination	5.2	4.9	5.0	4.7	5.4	4.3

3. Find Karl Pearson's correlation coefficient between density of population (per square km) and death rate (per thousand) from the following data.

City	A	B	C	D	E	F	G
Density (per sq. km)	750	600	350	500	200	700	850
Death rate (per thousand)	30	20	15	20	10	25	50

4. The following information is obtained to study the relationship between the advertisement cost and the sales of electric fans of the companies manufacturing electric fans. Find the correlation coefficient between advertisement cost and the sales by Karl Pearson's method.

Company	A	B	C	D	E	F
Advertisement Cost (lakh ₹)	140	120	80	100	80	180
Sales of electric fans (crore ₹)	35	45	15	40	20	50

5. A doctor obtains the following information for the weights of seven mothers and their children from a maternity home for his research to know the relation between the weights of mother and weights of their children at the time of birth.

Weight of mother (kg)	59	72	66	64	77	66	60
Weight of child (kg)	2.5	3.4	3.1	2.7	2.8	2.3	3.0

Find rank correlation coefficient between the weights of mother and child.

6. The following data is obtained to know the relation between maximum day temperature and the sale of ice-cream in Ahmedabad city.

Maximum Temperature (Celsius)	35	42	40	39	44	40	45	40
Sale of ice cream (kg)	600	680	750	630	920	750	900	720

Calculate the rank correlation coefficient.

7. An entrance test required to study abroad is conducted online. The marks obtained in Reasoning Ability and English Speaking in this online test (having negative marking system for wrong answer) by 5 students selected in a sample are given below.

Student	A	B	C	D	E
Marks in Reasoning Ability	5	5	5	5	5
Marks in English Speaking	2	–2	–2	0	2

Find the rank correlation coefficient between Reasoning Ability and ability in English Speaking.

8. Six dancers A, B, C, D, E and F in a dance competition were judged by two dance Gurus. The ranks assigned to the dancers are as follows.

Rank	1	2	3	4	5	6
By Guru 1	B	F	A	C	D	E
By Guru 2	F	A	C	B	E	D

Find the rank correlation coefficient between the judgement of the two Gurus.

9. The following data is obtained for two variables, inflation (X) and interest rate (Y).

$$n = 50,\ \Sigma x = 500,\ \Sigma y = 300,\ \Sigma x^2 = 5450,\ \Sigma y^2 = 2000,\ \Sigma xy = 3090$$

Later on, it was known that one pair of observation (10, 6) was included additionally by mistake. Find the correlation coefficient by excluding this pair of observations.

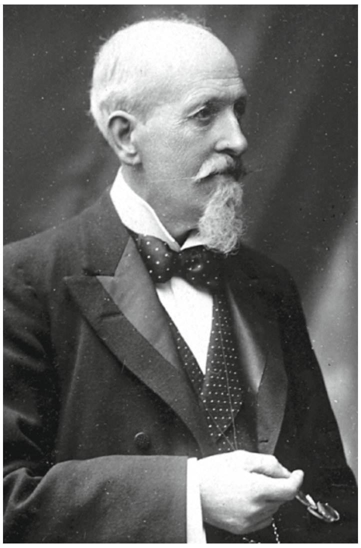
10. The information regarding sales (X) and expenses (Y) of 10 firms is given below.

$$\bar{x} = 58,\ \bar{y} = 14,\ \Sigma(x-65)^2 = 850,\ \Sigma(y-13)^2 = 32,\ \Sigma(x-65)(y-13) = 0$$

Find the correlation coefficient.

11. Daily calorie intake of ten persons is X and their weight is Y kg. The rank correlation coefficient from this information is 0.6. On subsequent verification, it was noticed that the difference of ranks of X and Y for one of the persons was taken as 2 instead of 4. Find the correct value of rank correlation coefficient.

12. The information of health index x and life expectancy y is obtained for 10 people. These data are ranked to find the rank correlation coefficient and the sum of squares of the ranks was found to be 42.5. It was also observed that health index 70 was repeated three times and life-expectancy 45 was repeated twice in the data. Find the rank correlation coefficient using this information.



Charles Edward Spearman
(1863 –1945)

Charles Edward Spearman was an English psychologist known for work in statistics, as a pioneer of factor analysis and for Spearman's rank correlation coefficient. He also did seminal work on models for human intelligence, including his theory that disparate cognitive test scores reflect a single General intelligence factor and coining the term g-factor.

After serving army for 15 years, he went on to study for a Ph.D. in experimental psychology. Spearman joined University College London and stayed there until he retired in 1931. Initially he was Reader and head of the small psychological laboratory. In 1911 he was promoted to the Grote professorship of the Philosophy of Mind and Logic. His title changed to Professor of Psychology in 1928 when a separate Department of Psychology was created.

His many published papers cover a wide field, but he is especially distinguished by his pioneer work in the application of mathematical methods to the analysis of the human mind and his original studies of correlation in this sphere.

•

"Prediction is very difficult, especially about the future."

– Niels Bohr

3

Linear Regression

Contents :

3.1 Introduction

In the previous chapter 2, we have studied the concept of correlation. We have seen whether the correlation between two variables is positive or negative is known by the correlation coefficient. Moreover, we get numerical measure of the closeness of the variables. But the coefficient of correlation fails to provide the expected value of one variable for the given value of the other variable. When some relation exists between two variables, many times it is necessary to obtain the approximate or estimated value of one variable for a known value of the other variable using this relation.

e.g. We know that there is a correlation between advertisement cost and sale of an item. Now, for some given amount of advertisement cost, if we want to know the corresponding expected sale then it is not possible to obtain it only by correlation. For this, it is necessary to use the concept of regression.

The literal meaning of regression is 'to avert' or 'return to the mean value'. The term regression was first used by a statistician Sir Francis Galton during his study of human inheritance. He had collected the information about the height of 1000 pairs of fathers and sons. He revealed the following interesting results.

(i) Tall fathers have tall sons and short fathers have short sons.

(ii) The average height of sons of a group of tall fathers is less than average height of group of tall fathers.

(iii) The average height of sons of a group of short fathers is greater than average height of group of short fathers.

So, it is clear from the above findings that the heights of sons show regressive tendency with respect to the height of their fathers. The existence of this tendency restricts the humans to split into two races of pigmies and giants. So, Sir Francis Galton has given the name regression to describe such relation.

Regression is a functional relation between two correlated variables. We shall study the concept of regression under the assumption that there exists a cause-effect relationship between two variables.

3.2 Linear Regression Model

A set of one or more equations representing a relation or a problem is called a model. A statistical model which describes the cause and effect relationship between two variables is called a regression model. Generally, out of two variables having cause-effect relationship, the causal variable is denoted by X. We shall call this variable as independent or explanatory variable and effect variable is denoted by Y. We shall call this variable as dependent or explained variable. Let us understand the meaning of independent variable and dependent variable from the following illustrations :

(i) In case of 'advertisement cost' and 'sales', generally, because of increase (decrease) in the 'advertisement cost', corresponding 'sales' also increases (decreases), so we shall take 'advertisement cost' as independent variable X and 'sales' as dependent variable Y.

(ii) In case of 'rainfall' and 'yield of rice' in some region, it is very clear that 'yield of rice' depends on 'rainfall'. So, we shall take 'rainfall' as independent variable X and 'yield of rice' as dependent variable Y.

In a regression model, the dependent variable Y is expressed in the form of an appropriate mathematical function of the independent variable X.

Now, we shall define a linear regression model as follows.

$$Y = \alpha + \beta X + u$$

Where, Y = Dependent Variable

X = Independent Variable

α = Constant

β = Constant

u = Disturbance Variable of the Model

The inadequacy of the linearity between two variables X and Y is shown by u. The perfect linear relation is possible in natural science like mathematics. So, the disturbance variable u obviously becomes 0 in such a case. In other words, when there is a perfect linear correlation between two variables X and Y then the regression model is $Y = \alpha + \beta X$. But we know that exact linear relation between the variables is not always possible in business, economics and social science as these correlated variables are also affected by other factors. Thus, when there is a partial correlation between the variables X and Y then the linear regression model is $Y = \alpha + \beta X + u$. From the above discussion, we can define linear regression in simple words as follows.

"A mathematical or functional relationship between two correlated variables which helps in estimating the value of dependent variable for some given (known) value of independent variable is called **Linear Regression.**''

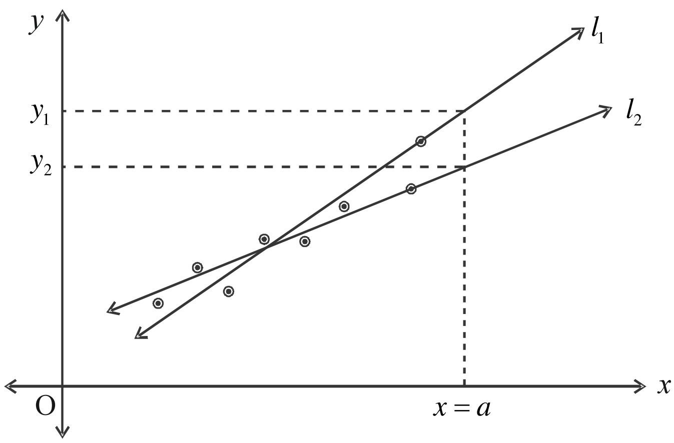
3.3 Fitting of Regression Line

In a scatter diagram of two correlated variables, if the points are clustered around a line, we can say that there is a linear regression. The method of obtaining such a line expressing the relation between two variables is called fitting of a regression line.

There are two methods for fitting a regression line : (1) Method of Scatter Diagram (2) Method of Least Squares.

3.3.1 Method of Scatter Diagram

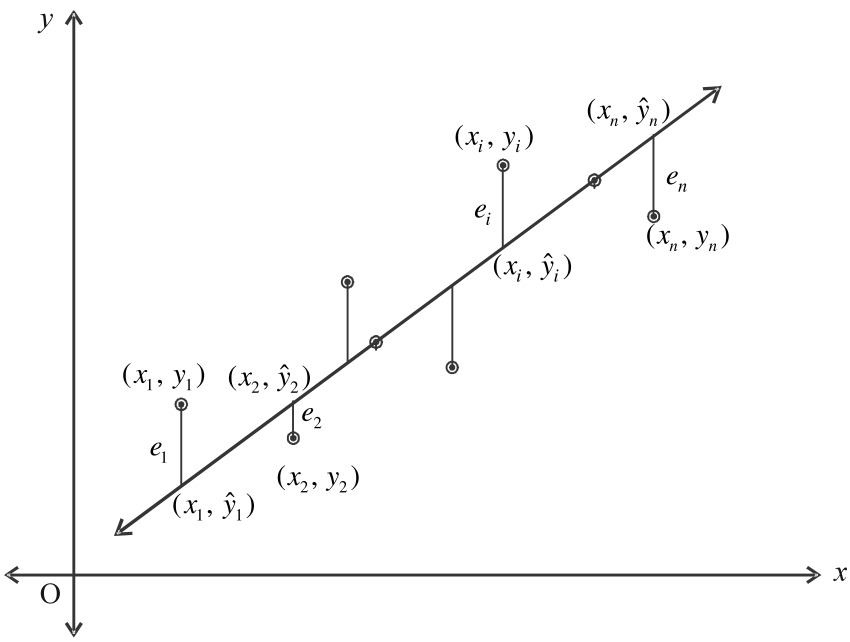
Suppose n ordered pairs of observations of two correlated variables X and Y are $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Using this data, we draw a scatter diagram. Now, a line is drawn in such a way that it is close to almost all the points of the scatter diagram. If Y is a dependent variable and X is an independent variable then such a line is called regression line of Y on X and an approximate value of dependent variable Y can be obtained for any given value of independent variable X from it. Since no computation is required to draw such a line, it is very easy and quick method of fitting a regression line. But there is a problem in doing so. Different persons may draw different lines. As a result, different persons may provide different estimates of the dependent variable Y for the same value of independent variable X. It can be seen very easily from the following scatter diagram.



Two different persons have drawn two different lines l_1 and l_2 in the following scatter diagram of the same data. We can see that for some value 'a' of independent variable X, corresponding estimated value is 'y_1' from line l_1 and it is 'y_2' from the line l_2. Thus we get different estimates for dependent variable Y from different lines for a single value of independent variable X. So, it can be said that this method is subjective. A line of regression drawn by this method is not the best fitted line because it does not guarantee the best estimate of the dependent variable. The method of least square is used to obtain such a best fitted regression line.

3.3.2 Method of Least Squares

Suppose n ordered pairs of observations of two correlated variables X (independent variable) and Y (dependent variable) are $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. We shall draw a scatter diagram for this data to understand the method of least squares.



If an equation of the best fitted line describing the linear regression between the variables X and Y is $\hat{y} = a + bx$ then the constants a and b of this line can be obtained by the method of least squares as follows.

Let the estimated values of variable Y corresponding to values $x_1, x_2, x_3, ..., x_n$ of variable X are $\hat{y}_1, \hat{y}_2, \hat{y}_3, ... \hat{y}_n$ from the line and the corresponding observed values of Y are $y_1, y_2, y_3, ... y_n$ respectively. Now, for some $X = x_i$, estimated value of Y from the line is $\hat{y}_i = a + bx_i$. The vertical distance (i.e. distance parallel to Y-axis) between observed value y_i and the estimated value $\hat{y}_i$ is called an error in the estimation. It is denoted by e_i.

$$\therefore\ e_i = y_i - \hat{y}_i = y_i - (a + bx_i) = y_i - a - bx_i$$

Where, $i = 1, 2, 3, ..., n$

Obviously, the error will be positive for points above the line the error will be negative for points below the line and it will be zero for the points which are on the line.

Now, the values of constants a and b of the fitted line $\hat{y} = a + bx$ (known as regression line of Y on X) are obtained in such a way that the sum of the squares of the errors is minimum.

i.e. $\Sigma e_i^2 = \Sigma(y_i - \hat{y}_i)^2 = \Sigma(y_i - a - bx_i)^2$ is minimum.

By ignoring the suffix i for convenience, we can get such values of a and b by a simple algebraic method, which are as follows.

$$b = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$= \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

And

$$a = \bar{y} - b\bar{x}$$

The line $\hat{y} = a + bx$ obtained by this method is a line passing as close as possible to the points of scatter diagram. The sum of squares of the errors is minimised while obtaining the regression line. Therefore, this method is called **'method of least squares'.**

The value of b obtained by this method is called the regression coefficient of the regression line of Y on X. b is also called slope of the regression line and the constant a is called intercept of the regression line.

Interpretation of regression coefficient *b*

b = the estimated change in the value of Y for a unit change in the value of X.

i.e. when $b > 0$, it means that a unit increase in the value of independent variable X implies an estimated increase of b units in the value of dependent variable Y.

when $b < 0$, it means that a unit increase in the value of independent variable X implies an estimated decrease of $|b|$ units in the value of dependent variable Y.

Note that the regression line obtained by the method of least squares is also known as the line of best fit.

Note : (1) The regression coefficient b can also be denoted by b_{yx}. If not required, generally we shall denote regression coefficient by b only.

(2) If all the points in a scatter diagram are on one line only then error will be zero for all the points. Hence, the estimated value $\hat{y}$ is same as its observed value y. So, the form of the regression line will be $y = a + bx$ in place of $\hat{y} = a + bx$. Naturally, in this situation r is 1 if $b > 0$ and r is -1 if $b < 0$.

Additional Information for understanding

Generally, only 'fitted line' is mentioned for the regression line obtained instead of 'best fitted line'.

Now, let us take some examples to obtain a regression line.

Illustration 1 : The following observations are obtained for life (years of usage) of cars and their average annual maintenance costs of a specific model of car of a particular company.

Life of cars (years)	2	4	6	8
Average annual maintenance cost (thousand ₹)	10	20	25	30

Obtain the regression line of maintenance cost on the life of cars. Also, estimate the maintenance cost if the life of a car is 10 years.

'Life of a car' is an independent variable. So, we shall denote it by variable X and 'maintenance cost' is dependent variable. So, we shall denote it by Y. Considering at the data, we shall prepare the following table for obtaining the regression line.

	Life of car (years) x	**Maintenance cost (thousand ₹) y**	xy	x^2
	2	10	20	4
	4	20	80	16
	6	25	150	36
	8	30	240	64
Total	**20**	**85**	**490**	**120**

$\bar{x} = \frac{\Sigma x}{n} = \frac{20}{4} = 5, \ \bar{y} = \frac{\Sigma y}{n} = \frac{85}{4} = 21.25$

Let us find the regression coefficient as follows.

$$b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{4(490) - (20)(85)}{4(120) - (20)^2}$$

$$= \frac{1960 - 1700}{480 - 400}$$

$$= \frac{260}{80}$$

$$= 3.25$$

$\therefore \ b = 3.25$

Now, putting the values of $\bar{x}, \bar{y}$ and b in the formula of a,

$$a = \bar{y} - b\bar{x}$$

$$= 21.25 - 3.25(5)$$

$$= 21.25 - 16.25$$

$\therefore \ a = 5$

So, the regression line of 'maintenance cost' (Y) on 'life of car' (X) is

$$\hat{y} = a + bx$$

$\therefore \ \hat{y} = 5 + 3.25\,x$

Putting $X = 10$,

$$\hat{y} = 5 + 3.25(10)$$

$$= 5 + 32.5 = 37.5$$

$\therefore \ \hat{y} = 37.5$

So, when the life of a car is 10 years then its estimated maintenance cost is ₹ 37.5 thousand **Note :** Since $b = 3.25$, we can say that every year (one unit change in X), the maintenance cost of the car increases by approximately ₹ 3.25 thousand (change in Y).

Illustration 2 : The monthly sale of different types of laptops (in hundred units) and its profit (in lakh ₹) for the last six months for a company is given below.

Month	1	2	3	4	5	6
No. of laptops sold (hundred units) x	5	7	5	12	8	3
Profit (lakh ₹) y	8	9	10	15	10	6

Obtain the regression line of *Y* on *X*. Also find the error in estimating Y for *X* = 7.

	No. of laptops sold (hundred units) x	Profit (lakh ₹) y	xy	x^2
	5	8	40	25
	7	9	63	49
	5	10	50	25
	12	15	180	144
	8	10	80	64
	3	6	18	9
Total	**40**	**58**	**431**	**316**

$\bar{x} = \frac{\Sigma x}{n} = \frac{40}{6} = 6.67$; $\bar{y} = \frac{\Sigma y}{n} = \frac{58}{6} = 9.67$

Let us find the regression coefficient b as follows.

$$b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{6(431) - (40)(58)}{6(316) - (40)^2}$$

$$= \frac{2586 - 2320}{1896 - 1600}$$

$$= \frac{266}{296}$$

$$= 0.8986$$

$$\simeq 0.90$$

$\therefore b \simeq 0.90$

By putting the values of $\bar{x}, \bar{y}$ and b in the formula of a,

$a = \bar{y} - b\bar{x}$

$= 9.67 - 0.90(6.67)$

$= 9.67 - 6.003$

$= 3.667$

$\therefore a \simeq 3.67$

So, regression line of Y on X is

$\hat{y} = a + bx$

$\therefore \hat{y} = 3.67 + 0.9x$

Now, to find the error for $X = 7$, first we obtain the estimated value of Y corresponding to it.

Putting $X = 7$,

$\hat{y} = 3.67 + 0.9(7)$

$= 3.67 + 6.3$

$\therefore\ \hat{y}$ = ₹ 9.97 lakh

Now, we can see from the available data that observed value of Y when $X = 7$ is 9.

$\therefore$ Error $e = y - \hat{y}$

$= 9 - 9.97$

$\therefore\ e$ = ₹ -0.97 lakh

Illustration 3 : In order to study the relationship between the repairing time of accident damaged cars and the cost of repair, the following information is collected.

Repairing time of a car (man hours)	32	40	25	29	35	43
Repairing cost (thousand ₹)	25	35	18	22	28	46

Obtain the regression line of Y (repairing cost) on X (repairing time). If the time taken to repair a car is 50 hours, find an estimate of the repairing cost.

Here, $n = 6, \bar{x} = \frac{\Sigma x}{n} = \frac{204}{6} = 34$ અને $\bar{y} = \frac{\Sigma y}{n} = \frac{174}{6} = 29$

	Repairing time (man hours) x	**Repairing cost (thousand ₹) y**	$x - \bar{x}$	$y - \bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
	32	25	–2	–4	8	4
	40	35	6	6	36	36
	25	18	–9	–11	99	81
	29	22	–5	–7	35	25
	35	28	1	–1	–1	1
	43	46	9	17	153	81
Total	**204**	**174**	**0**	**0**	**330**	**228**

$b = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

$= \frac{330}{228}$

$= 1.4474$

$\simeq 1.45$

$\therefore\ b \simeq 1.45$

Now, by putting the value of $\bar{x}$, $\bar{y}$ and b in the formula of a,

$a = \bar{y} - b\,\bar{x}$

$= 29 - 1.45(34)$

$= 29 - 49.3$

$\therefore\ a = -20.3$

So, the regression line of Y on X is

$$\hat{y} = a + bx$$

$$\therefore \quad \hat{y} = -20.3 + 1.45x$$

Putting $X = 50$,

$$\hat{y} = -20.3 + 1.45(50)$$

$$= -20.3 + 72.5$$

$$\therefore \quad \hat{y} = 52.2$$

So, when the repairing time is 50 hours, the estimated repairing cost is ₹ 52.2 thousand.

Exercise 3.1

1. From the following data of price (in ₹) and demand (in hundred units) of a commodity, obtain the regression line of demand on price. Also estimate the demand when price is 20 ₹.

Price (₹)	12	14	15	16	18	21
Demand (hundred units)	18	12	10	8	7	5

2. To study the relationship between the time of usage of cars and its average annual maintenance cost, the following information is obtained :

Car	1	2	3	4	5	6
Time of usage of a car (years) *x*	3	1	2	2	5	3
Average annual maintenance cost (thousand ₹) *y*	10	5	8	7	13	8

Obtain the regression line of Y on X. Find an estimate of average annual maintenance cost when the usage time of a car is 5 years. Also find its error.

3. The information for a year regarding the average rainfall (in cm) and total production of crop (in tons) of five districts is given below :

Average rainfall (cm)	25	32	38	29	31
Crop (tons)	84	90	95	88	93

Find the regression line of production of crop on rainfall and estimate the crop if average rainfall is 35 cm.

4. The following data gives the experience of machine operators and their performance ratings.

Operator	1	2	3	4	5	6	7	8
Experience (years) *x*	12	5	10	3	18	4	12	16
Performance rating *y*	83	75	80	78	89	68	88	87

Calculate the regression line of performance ratings on the experience and estimate the performance rating of an operator having 7 years of experience.

*

3.4 Utility of the Study of Regression

The following are some utilities of regression :

(1) We can determine a functional relation between two correlated variables.

(2) Once the functional relation is established, it can be used to predict the unknown value of dependent variable Y on the basis of known value of independent variable X.

(3) We can determine the approximate change in the value of dependent variable Y for a unit change in the value of independent variable X.

(4) We can determine the error in the estimation of dependent variable obtained by a regression line.

Regression is very useful for economists, planners, businessmen, administrators, researchers, etc.

Short-cut Method for computing Regression coefficient

When the values of X and Y are relatively large and / or fractional, it is difficult to calculate the terms like x^2, xy, etc. In such cases, an alternative formula can be used. It is based on the following property of regression coefficient.

Property : The regression coefficient is independent of change of origin but not of change of scale.

If $b = b_{yx}$ is a regression coefficient of a regression line of Y on X then using the above property, the following formulae can be written for the regression coefficient by short-cut method.

(1) If $u = x - A$ and $v = y - B$ then

$$b = b_{yx} = b_{vu} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2}$$

(2) If $u = \frac{x-A}{c_x}$ and $v = \frac{y-B}{c_y}$ then

$$b = b_{yx} = b_{vu} \cdot \frac{c_y}{c_x} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2} \times \frac{c_y}{c_x}$$

Here, A, B, c_x and c_y are constants and $c_x > 0, c_y > 0$.

Illustration 4 : In order to determine the relationship between monthly income (in thousand ₹) and monthly expenditure (in thousand ₹) of people of a group, a sample of seven persons is taken from that group and the following information is obtained.

Person	1	2	3	4	5	6	7
Monthly income (thousand ₹)	60	70	64	68	62	65	72
Monthly expenditure (thousand ₹)	50	59	57	50	53	58	60

Obtain the regression line of monthly expenditure on monthly income. If a person of the group has monthly income of ₹ 75 thousand, estimate his monthly expenditure.

Since the regression line of monthly expenditure on monthly income is to be obtained, we shall take 'monthly expenditure' as variable Y and 'monthly income' as variable X.

Here, $\bar{x} = \frac{\Sigma x}{n} = \frac{461}{7} = 65.86$ and $\bar{y} = \frac{\Sigma y}{n} = \frac{387}{7} = 55.29$

So, by taking $A = 65$ and $B = 55$, we can define u and v as follows.

$u = x - A = x - 65$ and $v = y - B = y - 55$

	Monthly income (thousand ₹) x	Monthly expenditure (thousand ₹) y	$u = x - 65$	$v = y - 55$	uv	u^2
	60	50	–5	–5	25	25
	70	59	5	4	20	25
	64	57	–1	2	–2	1
	68	50	3	–5	–15	9
	62	53	–3	–2	6	9
	65	58	0	3	0	0
	72	60	7	5	35	49
Total	**461**	**387**	**6**	**2**	**69**	**118**

b can be obtained by short-cut method as follows.

$$b = b_{yx} = b_{vu} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2}$$

$$= \frac{7(69) - (6)(2)}{7(118) - (6)^2}$$

$$= \frac{483 - 12}{826 - 36}$$

$$= \frac{471}{790}$$

$$= 0.5962$$

$$\therefore \quad b \simeq 0.60$$

Now, $a = \bar{y} - b\bar{x}$

$$= 55.29 - 0.60(65.86)$$

$$= 55.29 - 39.516$$

$$= 15.774$$

$$\therefore \quad a = 15.77$$

So, the regression line of Y on X is

$$\hat{y} = a + bx$$

$$\therefore \quad \hat{y} = 15.77 + 0.60x$$

Putting $X = 75$,

$$\hat{y} = 15.77 + 0.60(75)$$

$$= 15.77 + 45$$

$$= 60.77$$

$$\therefore \quad \hat{y} = 60.77$$

So, if a person has monthly income of ₹ 75 thousand, his approximate monthly expenditure is ₹ 60.77 thousand.

Illustration 5 : For the data given in illustration 1, obtain the regression line of maintenance cost (*Y*) on the life of cars (*X*) by using short-cut method.

Life of car (years) x	**2**	**4**	**6**	**8**
Annual maintenance cost (thousand ₹) y	**10**	**20**	**25**	**30**

All the values of X here are divisible by 2 and that of Y are divisible by 5. Moreover $\bar{x} = 5$ and $\bar{y} = 21.25$. So, we shall take $A = 4, B = 20, c_x = 2, c_y = 5$.

Now, let us define u and v as follows :

$$u = \frac{x-A}{c_x} = \frac{x-4}{2} \quad \text{and} \quad v = \frac{y-B}{c_y} = \frac{y-20}{5}$$

	x	y	$u = \frac{x-4}{2}$	$v = \frac{y-20}{5}$	uv	u^2
	2	10	–1	–2	2	1
	4	20	0	0	0	0
	6	25	1	1	1	1
	8	30	2	2	4	4
Total	**20**	**85**	**2**	**1**	**7**	**6**

$$b = b_{vu} \cdot \frac{c_y}{c_x} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2} \times \frac{c_y}{c_x}$$

$$= \frac{4(7)-2(1)}{4(6)-(2)^2} \times \frac{5}{2}$$

$$= \frac{28-2}{24-4} \times \frac{5}{2}$$

$$= \frac{26}{20} \times \frac{5}{2}$$

$$b = 3.25$$

Now, $a = \bar{y} - b\bar{x} = 21.25 - 3.25(5) = 21.25 - 16.25 = 5$

$\therefore$ The regression line of Y on X is

$\hat{y} = a + bx$

$\therefore \quad \hat{y} = 5 + 3.25x$

Note : We can see that, $b_{vu} = \frac{26}{20} = 1.3$ but when it is multiplied by $\frac{c_y}{c_x} = \frac{5}{2}$ then we get $b = 1.3 \times \frac{5}{2} =$ 3.25 (as obtained in illustration 1). So, we can understand that when the scale of variable X and/or Y are changed, it is necessary to multiply b_{vu} by $\frac{c_y}{c_x}$ to obtain b.

Illustration 6 : A sample of seven students is taken from the students coming from abroad in the current year to study in university of the Gujarat State. The information regarding their I.Q. and the marks obtained in an examination of 75 marks is given below.

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**
I.Q. x	85	95	100	90	110	125	70
Marks y	46	50	50	45	60	70	40

Obtain the regression line of Y on X and estimate the marks of a student whose I.Q. is 120. Also find the error in estimation when I.Q. is 100.

Here, $n = 7,\ \bar{x} = \frac{\Sigma x}{n} = \frac{675}{7} = 96.43,\ \bar{y} = \frac{\Sigma y}{n} = \frac{361}{7} = 51.57$

Since the values of X and Y are large, means are fractional and all the values of X are divisible by 5, we shall use short-cut method.

By taking $A = 95, B = 50, c_x = 5, c_y = 1$, we define u and v as follows.

$u = \frac{x-A}{c_x} = \frac{x-95}{5}$ and $v = \frac{y-B}{c_y} = \frac{y-50}{1} = y - 50$

	I.Q. x	**Marks** y	$u = \frac{x-95}{5}$	$v = y - 50$	uv	u^2
	85	46	–2	–4	8	4
	95	50	0	0	0	0
	100	50	1	0	0	1
	90	45	–1	–5	5	1
	110	60	3	10	30	9
	125	70	6	20	120	36
	70	40	–5	–10	50	25
Total	**675**	**361**	**2**	**11**	**213**	**76**

$$b = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2} \times \frac{c_y}{c_x}$$

$$= \frac{7(213) - (2)(11)}{7(76) - (2)^2} \times \frac{1}{5}$$

$$= \frac{1491 - 22}{532 - 4} \times \frac{1}{5}$$

$$= \frac{1469}{528} \times \frac{1}{5}$$

$$= \frac{1469}{2640}$$

$$= 0.5564$$

$\therefore \; b \simeq 0.56$

Now, $a = \overline{y} - b\overline{x}$

$$= 51.57 - 0.56(96.43)$$

$$= 51.57 - 54.0008$$

$$= -2.4308$$

$\therefore \; a \simeq -2.43$

So, the regression line of Y on X is

$$\hat{y} = a + bx$$

$$\therefore \quad \hat{y} = -2.43 + 0.56x$$

Putting $X = 120$,

$$\hat{y} = -2.43 + 0.56(120)$$

$$= -2.43 + 67.2$$

$\therefore \quad \hat{y} = 64.77$ marks.

So, when the I.Q. of a student is 120 then his marks are approximately 65.

Now, to obtain the error when I.Q. (X) = 100, first we have to find the estimate of Y i.e. $\hat{y}$.

$$\hat{y} = -2.43 + 0.56x$$

Taking X = 100,

$$\hat{y} = -2.43 + 0.56(100)$$

$$= -2.43 + 56$$

$\therefore \quad \hat{y} = 53.57$ marks

But the observed value of Y for X = 100 is 50. (See the given data)

$\therefore$ Error $e = y - \hat{y}$

$= 50 - 53.57$

$\therefore$ $e = -3.57$ marks

Note : It is necessary to keep in mind that the error can be obtained only for those values of independent variable (X), for which the observed value of dependent variable (Y) are known.

In this example, we can not obtain the error in estimating Y for $X = 120$ because the observed value of Y when $X = 120$ is not known.

Illustration 7 : From the data and calculation of illustration 12 of the chapter of linear correlation, obtain the regression line of profit on the sales. Estimate the profit when sales is ₹ 3 crore.

From the illustration, we know that

$$u = \frac{x-A}{c_x} = \frac{x-2}{0.1} \text{ and } v = \frac{y-B}{c_y} = \frac{y-5600}{100}$$

$\therefore c_x = 0.1$ and $c_y = 100$

Note that c_x is the divisor of $(x-A)$. So, though we have multiplied $(x-A)$ by 10 for simplicity of calculation, c_x is $\frac{1}{10}$ = 0.1.

($\because$ To multiply by 10 is same as to divide by $\frac{1}{10}$ = 0.1)

$$\text{Now } b = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2} \times \frac{c_y}{c_x}$$

$$= \frac{9(121)-(0)(1)}{9(60)-(0)^2} \times \frac{100}{0.1}$$

$$= \frac{1089}{540} \times \frac{100}{0.1}$$

$$= \frac{108900}{54}$$

$= 2016.6667$

$\therefore b \simeq 2016.67$

Now, $a = \bar{y} - b\bar{x}$

$= 5611.11 - 2016.67(2)$

$= 5611.11 - 4033.34$

$\therefore a = 1577.77$

So, the regression line of profit (Y) on the sales (X) is

$\hat{y} = a + bx$

$\therefore \hat{y} = 1577.77 + 2016.67x$

Putting $X = 3$,

$$\hat{y} = 1577.77 + 2016.67(3)$$

$$= 1577.77 + 6050.01$$

$$\therefore \quad \hat{y} = 7627.78$$

So, when sales is ₹ 3 crore then the estimated profit is 7627.78 (thousand ₹).

Activity

Collect the information of monthly income and monthly expenditure of your family from June to December of a year in which you are studying in standard 12. Obtain the regression line of monthly expenditure on the monthly income. Estimate the monthly expenditure of January of the successive year. Check the actual expenditure at the end of January and find the error in your estimation.

3.5 Regression coefficient from covariance and correlation coefficient

When the summary measures like mean, standard deviation (or variance), covariance, correlation coefficient are known for bivariate data of two variables X and Y, regression coefficient and the line of regression can be obtained as follows.

(1) When the measures like $\bar{x}, \bar{y}, s_x^2$ (or s_x), s_y^2 (or s_y) and Cov(x, y) are known,

$$b = \frac{\text{Covariance}(x, y)}{\text{Variance of } x} = \frac{Cov(x, y)}{s_x^2}$$

and $a = \bar{y} - b\bar{x}$

where, $$Cov(x, y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n} = \frac{\Sigma xy - n\,\bar{x}\,\bar{y}}{n}$$

$$s_x^2 = \frac{\Sigma(x-\bar{x})^2}{n} = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2 = \frac{\Sigma x^2}{n} - \bar{x}^2$$

$$s_y^2 = \frac{\Sigma(y-\bar{y})^2}{n} = \frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2 = \frac{\Sigma y^2}{n} - \bar{y}^2$$

(2) When the measures like $\bar{x}, \bar{y}, r, s_x$ (or s_x^2), and s_y (or s_y^2) are known,

$$b = r \cdot \frac{\text{S.D. of } y}{\text{S.D. of } x} = r \cdot \frac{s_y}{s_x}$$

and $a = \bar{y} - b\bar{x}$

The regression line of Y on X i.e. $\hat{y} = a + bx$ can be obtained by putting the values of a and b.

Now, we consider some examples in which some summary measures are known and the regression line is to be obtained.

Illustration 8 : The following measures are obtained to study the relation between rainfall in cm (*X*) and yield of Bajri in Quintal per Hectare (*Y*) in ten different regions during monsoon.

$n = 10, \bar{x} = 40, \bar{y} = 175, s_x = 12, Cov(x, y) = 360$

Obtain the regression line of yield *Y* on rainfall *X*.

Here, $Cov(x, y) = 360$ and $s_x = 12 \therefore s_x^2 = 144$

$$b = \frac{Cov(x, y)}{s_x^2}$$

$$= \frac{360}{144}$$

$$\therefore \quad b = 2.5$$

and $a = \bar{y} - b\bar{x}$

$$= 175 - 2.5(40)$$

$$= 175 - 100$$

$$\therefore \; a = 75$$

So, the regression line of *Y* on *X* is

$$\hat{y} = a + bx$$

$$\therefore \quad \hat{y} = 75 + 2.5x$$

Illustration 9 : To study the relation between two variables, yearly income (*X*) of a family and their yearly investment (*Y*) in mutual funds, the following information is shown for a sample of 100 families of a city.

***X* = Annual income of a family (lakh ₹)**

***Y* = Annual investment in mutual fund of a family (thousand ₹)**

$\bar{x} = 5.5, \; \bar{y} = 40.5, \; s_x = 1.2, \; s_y = 12.8, \; r = 0.65$

Obtain the regression line of annual investment in mutual fund of a family on their annual income. Estimate the annual investment in mutual fund of a family whose annual income is ₹ 4.5 lakh.

Here, $n = 100, \bar{x} = 5.5, \bar{y} = 40.5$

$s_x = 1.2, s_y = 12.8$ and $r = 0.65$

Now, $b = r \cdot \frac{s_y}{s_x}$

$$= 0.65 \times \frac{12.8}{1.2}$$

$$= 6.9333$$

$$\therefore \; b \simeq 6.93$$

And $a = \overline{y} - b\overline{x}$

$= 40.5 - 6.93\ (5.5)$

$= 40.5 - 38.115$

$= 2.385$

$\therefore \quad a \simeq 2.39$

So, the regression line of Y on X is

$\hat{y} = a + bx$

$\therefore \quad \hat{y} = 2.39 + 6.93x$

Putting $X = 4.5$,

$\hat{y} = 2.39 + 6.93(4.5)$

$= 2.39 + 31.185$

$= 33.575$

$\therefore \quad \hat{y} \simeq 33.58$

So, when annual income of a family is ₹ 4.5 lakh then estimated investment in mutual fund is ₹ 33.58 thousand.

Illustration 10 : The information of price (in ₹) of a ballpen and the supply of ballpen (in units) at the end of each month of a year for a company making ball pen is given below. Estimate the supply of ballpen when its price is ₹ 40.

Detail	**Price (x)**	**Supply (y)**
Average	30	500
Variance	25	10,000
$r = 0.8$		

Here, $\overline{x} = 30, \overline{y} = 500, s_x^2 = 25, s_y^2 = 10000$ and $r = 0.8$

Since $s_x^2 = 25$, $s_x = 5$

Since $s_y^2 = 10000$, $s_y = 100$

Since the supply Y is to be estimated for the price $X = 40$, we shall obtain the regression line of Y on X.

$$b = r \cdot \frac{s_y}{s_x}$$

$$= 0.8 \times \frac{100}{5}$$

$\therefore \quad b = 16$

$a = \overline{y} - b\overline{x}$

$= 500 - 16\ (30)$

$= 500 - 480$

$\therefore \quad a = 20$

So, the regression line of Y on X is

$\hat{y} = a + bx$

$\therefore \quad \hat{y} = 20 + 16x$

Putting $X = 40$,

$\hat{y} = 20 + 16(40)$

$= 20 + 640$

$\therefore \quad \hat{y} = 660$ units

So, when the price is ₹ 40, the estimate of supply is 660 units.

Illustration 11 : A person in a state of South India produces spoons from eatable materials. It can be eaten after using it. He launched such spoons for the purpose of selling in a state on an experimental level. The following results are obtained for the average price (in ₹) and its demand (in hundred units) for the last six months.

$\boldsymbol{n = 6, \Sigma x = 45, \Sigma y = 122, \Sigma x^2 = 439, \Sigma xy = 605}$

Obtain the regression line of the demand (Y) of spoons on the price (X) and estimate the demand of spoons when the price of a spoon is ₹ 10.

Here, $\overline{x} = \frac{\Sigma x}{n} = \frac{45}{6} = 7.5$, $\overline{y} = \frac{\Sigma y}{n} = \frac{122}{6} = 20.33$

$$b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{6(605) - (45)(122)}{6(439) - (45)^2}$$

$$= \frac{3630 - 5490}{2634 - 2025}$$

$$= \frac{-1860}{609}$$

$= -3.0542$

$\therefore \quad b \simeq -3.05$

$a = \overline{y} - b\overline{x}$

$= 20.33 - (-3.05)\ (7.5)$

$= 20.33 + 22.875$

$= 43.205$

$\therefore \quad a \simeq 43.21$

So, the regression line of Y on X is

$$\hat{y} = a + bx$$

$$\therefore \quad \hat{y} = 43.21 - 3.05x$$

Putting $X = 10$,

$$\hat{y} = 43.21 - 3.05(10)$$

$$= 43.21 - 30.5$$

$$\therefore \quad \hat{y} = 12.71$$

So, when the price is ₹ 10, estimated demand is 12.71 (hundred units).

Illustration 12 : The electricity is generated by windmill manufactured by a company. The following information is obtained by recording five observations regarding the velocity of wind (km per hour) and generation of electricity (in Watts) by a unit of the company.

Velocity of Wind = X km per hour

Electricity Generation = Y Watts

$$\bar{x} = 20,\ \bar{y} = 186,\ \Sigma xy = 23200,\ s_x^2 = 50$$

Obtain the regression line of electricity generation (Y) on velocity of wind (X). Estimate the electricity generation if the velocity of wind is 25 km per hour.

Here, $n = 5$, $\Sigma xy = 23200$, $\bar{x} = 20$, $\bar{y} = 186$ and $s_x^2 = 50$

$$\text{Now,} \quad b = \frac{Cov(x, y)}{s_x^2}$$

$$= \frac{\Sigma xy - n\,\bar{x}\,\bar{y}}{n \cdot s_x^2}$$

$$= \frac{23200 - 5(20)(186)}{5(50)}$$

$$= \frac{23200 - 18600}{250}$$

$$= \frac{4600}{250}$$

$$\therefore \quad b = 18.4$$

$$a = \bar{y} - b\bar{x}$$

$$= 186 - 18.4\ (20)$$

$$= 186 - 368$$

$$\therefore \quad a = -182$$

So, the regression line of Y on X is

$$\hat{y} = a + bx$$

$$\therefore \quad \hat{y} = -182 + 18.4x$$

Putting $X = 25$,

$$\hat{y} = -182 + 18.4(25)$$

$$= -182 + 460$$

$$\therefore \quad \hat{y} = 278$$

So, when the velocity of wind is 25 km per hour, approximately 278 watts electricity is generated.

Exercise 3.2

1. The following information is obtained from a study to know the effect of use of fertilizer on the yield of cotton.

Consumption of fertilizer (10 kg) x	28	35	25	24	20	25	20
Yield of cotton per hectare (Quintals) y	128	140	115	120	105	122	100

Obtain the regression line of Y on X and estimate the yield of cotton per hectare if 300 kg fertilizer is used.

2. To know the relationship between the heights of father and sons, obtain the regression line of height of son on the height of father from the following information of eight pairs of fathers and adult sons.

Height of father (cm) x	167	169	171	168	173	166	167	165
Height of son (cm) y	158	170	169	172	170	168	164	167

Estimate the height of a son whose father's height is 170 cm.

3. From the following information of altitude and the amount of effective Oxygen in air at the place, obtain the regression line of amount of effective Oxygen (Y) on the altitude (X). (305 meter $\simeq$ 1000 feet)

Altitude (305 meter) x	0	1	2	3	4	5	6
Effective Oxygen (%) y	20.9	20.1	19.4	17.9	17.9	17.3	16.6

If the altitude of a place is 7 units (1 unit = 305 meter), estimate the percentage of effective Oxygen in air at that place.

4. The following information is obtained to study the relation between the carpet area in a house and its monthly rent in a city.

Carpet area (square meter) x	55	60	75	80	100	120	140
Monthly rent (₹) y	18,000	19,000	20,000	20,000	25,000	30,000	50,000

Obtain the regression line of Y on X. Estimate the monthly rent of a house having carpet area of 110 square meter.

5. The following sample data is obtained to study the relation between the number of customers visiting a mall per day and the sales (ten thousand ₹).

No. of customers x	50	70	100	70	150	120
Sales (ten thousand ₹) y	2.0	2.0	2.5	1.4	4.0	2.5

Obtain the regression line of Y on X. Estimate the sales of a mall if 80 customers have visited the mall on a particular day.

6. The following information is given for ten firms running business of clothes in a city regarding their average annual profit (in lakh ₹) and average annual administrative cost (in lakh ₹).

Particulars	**Profit (in lakh ₹)** x	**Administrative Cost (in lakh ₹)** y
Mean	60	25
Standard Deviation	6	3
Covariance = 10.4		

Obtain the regression line of Y on X.

7. The following information is obtained to study the relationship between average rainfall (in cm) and the yield of maize (in quintal per hectare) in different talukaa of Gujarat.

Particulars	**Rainfall (cm)** x	**Yield of Maize (Quintal per Hectare)** y
Mean	82	180
Variance	64	225
Correlation coefficient = 0.82		

Estimate the yield of maize when the rainfall is 60 cm.

8. The following results are obtained to study the relation between the price of battery (cell) of wrist watch in rupees (X) and its supply in hundred units (Y).

$n = 10, \Sigma x = 130, \Sigma y = 220, \Sigma x^2 = 2288, \Sigma xy = 3467$

Obtain the regression line of Y on X and estimate the supply when price is ₹ 16.

9. The information regarding maximum temperature (X) and sale of ice-cream (Y) of six different days in summer for a city is given below.

Maximum Temperature = X (in Celsius)

Sale of Ice-cream = Y (in lakh ₹)

$\bar{x} = 40, \bar{y} = 1.2, \Sigma xy = 306, s_x^2 = 20$

Obtain the regression line of sale of ice-cream on maximum temperature. Estimate the sale of ice-cream if the maximum temperature on a day is 42 Celsius.

*

3.6 Coefficient of Determination

We know that regression is a functional relation between two correlated variables and it is useful to estimate the value of dependent variable for some given value of independent variable. The coefficient of determination is a measure to find the reliability of such an estimate.

Suppose the regression line of Y on X is $\hat{y} = a + bx$, then the square of the correlation coefficient between observed values of dependent variable y obtained from the observations and its estimated values $\hat{y}$ which are obtained from the regression line is called the **coefficient of determination.** It is denoted by R^2.

$$\therefore R^2 = \left[r\left(y, \hat{y}\right)\right]^2$$

It can be easily checked that R^2 is same as $r^2(x, y)$ or r^2.

$$R^2 = \left[r\left(y, \hat{y}\right)\right]^2$$

$$= \left[r\left(y, a + bx\right)\right]^2$$

$$= \left[r\left(y, x\right)\right]^2$$

$$= \left[r\left(x, y\right)\right]^2$$

$$\therefore R^2 = r^2$$

{ $\because$ r is independent of change of origin and scale, so from variable $\hat{y}(= a + bx)$, substracting a and then dividing by b, the value of r will not change. }

Since $R^2 = r^2$, we can say that the reliability of an estimate of dependent variable Y largely depends on the correlation coefficient r between two variables X and Y.

If $r = \pm 1$ then $R^2 = r^2 = 1$ and there is a perfect linear correlation between X and Y. So, we can say that the estimates of Y obtained from the regression line are 100 % reliable. But if $r = 0$ then $R^2 = r^2 = 0$ and there is no linear correlation between X and Y. So, we can say that the estimates of Y obtained from the regression line are not reliable.

It is clear from the above discussion that high value of R^2 shows that a good linear correlation exists between two variables X and Y. So, we can check whether the linearity assumption of regression is valid or not from the measure of coefficient of determination (R^2). If the value of R^2 is nearer to 1, the assumption of linearity of regression is valid. But if the value of R^2 is nearer to 0, the assumption of linearity of regression between X and Y is not valid.

How much variation in the dependent variable Y can be explained by the regression line, can be obtained from the coefficient of determination. e.g., If $r = 0.9$ for some data, then coefficient of determination = $(0.9)^2 = 0.81$ and therefore $r^2 \times 100\ \% = 81\ \%$. So, it can be said that out of total variation in variable Y, the explanation of 81 % variation is obtained from the regression line. So, we can say that the regression model used for the given data is suitable.

Illustration 13 : The following table shows the experience of technicians (in years) employed at various companies and their monthly salary (in thousand ₹).

Experience (years) x	12	8	16	20	5	14	10
Monthly Salary (thousand ₹) y	22	15	25	30	12	24	20

Calculate the coefficient of determination and check the validity of the linearity assumption of regression between the years of experience and the monthly salary.

Here, $n = 7$, $\bar{x} = \frac{\Sigma x}{n} = \frac{85}{7} = 12.14$, $\bar{y} = \frac{\Sigma y}{n} = \frac{148}{7} = 21.14$

	Experience (year) x	Monthly salary (thousand ₹) y	xy	x^2	y^2
	12	22	264	144	484
	8	15	120	64	225
	16	25	400	256	625
	20	30	600	400	900
	5	12	60	25	144
	14	24	336	196	576
	10	20	200	100	400
Total	**85**	**148**	**1980**	**1185**	**3354**

Now, $R^2 = r^2 = \left[\frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{n\Sigma y^2 - (\Sigma y)^2}}\right]^2$

$$= \left[\frac{7(1980) - (85)(148)}{\sqrt{7(1185) - (85)^2} \cdot \sqrt{7(3354) - (148)^2}}\right]^2$$

$$= \frac{[13860 - 12580]^2}{[8295 - 7225] \cdot [23478 - 21904]}$$

$$= \frac{(1280)^2}{(1070) \cdot (1574)}$$

$$= \frac{1638400}{1684180}$$

$$= 0.9728$$

$\therefore R^2 \simeq 0.97$

The value of R^2 is very near to 1. So, we can say that the linearity assumption of regression between the years of experience and the monthly salary is valid.

Note : For the above example, we can also compute R^2 by taking $u = x - A$ and $v = y - B$. Here, A and B are suitable constants.

Illustration 14 : In order to study the relationship between the density of population and the number of persons suffering from skin diseases, the following information is obtained for six cities regarding their density of population (per sq. km) and persons suffering from skin diseases (per thousand).

Density (per sq. km) x	**12,000**	**14,500**	**19,000**	**17,500**	**13,500**	**16,000**
Number of patients (per thousand) y	**80**	**60**	**90**	**80**	**40**	**30**

Obtain the regression line of Y **on** X**. Estimate the number of patients suffering from skin diseases if density of population of a city is 15000 (per sq.km). Examine the reliability of this regression model.**

Here, $n = 6$, $\bar{x} = \frac{\Sigma x}{n} = \frac{92500}{6} = 15416.67$; $\bar{y} = \frac{\Sigma y}{n} = \frac{380}{6} = 63.33$

We can see that the values of variable X are multiple of 500 and that of variable Y are of 10. So, by taking $A = 15000$, $B = 60$, $c_x = 500$, $c_y = 10$, we shall use short-cut method. Let us define u and v as follows.

$$u = \frac{x-A}{c_x} = \frac{x-15000}{500} \quad \text{and} \quad v = \frac{y-B}{c_y} = \frac{y-60}{10}$$

	Density (per sq. km) x	**Number of patients (per thousand)** y	$u = \frac{x-15000}{500}$	$v = \frac{y-60}{10}$	uv	u^2	v^2
	12000	80	–6	2	–12	36	4
	14500	60	–1	0	0	1	0
	19000	90	8	3	24	64	9
	17500	80	5	2	10	25	4
	13500	40	–3	–2	6	9	4
	16000	30	2	–3	–6	4	9
Total	**92500**	**380**	**5**	**2**	**22**	**139**	**30**

$$b = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2} \times \frac{c_y}{c_x}$$

$$= \frac{6(22) - (5)(2)}{6(139) - (5)^2} \times \frac{10}{500}$$

$$= \frac{132 - 10}{834 - 25} \times \frac{1}{50}$$

$$= \frac{122}{809} \times \frac{1}{50}$$

$$= \frac{122}{40450}$$

$\therefore b \simeq 0.003$

$$a = \overline{y} - b\overline{x}$$
$$= 63.33 - 0.003\,(15416.67)$$
$$= 63.33 - 46.25$$

$\therefore\ a = 17.08$

$\therefore$ The regression line of Y on X is

$$\hat{y} = a + bx$$

$\therefore\ \hat{y} = 17.08 + 0.003\,x$

Putting $X = 15000$,

$$\hat{y} = 17.08 + 0.003(15000)$$
$$= 17.08 + 45$$

$\therefore\ \hat{y} = 62.08$

So, when the density of a city is 15,000 then approximately $62.08 \simeq 62$ patients are suffering from skin diseases.

Now, reliability of a regression model can be examined by coefficient of determination R^2. So, we obtain it.

$$R^2 = r^2 = \left[\frac{n\Sigma uv - (\Sigma u)\,(\Sigma v)}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\cdot\sqrt{n\Sigma v^2 - (\Sigma v)^2}}\right]^2$$

$$= \frac{[6(22) - (5)\,(2)]^2}{\left[6(139) - (5)^2\right]\left[6(30) - (2)^2\right]}$$

$$= \frac{(122)^2}{(809)\,(176)}$$

$$= \frac{14884}{142384}$$

$$= 0.1045$$

$\therefore\ R^2 \simeq 0.10$

As the value of R^2 is very near to 0, it can not be said that the regression model is reliable.

3.7 Properties of Regression Coefficient

(1) Correlation coefficient r and regression coefficient b are either both positive or both negative. ($\because$ We know that standard deviations s_x and s_y are always non-negative and $-1 \le r \le 1$. So, from $b = r \cdot \frac{s_y}{s_x}$ it can be understood that the sign of b will be same as that of r.)

(2) Regression coefficient is independent of change of origin but not independent of change of scale. (This property is discussed in detail in the explanation of the short-cut method of calculation of regression coefficient.)

Note : The regression line of Y on X always passes through the point $(\overline{x}, \overline{y})$.

Illustration 15 : Six pairs of father-son are selected in a sample of an experiment to know the relation between the heights of fathers in cm (*X*) and the heights of their adult sons in cm (*Y*).

The following results are obtained from it.

$$\Sigma x = 1020,\ \Sigma y = 990,\ \Sigma(x-170)^2 = 60,\ \Sigma(y-165)^2 = 105$$

$$\Sigma(x-170)(y-165) = 45$$

Obtain the regression line of the heights of sons (*Y*) on the heights of fathers (*X*). Also verify the reliability of the regression model.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1020}{6} = 170$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{990}{6} = 165$$

$$\therefore \Sigma(x-170)^2 = \Sigma(x-\bar{x})^2 = 60$$

$$\Sigma(y-165)^2 = \Sigma(y-\bar{y})^2 = 105$$

$$\Sigma(x-170)(y-165) = \Sigma(x-\bar{x})(y-\bar{y}) = 45$$

$$\therefore b = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$= \frac{45}{60}$$

$$\therefore b = 0.75$$

$$a = \bar{y} - b\bar{x}$$

$$= 165 - 0.75(170)$$

$$= 165 - 127.5$$

$$\therefore a = 37.5$$

So, the regression line of Y on X is

$$\hat{y} = a + bx$$

$$\therefore \hat{y} = 37.5 + 0.75x$$

Now, to verify reliability of the regression model, let us obtain the coefficient of determination R^2.

$$R^2 = \left[\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\cdot\sqrt{\Sigma(y-\bar{y})^2}}\right]^2$$

$$= \frac{(45)^2}{(60)(105)}$$

$$= \frac{2025}{6300}$$

$$= 0.3214$$

$$\therefore R^2 \simeq 0.32$$

Since the value of R^2 is nearer to 0, it can not be said that the regression model is reliable.

Illustration 16 : (i) If the regression line of Y on X is $\hat{y} = 12 - 1.5x$ and the mean of X is 6, find the mean of Y. (ii) If the regression line of Y on X is $\hat{y} = 11.5 + 0.65x$ and $\bar{y} = 18$, find $\bar{x}$.

(i) We know that the regression line always passes through a point $(\bar{x}, \bar{y})$. So, the $\hat{y}$ obtained by putting $\bar{x}$ in place of x in the regression line is $\bar{y}$ or the x obtained by putting $\bar{y}$ in place of $\hat{y}$ is $\bar{x}$.

Putting $\bar{x} = 6$ in place of x in $\hat{y} = 12 - 1.5x$,

$\hat{y} = 12 - 1.5(6)$

$\therefore \hat{y} = 12 - 9$

$\therefore \hat{y} = 3$, so $\bar{y} = 3$

Therefore, the mean of Y is 3.

(ii) As per the above discussion, the value of x obtained by putting $\bar{y} = 18$ in place of $\hat{y}$ in $\hat{y} = 11.5 + 0.65x$, we get $\bar{x}$.

By putting $\hat{y} = \bar{y} = 18$ in $\hat{y} = 11.5 + 0.65x$,

$18 = 11.5 + 0.65x$

$\therefore 6.5 = 0.65x$

$\therefore x = \frac{6.5}{0.65}$

$\therefore x = 10$ So $\bar{x} = 10$

Therefore, the mean of X is 10.

Illustration 17 : (i) If $\bar{x} = 5, \bar{y} = 11$ and $b = 1.2$, obtain the regression line of Y on X. (ii) If $\bar{x} = 60, \bar{y} = 75$ and $s_x^2 : Cov(x, y) = 5:3$, obtain the regression line of Y on X and estimate y for $X = 65$ from it.

(i) Here, $b = 1.2$, $\bar{x} = 5$ and $\bar{y} = 11$.

Now, $a = \bar{y} - b\bar{x}$

$\therefore a = 11 - 1.2(5)$

$= 11 - 6$

$\therefore a = 5$

We get the regression line of Y on X as follows.

$\hat{y} = a + bx$

$\therefore \hat{y} = 5 + 1.2x$

(ii) Here, $\bar{x} = 60$, $\bar{y} = 75$ and $s_x^2 : Cov(x, y) = 5:3$

$s_x^2 : Cov(x, y) = 5:3$

$\therefore \frac{s_x^2}{Cov(x, y)} = \frac{5}{3}$ hence, $\frac{Cov(x, y)}{s_x^2} = \frac{3}{5}$

Now, $b = \frac{Cov(x, y)}{s_x^2} = \frac{3}{5} = 0.6$

and $a = \bar{y} - b\bar{x}$

$= 75 - 0.6(60)$

$= 75 - 36$

$\therefore a = 39$

We get the regression line of Y on X as follows.

$\hat{y} = a + bx$

$\therefore \hat{y} = 39 + 0.6x$

Putting $X = 65$,

$\hat{y} = 39 + 0.6x$

$= 39 + 0.6(65)$

$= 39 + 39$

$\therefore \hat{y} = 78$

So, for $X = 65$, the estimated value of Y is 78.

Illustration 18 : The fitted regression line of Y on X is $\hat{y} = 50 + 3.5x$. If an observation $(16, 108)$ is used in fitting of the line, find the error in estimating Y for $X = 16$. (ii) If one observation $(10, 30)$ is used in the fitting of the line $\hat{y} = 22 + 0.8x$, find the error in estimating Y for $X = 10$. What can you deduce from the value of the error ?

(i) Putting $x = 16$ in $\hat{y} = 50 + 3.5x$,

$\hat{y} = 50 + 3.5(16)$

$\therefore = 50 + 56$

$\therefore \hat{y} = 106$

And for $X = 16$, corresponding $Y = 108$ is observed.

$\therefore$ Error $e = y - \hat{y}$

$= 108 - 106$

$\therefore e = 2$

So, the error in estimating Y for $X = 16$ is 2.

(ii) Putting $X=10$ in $\hat{y}=22+0.8x$,

$$\hat{y}= 22+0.8(10)$$

$$= 22+8$$

$$\therefore\ \hat{y}=30$$

And for $X=10$, corresponding $Y=30$ is observed.

$$\therefore\ \text{Error}\ e = y-\hat{y}$$

$$=30-30$$

$$\therefore\ e=0$$

So, the error in estimating Y for $X=10$ is 0.

Since value of the error is 0, we can say that the point $(10, 30)$ lies on the fitted line $\hat{y}=22+0.8x$.

Note : For a regression line obtained by the method of least squares, the error is positive for the points above the line, negative for the points below the line and it is zero for the points which are on the line.

Illustration 19 : (i) If the regression line of Y on X is $\hat{y}=25+3x$ and $Cov(x, y)=48$, find the standard deviation of X. Also find coefficient of determination if the standard deviation of Y is 15. (ii) For the regression line given in the above question, how many units should be increased in the value of X to increase approximately 15 units in Y ?

(i) By comparing the regression line of y on x, $\hat{y}=25+3x$ with its general form $\hat{y}=a+bx$, we get regression coefficient $b=3$. Since $Cov(x, y)=48$ is given,

$$b=\frac{Cov(x, y)}{s_x^2}$$

$$\therefore\ 3 = \frac{48}{s_x^2}$$

$$\therefore\ s_x^2 = 16$$

$$\therefore\ s_x = 4$$

So, the standard deviation of X is 4.

Now, the standard deviation of Y, $s_y=15$ is given.

So, coefficient of determination $R^2 = \left[\frac{Cov(x,y)}{s_x \cdot s_y}\right]^2$

$$\therefore\ R^2 = \left[\frac{48}{4\times 15}\right]^2 = (0.8)^2 = 0.64$$

Second Method :

$b = r \cdot \dfrac{s_y}{s_x}$

$\therefore 3 = r \cdot \dfrac{15}{4}$

$\therefore r = \dfrac{3 \times 4}{15}$

$\therefore r = 0.8$

$\therefore R^2 = r^2 = (0.8)^2 = 0.64$

(ii) $\hat{y} = 25 + 3x$ and regression coefficient $b = 3$. It indicates that if the value of X is increased by one unit then estimated value of Y is increased by 3 units. So, if the value of Y is to be increased approximately by 15 units then the value of X should be increased by $\frac{15}{3} = 5$ units.

Illustration 20 : (i) If the regression line is $\hat{y} = \frac{x}{2} + 5$ and $s_y : s_x = 5:8$, find the coefficient of determination. (ii) If the regression line of Y on X is $4x + 5y - 65 = 0$, find the value of regression coefficient *b*.

(i) By comparing the regression line $\hat{y} = \frac{x}{2} + 5 = \frac{1}{2} \cdot x + 5$ with its general form $\hat{y} = a + bx$, we get $b = \frac{1}{2}$.

Now, $s_y : s_x = 5:8$

$\therefore \dfrac{s_y}{s_x} = \dfrac{5}{8}$

and $b = r \cdot \dfrac{s_y}{s_x}$

$\therefore \dfrac{1}{2} = r \cdot \dfrac{5}{8}$

$\therefore r = \dfrac{1}{2} \times \dfrac{8}{5}$

$\therefore r = 0.8$

$\therefore$ The coefficient of determination $R^2 = r^2 = (0.8)^2 = 0.64$.

(ii) The regression line of Y on X, $4x + 5y - 65 = 0$ is given.

Now, we convert it into its general form.

$4x + 5y - 65 = 0$

$\therefore 5y = 65 - 4x$

$\therefore y = \dfrac{65 - 4x}{5}$

$\therefore y = \dfrac{65}{5} - \dfrac{4x}{5}$

$\therefore y = 13 - 0.8x$

By comparing it with $\hat{y} = a + bx$, we get $b = -0.8$.

Illustration 21 :

(i) If $b_{yx} = 0.85$, $u = x - 15$ and $v = y - 20$, find the value of b_{vu}.

(ii) If $u = \frac{x-5}{3}$, $v = \frac{y-8}{5}$ and $b_{yx} = 0.9$, find the value of b_{vu}.

(iii) If $u = 10(x - 4.5)$, $v = \frac{y-50}{10}$ and $b_{yx} = 0.25$, find the value of b_{vu}.

(iv) If $u = 5(x - 40)$, $v = 2(y - 18)$ and $b_{yx} = 1.6$, find the value of b_{vu}.

For the solution of all the questions given above, we shall use the following property of the regression coefficient.

- If $u = x - A$ and $v = y - B$ then $b_{yx} = b_{vu}$

- If $u = \frac{x-A}{c_x}$ and $v = \frac{y-B}{c_y}$ then $b_{yx} = b_{vu} \cdot \frac{c_y}{c_x}$

(i) Since $u = x - 15 = x - A$ and $v = y - 20 = y - B$

$\therefore\ b_{vu} = b_{yx} = 0.85$

(ii) Since $u = \frac{x-5}{3} = \frac{x-A}{c_x}$ and $v = \frac{y-8}{5} = \frac{y-B}{c_y}$

$$b_{yx} = b_{vu} \cdot \frac{c_y}{c_x} \quad \therefore\ b_{vu} = b_{yx} \cdot \frac{c_x}{c_y} = 0.9 \times \frac{3}{5} = 0.54$$

(iii) Since $u = 10(x - 4.5) = \frac{x-4.5}{\frac{1}{10}} = \frac{x-A}{c_x}$ and $v = \frac{y-50}{10} = \frac{y-B}{c_y}$

$$b_{yx} = b_{vu} \cdot \frac{c_y}{c_x} \quad \therefore\ b_{vu} = b_{yx} \cdot \frac{c_x}{c_y} = 0.25 \times \frac{\left(\frac{1}{10}\right)}{10} = 0.25 \times \frac{1}{100} = 0.0025$$

(iv) Since $u = 5(x - 40) = \frac{x-40}{\frac{1}{5}} = \frac{x-A}{c_x}$ and $v = 2(y - 18) = \frac{y-18}{\frac{1}{2}} = \frac{y-B}{c_y}$

$$b_{yx} = b_{vu} \cdot \frac{c_y}{c_x} \quad \therefore\ b_{vu} = b_{yx} \cdot \frac{c_x}{c_y} = 1.6 \times \frac{\left(\frac{1}{5}\right)}{\left(\frac{1}{2}\right)} = 1.6 \times \frac{2}{5} = 0.64.$$

3.8 Precautions while using Regression

We know that regression is a functional relation between two correlated variables and hence we can estimate the value of dependent variable from it. The regression analysis is very useful in decision making in the practical fields like economics, trade, industry, education, psychology, sociology, medicine, planning etc. Inspite of vast application of regression analysis, some precautions are necessary while using it.

(1) The reliability of the estimate can be verified by the coefficient of determination $\left(R^2\right)$. So, we should use the estimate only after ascertaining the linearity of regression by the coefficient of determination.

(2) Another point which is necessary to keep in mind while using the regression analysis is, the regression relation obtained by the scatter diagram or by the method of least squares should not be used for the values which are very far from the given values of the independent variable.

e.g. If for some data, there is a high degree of correlation between the rainfall and the yield of wheat, we can say that as rainfall increases, yield of wheat also increases. Now, using the regression relation obtained from the given data, if for some value of rainfall, corresponding yield of wheat is to be estimated then the estimate of yield of wheat can be proper only when the value of rainfall is around the given values of rainfall. If there is heavy rain then the crop may get damaged and the yield of wheat may also decrease. In such a case, the dependent variable (yield) estimated from the above mentioned regression relation may be wrong.

Summary

- The concept of regression is studied under the assumption that two variables under study have cause-effect relationship.
- Regression : Functional relation between two related variables.
- Linear Regression : Functional relation between two related variables in which the change in the values of the variables are approximately in constant proportion and this relationship can be determined by a straight line.
- The value of the dependent variable can be estimated for some known value of the independent variable by using regression.
- Regression coefficient : The approximate change in the value of dependent variable for a unit change in the value of independent variable. It is also known as slope of the regression line.
- Error : Mistake occurring in estimating the value of the dependent variable.
- Coefficient of Determination : It is the square of the correlation coefficient between the observed value of dependent variable Y and its estimated values. In case of two variables, it is same as square of the coefficient of correlation between independent variable X and dependent variable Y.
- Using the coefficient determination, how much variation in the dependent variable Y is explained by the regression line can be known and the reliability of the regression model can also be known.
- The regression relation should not be used for the values which are very far from the given values of the independent variable.

List of Formulae :

Equation of Regression Line

$\hat{y} = a + bx$

Where, $b = b_{yx} =$ Regression Coefficient

(1) $b = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

(2) $b = \frac{n\Sigma xy-(\Sigma x)(\Sigma y)}{n\Sigma x^2-(\Sigma x)^2}$

(3) $b = \frac{n\Sigma uv-(\Sigma u)(\Sigma v)}{n\Sigma u^2-(\Sigma u)^2}$ Here, $u = x - A$ and $v = y - B$

(4) $b = \frac{n\Sigma uv-(\Sigma u)(\Sigma v)}{n\Sigma u^2-(\Sigma u)^2} \times \frac{c_y}{c_x}$ Here, $u = \frac{x-A}{c_x}$ and $v = \frac{y-B}{c_y}$

(5) $b = r \cdot \frac{s_y}{s_x}$

(6) $b = \frac{Cov(x, y)}{s_x^2}$

(7) $a = \bar{y} - b\bar{x}$

(8) Coefficient of Determination $R^2 = [r(y, \hat{y})]^2 = [r(x, y)]^2 = r^2$

Exercise 3

Section A

Find the correct option for the following multiple choice questions :

1. Which of the following indicates the functional relation between the two variables ?

(a) Correlation (b) Regression (c) Mean (d) Variance

2. The best fitted line of regression can be obtained by which method ?

(a) Least Square Method (b) Karl Pearson's Method

(c) Maximum Square Method (d) Bowley's Method

3. In usual notation, what is b_{yx} ?

(a) Intercept (b) Dependent Variable

(c) The approximate change in the value of Y for a unit change in the value of X.

(d) The approximate change in the value of X for a unit change in the value of Y.

4. Which of the following is correct ?

(a) $b_{yx} = r \cdot \frac{s_x}{s_y}$ (b) $b_{yx} = r \cdot \frac{s_y^2}{s_x^2}$ (c) $b_{yx} = \frac{Cov(x, y)}{s_y^2}$ (d) $b_{yx} = r \cdot \frac{s_y}{s_x}$

5. The regression line always passes through which point ?

(a) $(\bar{x}, \bar{y})$ (b) $(0, \bar{y})$ (c) $(\bar{x}, 0)$ (d) $(0, 0)$

6. What is error e in estimation in case of line of regression of Y on X ?

(a) $y - \hat{y}$ (b) $\hat{x} - \hat{y}$ (c) $x - \hat{x}$ (d) $\hat{y} - \hat{x}$

7. Which regression line is used if the sale of a commodity depends on its advertisement cost ?

(a) Regression line of advertisement cost on sale

(b) Regression line of advertisement cost on advertisement cost

(c) Regression line of sales on advertisement cost

(d) Regression line of sales on sales

8. Which of the following is a regression line of Y on X ?

(a) $\hat{y} = a + bx + cx^2$ (b) $\hat{x} = c + by$ (c) $\hat{y} = a + bx$ (d) $\hat{y} = a + bx^2$

9. For which value of the correlation coefficient (r), the regression coefficient becomes zero ?

(a) 1 (b) -1 (c) $\frac{1}{2}$ (d) 0

10. What is coefficient of determination in the study of regression for two variables ?

(a) Product of two standard deviations (b) Square of correlation coefficient

(c) Square of covariance (d) Product of two variances

11. If the regression line is $\hat{y} = 10 + 3x$, what is the estimate of Y for $X = 20$?

(a) 13 (b) 60 (c) 70 (d) 203

12. What is the value of b_{yx} if the regression line is $2x + 3y - 50 = 0$?

(a) $\frac{3}{2}$ (b) $-\frac{3}{2}$ (c) $-\frac{2}{3}$ (d) 2

13. The regression line of Y on X is $\hat{y} = 30 - 1.5x$. What is the value of $\bar{y}$ if $\bar{x} = 10$?

(a) 28.5 (b) 20 (c) 15 (d) 45

14. If $u = \frac{x-15}{10}$ and $v = \frac{y-50}{2}$ and $b_{yx} = 7.5$, What is the value of b_{vu} ?

(a) 7.5 (b) 1.5 (c) 37.5 (d) 150

15. If $r = 0.8$, how much part of the total variation in the dependent variable can be explained by the regression model ?

(a) 80 % (b) 64 % (c) 36 % (d) 20 %

Section B

Answer the following questions in one sentence :

1. Define : Linear Regression
2. Define : Regression Coefficient
3. State the Linear Regression model.
4. What is an error in context with a regression line ?
5. Give the name of a method to obtain the best fitted regression line.
6. The regression coefficient is independent of which transformation ?
7. The regression coefficient is not independent of which transformation ?
8. What is the value of error if a sample point is on the fitted line ?
9. Will the regression coefficient change if the values of both the variables are doubled with the help of transformation of scale ?
10. If $r = 0.5, s_x = 2, s_y = 4$, what is the value of b_{yx} ?
11. If a regression line is $\hat{y} = 31.5 + 1.85x$, estimate Y for $X = 10$.
12. If Y and X have the relation $y = a + bx$, where $b > 0$ then what is the value of r ?
13. If $y = 5 - 3x$ is the relation between Y and X then what is the value of r ?

Section C

Answer the following questions :

1. What are the constants a and b in the regression line $\hat{y} = a + bx$?
2. The fitted regression line of Y on X is $\hat{y} = 23.2 - 1.2x$ and one of the observations used in fitting of the line is (6, 17). Find the error in estimating Y for $X = 6$.
3. If $\bar{x} = 30, \bar{y} = 20$ and $b = 0.6$, find the intercept of the regression line of Y on X and write equation of the line.
4. Interpret $b_{yx} = 5$.
5. If $b = 1.5, r = 0.8$ and standard deviation of X is 1.6, find the standard deviation of Y.
6. If the regression coefficient of the regression line of Y on X is 0.6 and the standard deviations of X and Y are 5 and 3 respectively, find the coefficient of determination.
7. If the regression line of Y on X is $\hat{y} = 35 + 2x$ and $Cov(x, y) = 50$, find the standard deviation of X.
8. For the regression line given in the previous question (7), if the value of Y is to be increased by 10 units, how many units should be increased in the value of X ?
9. If $\bar{x} = 10, \bar{y} = 25, \Sigma(x-10)(y-25) = 120$ and $\Sigma(x-10)^2 = 100$, find the values of a and b for the regression line of Y on X.
10. If $b_{yx} = 0.75, u = 6(x-20)$ and $v = 2(y-15)$ for the data in the study of a regression line then find the value of b_{vu}.

Section D

Answer the following questions :

1. Explain the statement, "There is a cause and effect relationship between two variables" by giving a suitable example. Also define independent variable and dependent variable.
2. Explain the method of scatter diagram for fitting a line of regression and state its limitation.
3. Explain the method of least square for fitting a regression line.
4. State the utility of regression.
5. State properties of regression coefficient. Also state the point through which a regression line always passes.
6. Explain : coefficient of determination
7. State precautions which are necessary while using the regression.
8. For two related variables X and Y, $\Sigma(x-\bar{x})^2 = 80$, $\Sigma(x-\bar{x})(y-\bar{y}) = 60$, $\bar{x} = 8$, $\bar{y} = 10$. Obtain the regression line of Y on X.
9. If $\bar{x} = 30$, $\bar{y} = 50$, $r = 0.8$ and the standard deviations of X and Y are 2 and 5 respectively, obtain the regression line of Y on X.
10. If the regression line of Y on X is $\hat{y} = 11 + 3x$ and $s_x : s_y = 3:10$, find the coefficient of determination.
11. In usual notations, $n = 7$, $\Sigma u = 2$, $\Sigma v = 25$, $\Sigma u^2 = 160$ and $\Sigma uv = 409$. Obtain the regression coefficient of a regression line of Y on X and interpret it.
12. If $b_{yx} = 0.8$ then find the value of b_{vu} for the following u and v.

    (i) $u = x - 105$ and $v = y - 90$

    (ii) $u = \frac{x-1400}{100}$ and $v = \frac{y-750}{50}$

    (iii) $u = 10(x - 4.6)$ and $v = y - 75$
13. The following results are obtained for a bivariate data.

Particulars	***x***	***y***
No. of observations	8	
Mean	100	100
The sum of squares of deviations taken from mean	130	145
The sum of product of deviations taken from mean	115	

Obtain the regression line of Y on X.

Section E

Solve the following :

1. A manager of the an I.T. company has collected the following information regarding the years of job and monthly income of seven marketing executives.

Years of job	10	6	8	5	9	7	11
Monthly income (ten thousand ₹)	11	7	9	5	6	8	10

Obtain the regression line of the monthly income on the years of job of the marketing executives.

2. The information collected regarding price (in ₹) of a commodity and its supply (in hundred units) is as follows.

Price (₹)	59	60	61	62	64	57	58	59
Supply (hundred units)	78	82	82	79	81	77	78	75

Obtain the regression line of the supply on the price.

3. The following information is obtained for monthly advertisement cost and the sales of the last year for a company providing online shopping.

Particulars	**Advertisement cost (ten thousand ₹)**	**Sales (lakh ₹)**
Mean	10	90
Standard Deviation	3	12
$r = 0.8$		

Obtain the regression line of the sales on the advertisement cost.

4. The following results are obtained from the information of average rain and yield of a crop per acre in the last ten years of an arid region.

Particulars	**Rainfall (cm)**	**Yield of crop (kg)**
Mean	18	970
Standard Deviation	2	38
Correlation Coefficient = 0.6		

Estimate the yield of the crop if it rains 20 cms.

5. The information of investment (in lakh ₹) and its market price (in lakh ₹) after six months in share market in the last seven years for a Mutual Fund Company is obtained as follows.

Particulars	**Investment (lakh ₹)** x	**Market price after six months (lakh ₹)** y
Mean	40	50
Variance	100	256
Covariance = 80		

Obtain the regression line of Y on X and estimate the market price in the share market after six months if there is an investment of ₹ 45 lakh in a year.

Section F

Solve the following :

1. Obtain the regression line of the demand on the price using the following information collected for the demand and the price of a commodity. Estimate the demand of the commodity if price is ₹ 40.

Price (₹)	38	36	37	37	36	38	39	36	38
Demand (hundred units)	12	18	15	12	17	13	13	15	12

2. The information regarding the experience (in years) of eight workers on a machine and their performance ratings based on the nondefective units they manufactured in every 100 units is as follows.

Experience of worker (years)	5	12	15	8	20	18	22	25
Performance rating	80	82	85	81	90	90	95	97

Obtain the regression line of the performance rating on the experience and estimate the performance rating if a worker has an experience of 17 years.

3. The information regarding daily income (in ₹) and expenditure (in ₹) of five labour families earning by daily work.

Daily income (₹)	200	300	400	600	900
Expenditure (₹)	180	270	320	480	700

Obtain the regression line of the expenditure on the daily income. Estimate the expenditure of a family having daily income of ₹ 500.

4. The following information is collected by a firm to know the effect of an advertisement campaign.

Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Advertisement cost (ten thousand ₹)	12	15	15	23	24	38	42	48
Sales (crore ₹)	5	5.6	5.8	7	7.2	8.8	9.2	9.5

Obtain the regression line of sales on the advertisement cost. Estimate the sales when the advertisement cost is ₹ 5,00,000.

5. The information of eight construction companies regarding the number of contracts received in a year and the annual profit is as follows.

No. of contracts	2	5	9	12	6	4	8	10
Annual profit (lakh ₹)	100	300	700	1000	350	250	700	750

Obtain the regression line of the annual profit on the number of contracts. Verify the reliability of the regression model.

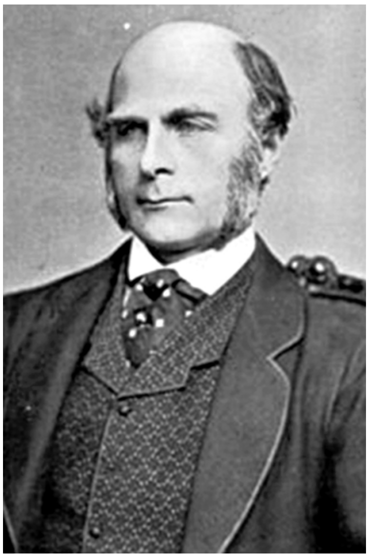
6. Obtain the regression line of Y on X from the following data and estimate Y for $X = 30$.

$n = 10, \Sigma x = 250, \Sigma y = 300, \Sigma xy = 7900, \Sigma x^2 = 6500$

7. The following results are obtained for a data.

$n = 12, \Sigma x = 30, \Sigma y = 5, \Sigma x^2 = 670, \Sigma xy = 344$

Later on, it was known that one pair (10, 14) was wrongly taken as (11, 4). By correcting the above measures, obtain the regression line of Y on X. Estimate Y for $X = 5$.



Sir Francis Galton
(1822 –1911)

Sir Francis Galton was an English Victorian statistician, progressive, polymath, sociologist, psychologist, anthropologist, eugenicist, tropical explorer, geographer, inventor, meteorologist, protogeneticist and psychometrician. He was knighted in 1909.

Galton produced over 340 papers and books. He also created the statistical concept of correlation and widely promoted regression towards the mean. He was the first to apply statistical methods to the study of human differences and inheritance of intelligence, and introduced the use of questionnaires and surveys for collecting data on human communities, which he needed for genealogical and biographical works and for his anthropometric studies.

He was a pioneer in eugenics, coining the term itself and the phrase "nature versus nature". His book Hereditary Genius (1869) was the first social scientific attempt to study genius and greatness.

As an investigator of the human mind, he founded psychometrics (the science of measuring mental faculties) and differential psychology and the lexical hypothesis of personality. He devised a method for classifying fingerprints that proved useful in forensic science.

"Imperfect prediction, despite being imperfect can be valuable for decision making process."

– Michael Kattan

4

(Time Series)

Contents :

4.1 Time Series

Introduction

Two related variables are studied by different methods in Statistics. A special method is applied to study the dependent variable among these related variables by taking time as an independent variable. The data related to values of the variable changing with time are studied in Economics, Sociology and Business Statistics. For example, population of a country, agricultural production, wholesale price index, unemployment statistics, import-export information, annual production of a certain factory, data from share market, bank interest rates, the hourly temperature measured in a city, etc. are presented with respect to time. These data are called time series as they are dependent on time.

Meaning of Time Series

The statistical data collected at specific intervals of time and arranged in a chronological order is called Time Series. Time series consists of values of a variable associated with time. The estimated value of this variable for the future can be obtained if the values of such a variable are studied over a long period of time. These forecasts are very useful for future planning. For example, the direction, proportion and pattern of variations in the population of a certain region can be known by studying its time series. The necessary infrastructure, medical facilities, employment opportunities, education can be planned for the people of this region in future. The fluctuations in the prices of shares can be known by studying the time series of share prices of different companies and the investors can decide about buying or selling shares. The temperature measured at different places and time as well as the data of rainfall indicate the global changes in the weather which is useful to form policies for conservation of environment. In recent times, time series is extensively used in different methods of business analytics.

The data regarding changes in a variable at specific time interval are shown in the time series. The unit of time is dependent on the variable under study. For example, population data are obtained every ten years, data about total sales tax collected are available annually, quarterly interests are calculated in banks, monthly profits of shops are given, the time for bacterial growth is in hours, etc.

We will see the following illustrations of time series :

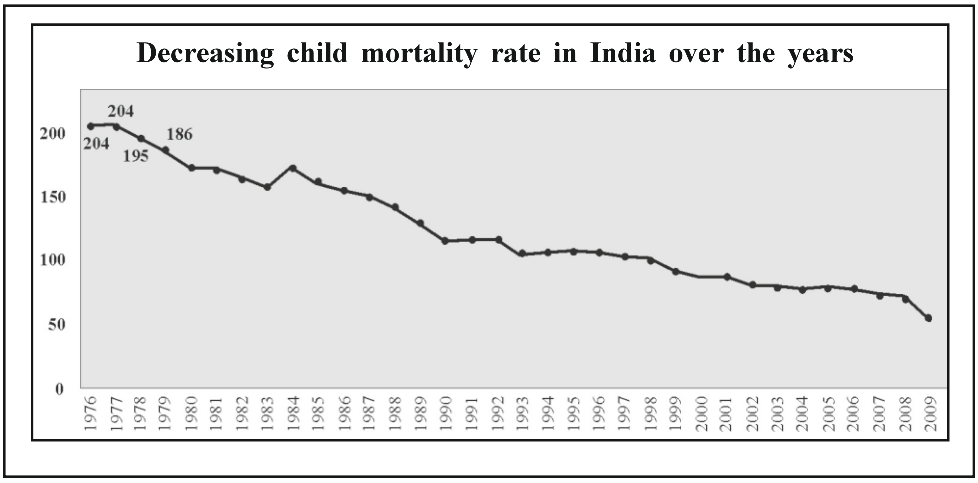
(1)

MACRO FUNDAMENTALS				(in %)
Year	**GDP Growth**	**Investment Growth**	**Average WPI Inflation**	**CAD (As % of GDP)**
2002-03	4.0	–0.4	3.4	0
2003-04	8.1	10.6	5.5	0
2004-05	7.0	24.0	6.5	0.4
2005-06	9.5	16.2	4.5	1.2
2006-07	9.6	13.8	6.6	1.0
2007-08	9.3	16.2	4.7	1.3
2008-09	6.7	3.5	8.1	2.3
2009-10	8.6	7.7	3.8	2.8
2010-11	9.3	14.0	9.6	2.8
2011-12	6.2	4.4	8.9	4.2
2012-13	5.4*	2.3*	7.6**	4.7*

* April-September ** April-December

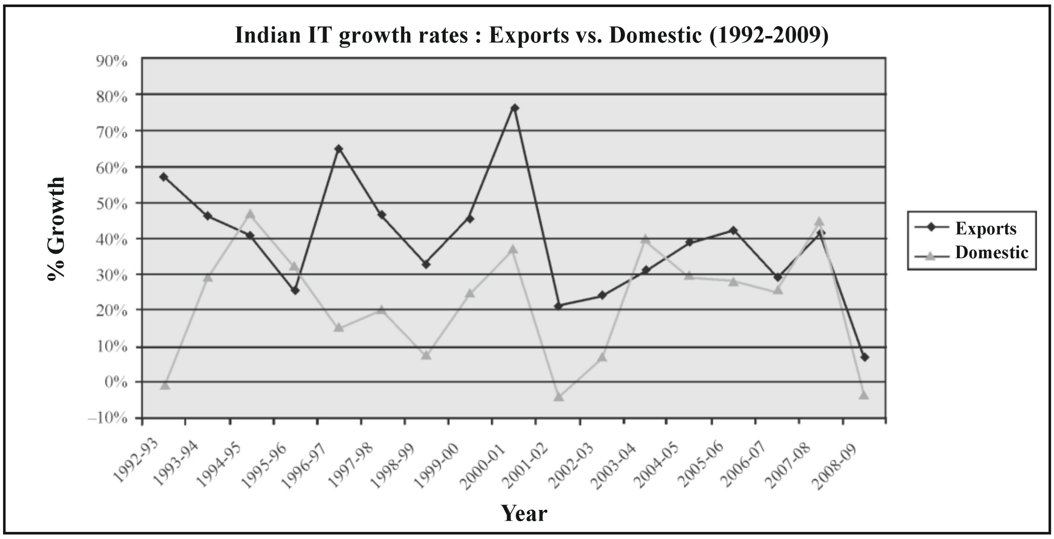
This time series gives information about macro fundamentals of different years which includes growth in Gross Domestic Product (GDP) along with percentage investment growth, Wholesale Price Index (WPI) and Current Account Deficit (CAD).

(2)



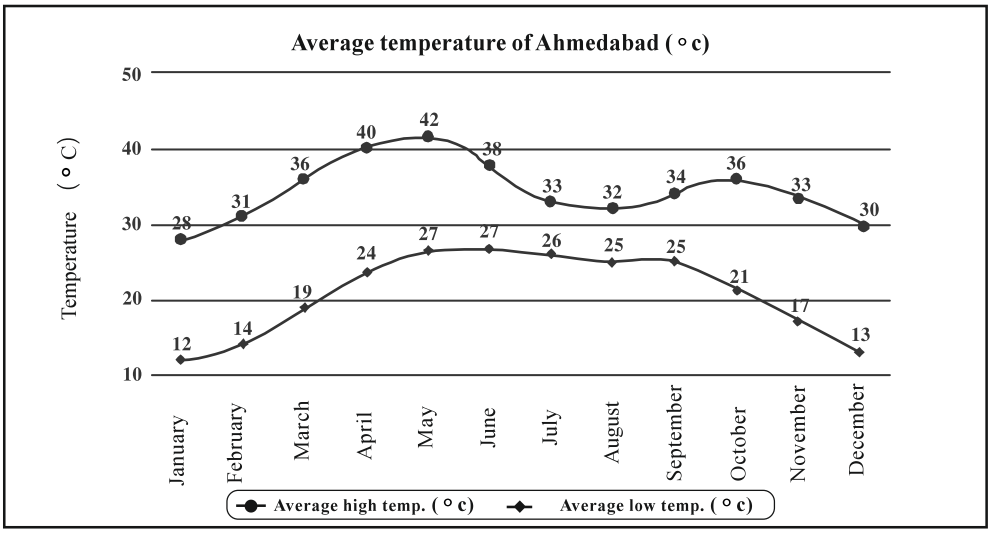
Time Series showing death rate among new-born children in India over different years.

(3)



A comparative study of two time series showing growth in IT sector.

(4)



Time Series showing average monthly maximum and minimum temperatures in Ahmedabad.

Importance of Time Series

The information collected in the form of time series is extremely essential in modern era due to increasing uncertainties in trade and business activities. The study of time series gains importance due to following reasons.

(1) The direction and pattern of variations in the values of the series can be known from the past data.

(2) The variations in the future can be estimated from the extent of variation in the values of the series.

(3) Important decisions can be taken from the estimated values of the future and industrial as well as government policies can be easily framed.

(4) Two or more industrialists or government institutions can make a comparative study from the data of the time series obtained by them.

Definition of Time Series

Time series is defined as follows :

'A time series is a set of observations taken at specified time periods.'

Usually these observations are taken at equal intervals of time.

The time is taken as an independent variable in the time series which will be denoted by t and the dependent variable associated with it will be denoted by y_t. Thus, we shall represent the time series for different units of time as follows :

Time t	1	2	3	...	n
Variable y_t	y_1	y_2	y_3	...	y_n

Uses of Time Series

The variations in the variable of a time series which changes with time are not caused by one specific reason. The variable of a time series is influenced by various factors and all these factors have an effect on the given variable. For example, the price of wheat in the wholesale market changes with time due to various different reasons such as the production of wheat at that time, demand of wheat, the cost of transporting the production to the market, etc. Each of these factors is dependent on other forces. For instance, the production of wheat is affected by various factors like the weather at that time, irrigation facility, the quality of seeds. It is necessary to study the various ways in which these factors affect the variable of the time series. Such a detailed study conducted for the time series is called as the analysis of time series which is done in the following two stages :

(1) To identify the various factors affecting the variable of the time series.

(2) To segregate these factors and determine the extent of effect of each factor on the given variable.

The analysis of time series done in this way is useful in trade, science, social and political fields as follows :

(1) It is possible to know the past situation and use it to obtain the type and measure of the variation.

(2) It is possible to estimate the value of the variable in future using statistical methods.

(3) Proper decisions can be taken for the future using the estimated values and activities can be planned accordingly.

(4) A comparative study can be carried out for the variations in the given variable at different places or time intervals.

(5) The estimates obtained from the past data can be compared with the present values and the reasons for the discrepancies between them can be investigated.

Activity

Collect the information with the help of your teachers about the percentage of students passing 12th standard from your school in the last 10 years and present it in the form of a time series.

4.2 Components of Time Series

We saw that there is a composite effect of many factors on the variable of the time series which brings fluctuations in the values of the variable. After observing different time series, it is known that the variations in the variable exhibit a specific pattern. The time series can be decomposed in the following components based on this pattern :

(1) Long-term Component or Trend : The variation seen in the variable of a time series over a very long period of time is the effect of long-term component or trend. The variable of a time series is generally found to have a continuous increase or decrease. This phenomenon is due to trend. For example, decreasing value of rupee in the international market, increasing usage of mobile phones, rising population of the country, decreasing death rates, etc. The intermittent short-term variations are ignored as the long-term variations in the time series are studied in the trend. The overall changes taking place in the variable of the series are considered here. The factors responsible for these changes produce a very slow pace of variation. For example, the number of literates are increasing in India but this change is taking place slowly in the last 60-70 years. Generally, the causes of such variations are the changing customs in the society, changes in tastes and choices of the people, technological changes in the industry, etc.

The trend of a time series is experienced after a very long time where 'long time' is a completely relative term. An interval of 10-15 years is required to know the trend in agricultural yield or industrial production whereas it may be clear within 4-5 years in the sale of electronic goods. The trend in the series having almost continuous increase or decrease is called a linear trend, which is generally observed in most of the time series. The data in economics, commerce and trade have series in which the rate of increase or decrease of the values does not remain constant. The rate of increase in the values of such series is very slow initially which goes on increasing slowly. The values stabilize after a certain interval of time and then start decreasing gradually. The trend of the series in such a situation is said to be Non-linear or Curvi-linear.

We will denote the component showing trend in the variable y_t of the time series as 'T_t'.

(2) Seasonal Component : The variation occurring in the time series variable almost regularly over a very short period of time is the effect of seasonal componont. The period of oscillation of such variations is usually less than a year. It is necessary to record the short-tem values in the time series to study these variations. It is not possible to get the information about the seasonal component if the yearly values of the given variable are available. The seasonal component affects the time series as follows :

(i) Effect of natural factors : The variations in the values of the time series occur in association with the seasons or weather fluctuations. Such variations occur at almost regular intervals. For example, the demand of fans, coolers or A.C. increases during summer whereas the demand of woollens increases in winter, the market prices reduce when the new crop is ready, etc.

(ii) Effect of man-made factors : The variations occurring regularly in less than one year period are caused by the festivals, customs in the society, habits of people, etc. For example, the purchase of ornaments increases during marriage season, kites are in demand during Uttarayan, the number of customers increases at the restaurants or theatres during weekends, increase in the purchase of clothing and gift articles during festivals, etc.

As the period of oscillation of these types of variations is almost certain, they are called regular variations. If the time and measure of these variations are known then the traders, producers are benefitted as higher profit can be earned by a control over their inventory.

We shall denote this short-term component of the time series by 'S_t'.

(3) Cyclical Component : The variations occurring in the time series variable at approximately regular intervals of more than one year are the effects of cyclical component. The variations occurring due to this component are less regular as compared to seasonal component. The period of oscillation of these variations can be 2 to 10 years and in specific circumstances it can also be 10-15 years. The cyclical component is also considered as a short-term component as the time interval of variations due to this component is less than that of the time for the whole series which can be 40-50 years or even more. The cycles of boom and recession are examples of these variations. These cycles pass through the four stages namely, depression, recovery, boom, recession. These variations are found in the time series of trade and financial matters such as production, price of an item, prices of shares in share market, investment, etc. The traders can plan suitably with the help of estimate of time and measure of these variations.

This component of the time series is denoted by 'C_t'.

(4) Random or Irregular Component : The effect of irregular or random component is also seen in the variations of the variable of time series in addition to the approximately regular short-term components like seasonal and cyclical which also give short-term effect. If the values in the series change due to sudden and unpredictable causes, the changes are called as random variations. The time-interval and effect of this variation is not certain. The variation which cannot be attributed to any one of the trend, seasonal component, cyclical component is the effect of random component. These fluctuations appear completely unexpected and are irregular. It cannot be predicted, it does repeat regularly and cannot be controlled. This variation occurs due to natural disasters like earthquake, floods or due to man-made predicaments like war, strike, political upheaval. The estimates can contain error due to this component.

This component is denoted by 'R_t'.

The value of the variable of the time series y_t based on time t is determined with the combined effect of trend (T_t), seasonal component (S_t), cyclical component (C_t) and random component (R_t). This relationship is shown as follows in the additive model of the time series :

$$y_t = T_t + S_t + C_t + R_t$$

The seasonal component (S_t) does not appear if the yearly values of the variable are given for the time series. To find the effect of each component, trend (T_t) is found first using the given values y_t. After substracting it from y_t, the residual variation shows short-term components $\left(S_t, C_t, R_t\right)$. Then seasonal component (if available) and cyclical components are found. Random component is found in the end as $R_t = y_t - \left(T_t + S_t + C_t\right)$. The future estimate of the variable $(\hat{y}_t)$ is found by estimating the trend value and then adding the effect of each component at the given time (t) as mentioned above.

Activity

Prepare a time series of units of electricity consumption for the past one year from the electricity bills of your house. Identify the component of time series showing its effect on the variation in the variable of this series.

4.3 Methods for Determining Trend

Trend is an important component of time series. We will study the following methods to estimate it :

4.3.1 Graphical Method

This is the easiest method to find trend. The points are plotted on the graph paper by taking the independent variable, time (t), on X-axis and the dependent variable y_t on Y-axis. These points are joined in their order by line segments. This shows the variation in the values of the variable. A smooth curve is then drawn through the middle of the points by personal judgement. This curve shows the trend by ignoring the short-term fluctuations in the series. The future estimates are obtained by extending the curve thus drawn.

The merits and limitations of graphical method are as follows.

Merits :

(1) This method is easy to understand and use.

(2) The trend can be found without any mathematical formula or calculations.

(3) This method can be used even if the trend is not linear.

(4) The judgement about the type of curve to be fitted for obtaining trend can be given by this method.

Limitations :

(1) It is possible that different people draw different curves. Hence, the uniformity is not maintained in the trend and its estimates.

(2) The estimates cannot be accurate as this is not a mathematical method and it is not possible to know the reliability of the estimates.

Illustration 1 : The yearly production (in tons) of a factory is as follows. Obtain the trend using graphical method.

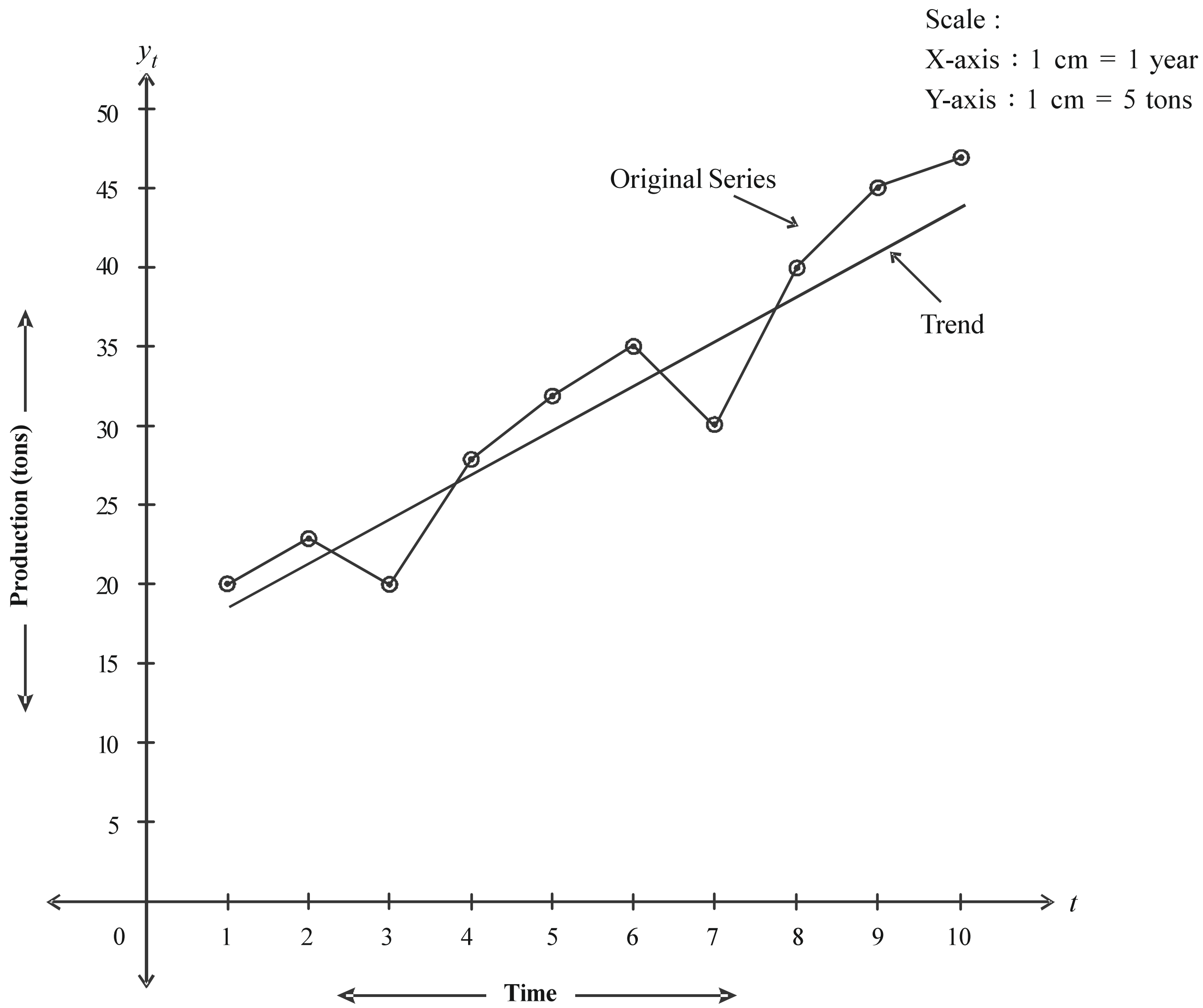
Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Production (tons)	20	23	20	28	32	35	30	40	45	47

We will represent these data as the following time series :

Time t	1	2	3	4	5	6	7	8	9	10
Production (ton) y_t	20	23	20	28	32	35	30	40	45	47

We will plot these points on a graph by taking t on X-axis and production y_t on Y-axis. The pattern of points indicates that linear trend is more suitable.

Downloaded from https:// www.studiestoday.com



The line passing through the middle of the points shows trend.

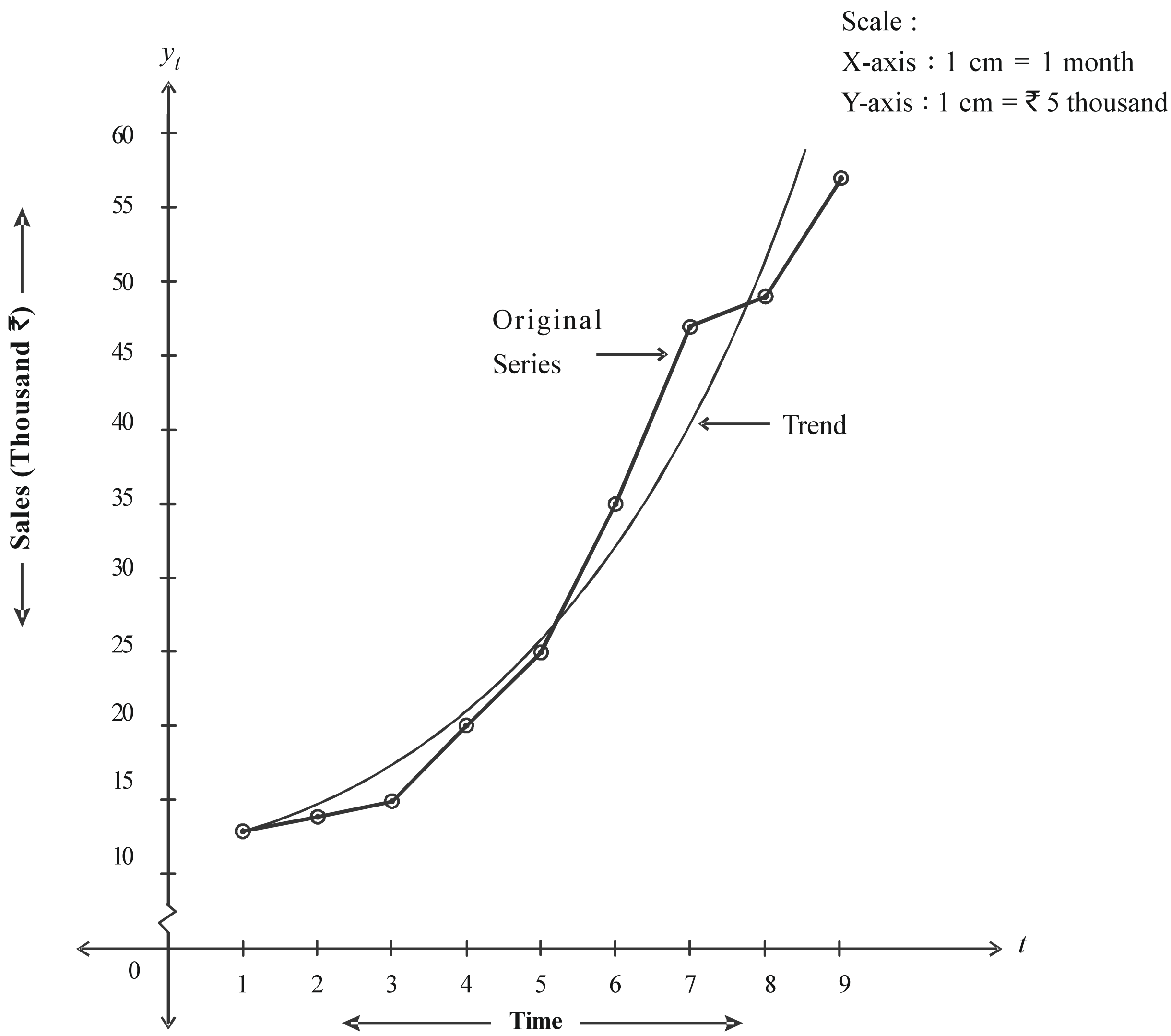
Illustration 2 : The data about monthly sales (in thousand ₹) of a company are given in the following table. Obtain trend using grapical method.

Month	Jan.	Feb.	March	April	May	June	July	August	Sept.
Sales (thousand ₹)	13	14	15	20	25	35	47	49	57

We will take the following time series for these data :

Time t	1	2	3	4	5	6	7	8	9
Sale (thousand ₹) y_t	13	14	15	20	25	35	47	49	57

We will plot these points on a graph by taking t on X-axis and sales y_t on Y-axis. It indicates that the non-linear trend is more suitable for these data.



The curve passing through the middle of the points on the graph shows trend.

EXERCISE 4.1

1. The information about the capacity (in lakh tons) to load ships at a port each year is given below. Find the linear trend using graphical method.

Year	2008	2009	2010	2011	2012	2013	2014	2015
Capacity (lakh tons)	90	97	108	111	127	148	169	200

2. The number of tourists (in thousand) visiting a certain tourist place is as follows. Find the trend using a suitable graph.

Year	2010	2011	2012	2013	2014	2015	2016
No. of tourists (thousand)	5	7	10	14	30	41	50

3. The data regarding number of girls (y_t) per 1000 boys in the age group 0-6 years of a state are given in the following table. Obtain the linear trend using graphical method.

Year	1961	1971	1981	1991	2001	2011
y_t	956	948	947	928	883	890

4. The data about the closing prices of shares of a company for 10 days are given in the following table. Obtain the trend using graphical method.

Day	1	2	3	4	5	6	7	8	9	10
Price of share (₹)	297	300	304	299	324	320	318	324	329	328

*

4.3.2 Method of Least Squares

As a limitation of the graphical method we saw that if a mathematical technique is not used then the trend and the estimates obtained from it change from person to person and its reliability cannot be known. If we want to find the linear trend of the time series by a mathematical method, we will require a specific linear equation which represents trend. We have studied the method of least squares in the chapter of regression to fit a linear equation to the given data which will be used to find the linear trend in a time series.

Suppose the values of the variable y_t in the time series are available based on time t. We shall use the linear model $y_t = \alpha + \beta t + u_t$ (where u_t is disturbance variable) to represent the relation between them. The estimated values $\hat{y}_t$ of y_t can be found by fitting this model using the method of least squares. We will use the equation $\hat{y}_t = a + bt$ for this as shown in chapter 3.

We will ignore the suffix t in y_t for simplicity and consider $\hat{y} = a + bt$. The dependent variable is y for the independent variable t.

The constants a and b are obtained by the method of least squares as follows :

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2} \quad \text{and} \quad a = \bar{y} - b\bar{t},$$

where n = no. of observations

The linear equation thus obtained is the best linear equation for the given data.

The estimate of trend for the future is obtained using this linear equation.

Note : Other equations like polynomial, exponential equations can also be fitted besides the linear equation to find trend.

The merits and limitations of the method of least squares are as follows.

Merits :

(1) This method is absolutely mathematical and hence the future estimates do not change subjectively with the person.

(2) The trend estimates can be obtained by this method for each of the given values of t.

(3) The trend estimates can also be obtained for intermediate periods as the trend values are obtained using an equation. For example, the trend estimate for the period in the centre of the second and third year can be found by taking $t = 2.5$.

Limitations :

(1) This method requires extensive calculations to find trend.

(2) The reliability of the estimated values obtained by this method is less if an appropriate type of trend curve and its suitable equation is not fitted.

Illustration 3 : The profit earned (in lakh ₹) by a company making computers is as follows. Find the linear equation for the trend from these data by least square method and estimate the profit for the year 2017.

Year	2011	2012	2013	2014	2015
Profit (Lakh ₹)	31	35	39	41	44

The values of profit are given for $n = 5$ years. We will thus denote the given years as $t = 1, 2, \ldots, 5$ respectively.

Calculation for fitting linear trend

Year	Profit y	t	t^2	ty
2011	31	1	1	31
2012	35	2	4	70
2013	39	3	9	117
2014	41	4	16	164
2015	44	5	25	220
Total	**190**	**15**	**55**	**602**

$$\bar{t} = \frac{\Sigma t}{n} = \frac{15}{5} = 3, \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{190}{5} = 38$$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{5 \times 602 - 15 \times 190}{5 \times 55 - (15)^2}$$

$$= \frac{3010 - 2850}{275 - 225}$$

$$= \frac{160}{50}$$

$$= 3.2$$

$$a = \bar{y} - b\bar{t}$$

$$= 38 - 3.2 \times 3$$

$$= 38 - 9.6$$

$$= 28.4$$

Equation for trend $\hat{y} = a + bt$

$$\therefore \hat{y} = 28.4 + 3.2\,t$$

We take t = 7 for the year 2017.

$$\therefore \quad \hat{y} = 28.4 + 3.2 \times 7$$
$$= 28.4 + 22.4$$
$$= 50.8$$
$$\therefore \quad \hat{y} = ₹\ 50.8 \text{ lakh}$$

Thus, the estimated trend value of profit for the year 2017 is ₹ 50.8 lakh.

Illustration 4 : The dropout rate of students of standard 1 to 5 from primary schools of a district is as follows :

Year	2009 – 10	2010 – 11	2011 – 12	2012 – 13	2013 – 14	2014 – 15	2015 – 16
Dropout rate	3.24	2.98	2.29	2.20	2.09	2.07	2.04

Estimate the dropout rate for students from standard 1 to 5 for the year 2016-17 and 2017-18 by fitting a linear equation for trend.

The data are given for n = 7 years. We will thus denote the given years as t = 1, 2, ... , 7 respectively.

Calculations for fitting linear trend

Year	**Dropout rate** y	t	t^2	ty
2009-10	3.24	1	1	3.24
2010-11	2.98	2	4	5.96
2011-12	2.29	3	9	6.87
2012-13	2.20	4	16	8.80
2013-14	2.09	5	25	10.45
2014-15	2.07	6	36	12.42
2015-16	2.04	7	49	14.28
Total	**16.91**	**28**	**140**	**62.02**

$$\bar{t} = \frac{\Sigma t}{n} = \frac{28}{7} = 4, \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{16.91}{7} = 2.4157 \simeq 2.42$$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{7 \times 62.02 - 28 \times 16.91}{7 \times 140 - (28)^2}$$

$$= \frac{434.14 - 473.48}{980 - 784}$$

$$= \frac{-39.34}{196}$$

$$= -0.2007$$

$$\simeq -0.2$$

$$a = \overline{y} - b\overline{t}$$

$$= 2.42 - (-0.2) \times 4$$

$$= 2.42 + 0.8$$

$$= 3.22$$

Equation for trend $\hat{y} = a + bt$

$$\therefore \hat{y} = 3.22 + (-0.2)\, t$$

$$= 3.22 - 0.2\, t$$

We take $t = 8$ for the year 2016-17.

$$\therefore \hat{y} = 3.22 - 0.2 \times 8$$

$$= 3.22 - 1.6$$

$$= 1.62$$

We take $t = 9$ for the year 2017-18.

$$\therefore \hat{y} = 3.22 - 0.2 \times 9$$

$$= 3.22 - 1.8$$

$$= 1.42$$

Thus, the estimates of dropout rates for the students of standard 1 to 5 in this district for the years 2016-17 and 2017-18 are 1.62 and 1.42 respectively.

Illustration 5 : The data of population (in lakh) of a taluka are given in the following table. Fit a linear equation for the data and find the trend value for each year. Also find the trend estimate for the population in the year 2021.

Year	1951	1961	1971	1981	1991	2001	2011
Population (lakh)	15.1	16.9	18.7	20.1	21.6	25.7	27.1

The data about population are given which are associated with each decade. We will take $t = 1, 2, \ldots, 7$ respectively for the given years. Hence, we get $n = 7$.

Calculation for fitting linear trend

Year	Population (lakh) y	t	t^2	ty	Trend values $\hat{y} = 12.66 + 2.02\ t$
1951	15.1	1	1	15.1	14.68
1961	16.9	2	4	33.8	16.7
1971	18.7	3	9	56.1	18.72
1981	20.1	4	16	80.4	20.74
1991	21.6	5	25	108	22.76
2001	25.7	6	36	154.2	24.78
2011	27.1	7	49	189.7	26.8
Total	**145.2**	**28**	**140**	**637.3**	

$$\bar{t} = \frac{\Sigma t}{n} = \frac{28}{7} = 4\ , \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{145.2}{7} = 20.7429 \simeq 20.74$$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{7 \times 637.3 - 28 \times 145.2}{7 \times 140 - (28)^2}$$

$$= \frac{4461.1 - 4065.6}{980 - 784}$$

$$= \frac{395.5}{196}$$

$$= 2.0179$$

$$\simeq 2.02$$

$$a = \bar{y} - b\bar{t}$$

$$= 20.74 - 2.02 \times 4$$

$$= 20.74 - 8.08$$

$$= 12.66$$

Equation for trend $\hat{y} = a + bt$

$$\therefore \hat{y} = 12.66 + 2.02\, t$$

We take $t = 1, 2, ..., 7$ for each of the given year respectively to find the values of trend. Taking $t = 1$,

$$\hat{y} = 12.66 + 2.02 \times 1$$
$$= 12.66 + 2.02$$
$$= 14.68$$
$$\therefore \quad \hat{y} = 14.68 \text{ lakh}$$

Similarly we will take $t = 2, 3, ..., 7$ to find the remaining values of trend and show them in the table.

It can be seen here that the values of $\hat{y}$ increase successively by 2.02.

We take $t = 8$ for the year 2021

$$\hat{y} = 12.66 + 2.02 \times 8$$
$$= 12.66 + 16.16$$
$$= 28.82$$
$$\therefore \quad \hat{y} = 28.82 \text{ lakh}$$

Thus, the estimate for trend value for the population for the year 2021 is 28.82 lakh.

Illustration 6 : The data about monthly sales (in thousand ₹) of a company are given in the following table. Fit a linear trend and show it grapically. Estimate the sale for the month of August using the equation obtained.

Month	January	February	March	April	May	June
Sale (thousand ₹)	80	85	90	76	82	88

The data for $n = 6$ months are given here. Hence, we will take $t = 1, 2, ..., 6$ for the given months respectively.

Calculation for fitting linear Trend

Month	**Sales y (thousand ₹)**	t	t^2	ty	$\hat{y} = 81.79 + 0.49\, t$
January	80	1	1	80	82.28
February	85	2	4	170	82.77
March	90	3	9	270	83.26
April	76	4	16	304	83.75
May	82	5	25	410	84.24
June	88	6	36	528	84.73
Total	**501**	**21**	**91**	**1762**	

$$\bar{t} = \frac{\Sigma t}{n} = \frac{21}{6} = 3.5, \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{501}{6} = 83.5$$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{6 \times 1762 - 21 \times 501}{6 \times 91 - (21)^2}$$

$$= \frac{10572 - 10521}{546 - 441}$$

$$= \frac{51}{105}$$

$$= 0.4857$$

$$\simeq 0.49$$

$$a = \bar{y} - b\bar{t}$$

$$= 83.5 - 0.49 \times 3.5$$

$$= 83.5 - 1.715$$

$$= 81.785$$

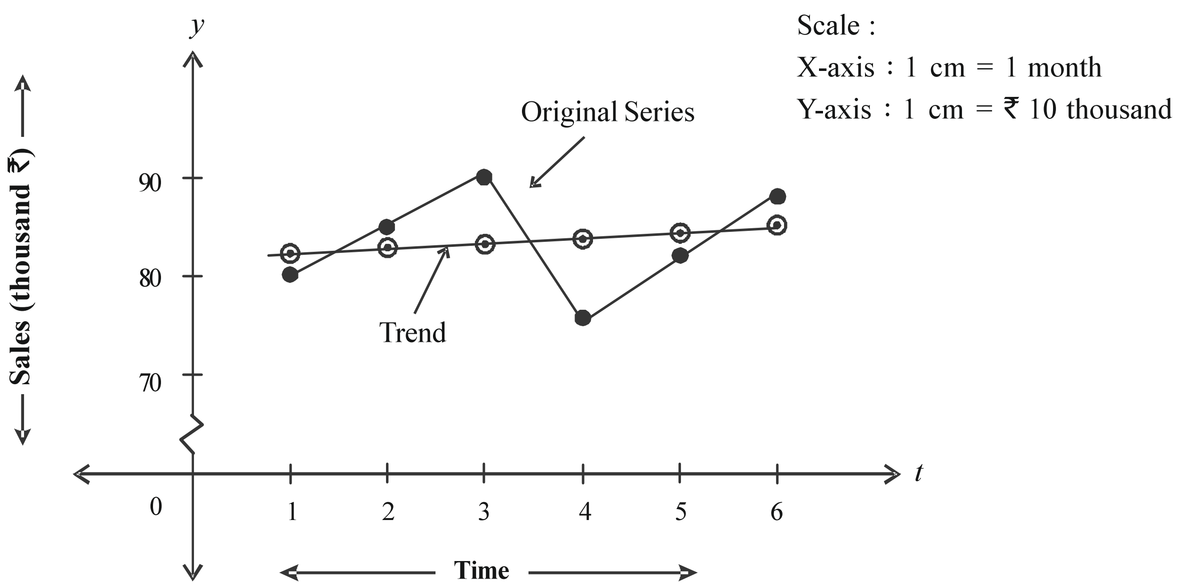
$$\simeq 81.79$$

Equation for trend $\hat{y} = a + bt$

$$\therefore \hat{y} = 81.79 + 0.49\,t$$

By substituting $t = 1, 2, ... , 6$ successively, we get the corresponding values of $\hat{y}$ which are shown in the table.

The trend values and the values in the given series can be shown in the graph as follows :



Now we take $t = 8$ for the month of August

$$\hat{y} = 81.79 + 0.49 \times 8$$

$$= 81.79 + 3.92$$

$$= 85.71$$

$$\therefore \hat{y} = ₹\ 85.71 \text{ thousand}$$

Thus, the trend estimate for sales of this company in the month of August is ₹ 85.71 thousand.

Note : It is not necessary to take all the values of $\hat{y}$ to draw the line representing the linear equation. The equation of trend can be shown in the graph by joining the values of $\hat{y}$ corresponding to any two values among t = 1, 2, ... , 6.

Illustration 7 : Obtain the linear equation for trend for a time series with $n = 8$, $\Sigma y = 344$, $\Sigma ty = 1342$

Since $n = 8$, we take t = 1, 2, ... , 8 Hence, we get $\Sigma t = 1 + 2 + ... + 8 = 36$ and $\Sigma t^2 = 1^2 + 2^2 + ... + 8^2 = 1 + 4 + ... + 64 = 204$.

$$\bar{t} = \frac{\Sigma t}{n} = \frac{36}{8} = 4.5, \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{344}{8} = 43$$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{8 \times 1342 - 36 \times 344}{8 \times 204 - (36)^2}$$

$$= \frac{10736 - 12384}{1632 - 1296}$$

$$= \frac{-1648}{336}$$

$$= -4.9048$$

$$\simeq -4.9$$

$$a = \bar{y} - b\bar{t}$$

$$= 43 - (-4.9) \times 4.5$$

$$= 43 + 22.05$$

$$= 65.05$$

Equation for trend $\hat{y} = a + bt$

$$= 65.05 + (-4.9)\,t$$

$$= 65.05 - 4.9\,t$$

EXERCISE 4.2

1. The information about death rate of a state in different years is given in the following table. Fit a linear equation to find trend and hence estimate the death rate for the year 2017.

Year	2009	2010	2011	2012	2013	2014	2015
Death rate	7.6	6.9	7.1	7.3	7.2	6.9	6.9

2. The data about Cost Inflation Index (CII) declared by the central government are as follows. The year 1981-82 is the base for this index. Find the estimate of this index for the year 2015-16 by fitting the linear equation to these data.

Year	2007 – 08	2008 – 09	2009 – 10	2010 – 11	2011 – 12	2012 – 13	2013 – 14	2014 – 15
CII	551	582	632	711	785	852	939	1024

3. The number of two wheelers registered (in thousand) in a city in different years is as follows. Use the method of fitting linear equation to these data to obtain the estimates for the number of vehicles registered in the year 2016 and 2017. Also find the trend values for each year.

Year	2010	2011	2012	2013	2014	2015
No. of vehicles (thousand)	69	75	82	91	101	115

4. The average age of women (in years) at the time of marriage obtained from the data of different census surveys in India are given in the following table. Fit an equation for a linear trend from the data and show it on a graph. Find the estimate for the value of the given variable for the year 2021 using the linear equation.

Year of census survey	1971	1981	1991	2001	2011
Average age of women at marriage (years)	17.7	18.7	19.3	20.2	22.2

*

4.3.3 Method of Moving Averages

The method of moving averages is very useful to find trend by eliminating the effect of short-term variations. The short-term variations are usually regular and have repetitions. The period of repetition of these variations can be found by past experience or other techniques and the average is found for the number of observations corresponding to this period. Since the average value lies in the center, we get the values that are free from the short-term fluctuations which show the trend.

Suppose, the values of the dependent variable in the given time series are $y_1, y_2, ..., y_n$ for time t = 1, 2, ... , n respectively and the interval for short-term cyclical fluctuations is 3 years. The mean of first three observations y_1, y_2, y_3 is found as $\frac{y_1 + y_2 + y_3}{3}$ and it is written against the center of these three values which is y_2. Further, the mean of successive three values y_2, y_3, y_4 is found as $\frac{y_2 + y_3 + y_4}{3}$ and it is written against y_3 which is the center of these three values. Similarly, all the means are calculated till the last value from the given values of the variable is included. The averages thus calculated are called three yearly moving averages which indicate the trend.

It is not necessary that the unit of time in every time series is year and time interval for repetitions in the pattern of the values of the variable may not be necessarily three years. The moving averages will be denoted according to the unit of time. For example, 5 days moving averages, three monthly moving averages, 4 weekly moving averages, etc. The unit for time is taken as 'year' for the discussion here.

While calculating for the given data, we first find the total of values of the variable in accordance with the time interval for the average. After finding the first total $y_1 + y_2 + y_3$, the next total namely $y_2 + y_3 + y_4$ is found by subtracting y_1 from the above total and then adding y_4 to it. All the successive totals are found in this manner and each total is divided by 3 to obtain three yearly moving averages.

Note : The first three yearly moving average is written against y_2 and thus the moving average against y_1 i.e. the trend value at that time cannot be obtained. Similarly, the trend value corresponding to y_n cannot be obtained.

Illustration 8 : The number of accounts opened in different weeks in a branch of a certain bank are given below. Find the trend using three-weekly moving averages.

Week	1	2	3	4	5	6	7	8	9	10
No. of accounts opened	26	27	26	25	22	24	25	23	22	21

Calculation for three weekly moving averages

Week t	No. of accounts opened y	Three weekly moving total	Three weekly moving average
1	26	—	—
2	27	26 + 27 + 26 = 79	$\frac{79}{3}$ = 26.33
3	26	79 – 26 + 25 = 78	$\frac{78}{3}$ = 26
4	25	78 – 27 + 22 = 73	$\frac{73}{3}$ = 24.33
5	22	73 – 26 + 24 = 71	$\frac{71}{3}$ = 23.67
6	24	71 – 25 + 25 = 71	$\frac{71}{3}$ = 23.67
7	25	71 – 22 + 23 = 72	$\frac{72}{3}$ = 24
8	23	72 – 24 + 22 = 70	$\frac{70}{3}$ = 23.33
9	22	70 – 25 + 21 = 66	$\frac{66}{3}$ = 22
10	21	—	—

Illustration 9 : Find the trend using five yearly moving averages for the following data about yearly production (in tons) of a factory.

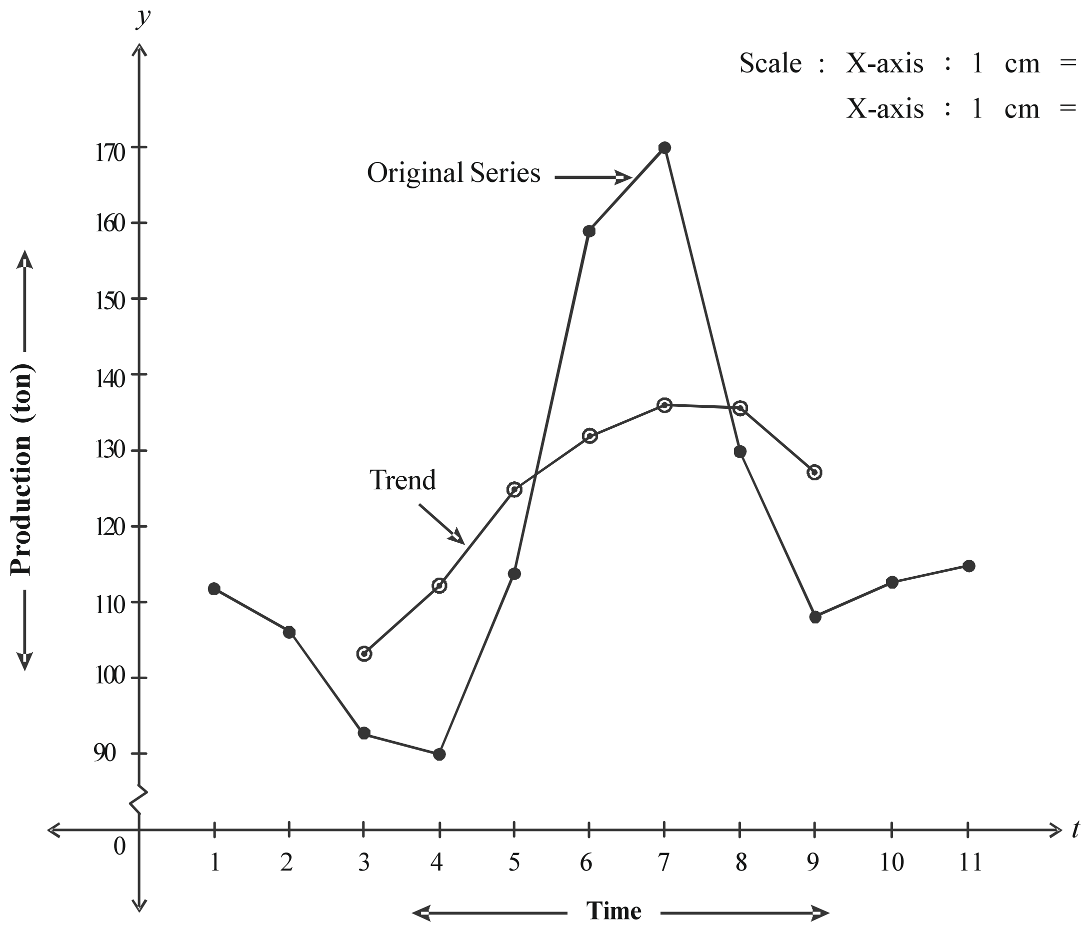
Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Production (tons)	112	106	93	90	114	159	170	130	108	113	115

Calculation for five yearly moving averages

Year	Production y	t	Five yearly moving total	Five yearly moving average (trend)
2006	112	1	—	—
2007	106	2	—	—
2008	93	3	$112+106+93+90+114=515$	$\frac{515}{5}=103$
2009	90	4	$515-112+159=562$	$\frac{562}{5}=112.4$
2010	114	5	$562-106+170=626$	$\frac{626}{5}=125.2$
2011	159	6	$626-93+130=663$	$\frac{663}{5}=132.6$
2012	170	7	$663-90+108=681$	$\frac{681}{5}=136.2$
2013	130	8	$681-114+113=680$	$\frac{680}{5}=136$
2014	108	9	$680-159+115=636$	$\frac{636}{5}=127.2$
2015	113	10	—	—
2016	115	11	—	—

Additional information for understanding

We shall show the values of the variable and the trend obtained by five yearly moving averages to understand the trend found by this method.



The curve passing through all the averages shows the trend.

If the interval of time for the moving averages is odd number like 3, 5, 7, ... then the trend is found as shown earlier. But the calculation of moving averages becomes difficult if this interval is an even number.

Suppose four yearly moving averages are to be found. First four yearly average will be found as $\frac{y_1 + y_2 + y_3 + y_4}{4}$. As the center for these four values is between y_2 and y_3, this average will be written at that position. Similarly, the successive averages namely $\frac{y_2 + y_3 + y_4 + y_5}{4}$, $\frac{y_3 + y_4 + y_5 + y_6}{4}$, ... will be found and written between y_3 and y_4, between y_4 and y_5, ... respectively. Since these averages are in between two years, the average of each pair of averages is found and it is written between two moving averages. Thus, the average value of the first two averages shown above will be written against y_3. The averages thus obtained are called as four yearly moving averages. The processes of finding an average is to be done twice here. To simplify these calculations, four yearly totals are obtained first and then totals of pairs of years are found. As these totals involve 8 values, each total is divided by 8 which gives the four yearly averages mentioned above.

Whenever the time of the cycles of short-term variations is an even number, moving averages are obtained by first finding the moving totals and then the pairwise totals as shown in the above method.

Illustration 10 : Find the trend using four monthly moving averages for the following data showing monthly sales (in lakh ₹) of a shop.

Month	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
Sales (lakh ₹)	5	3	7	6	4	8	9	10	8	9

Calculation of four monthly moving averages

Month	**Sales (lakh ₹) y**	t	**Four monthly moving total**	**Pairwise total**	**Four monthly moving average**
March	5	1		—	—
			—		
April	3	2		—	—
			$5+3+7+6=21$		
May	7	3		$21+20=41$	$\frac{41}{8}=5.13$
			$21-5+4=20$		
June	6	4		$20+25=45$	$\frac{45}{8}=5.63$
			$20-3+8=25$		
July	4	5		$25+27=52$	$\frac{52}{8}=6.5$
			$25-7+9=27$		
August	8	6		$27+31=58$	$\frac{58}{8}=7.25$
			$27-6+10=31$		
September	9	7		$31+35=66$	$\frac{66}{8}=8.25$
			$31-4+8=35$		
October	10	8		$35+36=71$	$\frac{71}{8}=8.88$
			$35-8+9=36$		
November	8	9		—	—
			—		
December	9	10		—	—

The trend of the time series is shown by the four monthly moving averages.

Merits and limitations of the method of moving averages are as follows :

Merits :

(1) The effect of short-term component is eliminated to a large extent using the averages and the trend of the series is obtained.

(2) The calculation is easy to understand as it is comparatively less and simple.

Limitations :

(1) The trend obtained by this method is not accurate if the interval for the moving averages is not chosen correctly.

(2) The estimates of trend for some initial and last time periods cannot be obtained.

(3) A specific mathematical formula is not obtained for future estimates.

EXERCISE 4.3

1. Find the trend by three yearly moving averages from the following data about the sales (in ten lakh ₹) of a company.

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Sales (ten lakh) ₹	3	4	8	6	7	11	9	10	14	12

2. The average monthly closing prices of shares of a company in the year 2016 are given in the following table. Find the trend using four monthly moving averages.

Month	January	February	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
Share price (₹)	253	231	350	261	262	266	263	261	281	278	278	272

3. Find the trend using five yearly moving averages from the following data of profit (in lakh ₹) of a trader in different years.

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015
Profit (lakh ₹)	15	14	18	20	17	24	27	25	23

4. The wholesale price index numbers for different quarters (Q) of a year are obtained as follows. Find the trend by four quarterly moving averages.

Year	2013				2014				2015			
Quarter	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
Index No.	110	110	125	135	145	152	155	168	131	124	132	153

*

Summary

- The data collected and arranged according to time is called Time Series.
- It is necessary to analyse the time series to find the future estimates of the given variable.
- There are four main components affecting the values of the variable in a time series :

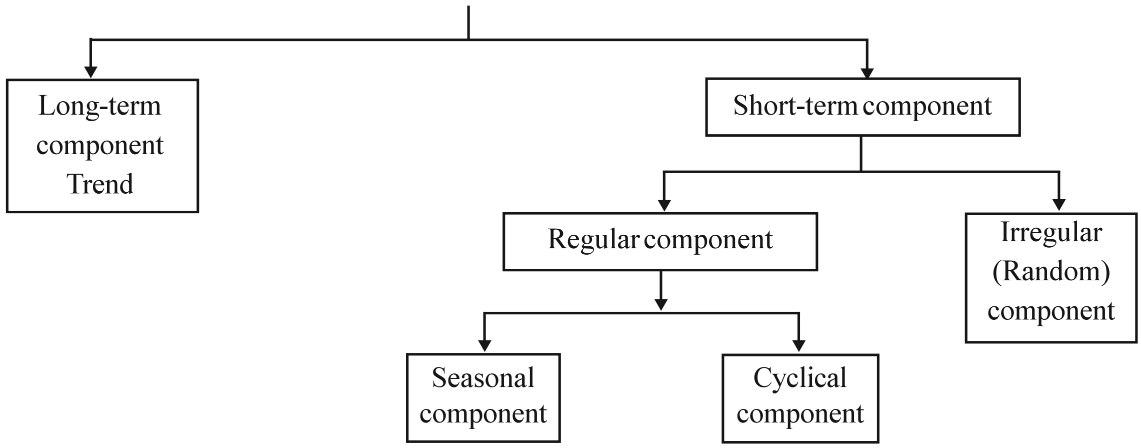
  (1) Long-term Component (Trend) (2) Seasonal Component

  (3) Cyclical Component (4) Random (Irregular) Component
- Short-term fluctuations are found in the time series due to seasonal component, cyclical component and random component.
- Seasonal and cyclical fluctuations repeat almost regularly.
- Three methods of measuring trend :

  (1) Graphical Method (2) Method of least squares (3) Method of moving average

Chapter at a glance

Time Series



List of formulae :

For fitting a linear equation $\hat{y} = a + bt$ to the given data

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}, \qquad a = \bar{y} - b\bar{t}$$

Exercise 4

Section A

Find the correct option for the following multiple choice questions :

1. Which type of variations are produced in the time series variable due to seasonal component ?

(a) Long-term (b) Irregular (c) Regular (d) Zero

2. Which variation is shown in 'decrease in the production of a company' due to strike ?

(a) Random (b) Trend (c) Seasonal (d) Cyclical

3. Name the method for fitting the linear equation to find linear trend.

(a) Graphical Method (b) Method of least squares

(c) Method of moving average (d) Method of partial average

4. How do you show the additive model of the time series ?

(a) $y_t = T_t + S_t + C_t - R_t$ (b) $y_t = T_t + S_t + C_t + R_t$

(c) $y_t = T_t \times S_t + C_t \times R_t$ (d) $y_t = S_t + C_t + R_t$

5. State the independent variable of time series.

(a) y_t (b) S_t (c) t (d) x_t

6. Which component of the time series is impossible to predict ?

(a) Random component (b) Trend (c) Seasonal component (d) Cyclical component

7. Which of the following variations are due to cyclical component ?

(a) Rise in demand during winter

(b) Decrease in the share prices due to recession in share market

(c) Decrease in the agricultural produce due to excessive rains

(d) Continuously decreasing death rate

8. The trend equation obtained from a time series from January 2016 to December 2016 is $\hat{y} = 30.1 + 1.5\,t$. Find the value of trend for April 2016.

(a) 30.1 (b) 34.6 (c) 36.1 (d) 33.1

9. Which of the following fluctuations is the effect of seasonal component ?

(a) Increase in the migration to cities from rural areas

(b) Increasing number of vehicles on roads in a city

(c) Increase in the number of tourists during school vacation

(d) Increased death rate during a certain epidemic

10. Which method of finding trend is best to eliminate the effect of repetitive short-term variations ?

(a) Graphical Method (b) Method of least squares

(c) Karl Pearson's method (d) Method of moving average

Section B

Answer the following questions in one sentence :

1. Give an example of time series having decreasing trend.
2. What is a time series ?
3. Which of the components of time series produce short-term variations ?
4. What is meant by analysis of time series ?
5. What is the notation to show the cyclical component of the time series ?
6. State the names of methods of measuring trend.
7. The effect of which component indicates fluctuations repeating within one year ?
8. State the components of time series.
9. When is the method of moving average more useful to find trend ?
10. The linear equation fitted using the data of 7 weeks for a variable y is $\hat{y} = 25.1 - 1.3\,t$. Estimate the value of y for the eighth week.

Section C

Answer the following questions :

1. Describe the additive model of time series.
2. What is meant by cyclical component ?
3. How does seasonal component differ from the cyclical component ?
4. Explain the irregular component.
5. State the limitations of graphical method.
6. Explain the meaning of moving average.
7. Define time series.
8. State the merits of the method of moving average to measure trend.
9. Describe the graphical method to measure trend.

Section D

Answer the following questions :

1. Explain the importance of time series.
2. State the uses of analysis of time series.
3. What is meant by trend of a time series ? Explain with an illustration.
4. Write a short note on seasonal component.
5. Explain the method of fitting a linear equation to the given data using the method of least squares.
6. State the merits and limitations of the method of least squares.
7. Describe the method of moving average to find trend.
8. Discuss the limitations of the method of moving average.
9. The following time series shows the daily production of a factory. Find the trend using graphical method.

Day	1	2	3	4	5	6	7	8	9	10
Production (units)	21	22	23	25	24	22	25	26	27	26

10. Fit a linear equation from the following data for variable (y) of a time series.
    $n = 4, \quad \Sigma y = 270, \quad \Sigma ty = 734$
11. The data collected about the demand of a commodity from a store are as follows. Find the trend using three monthly moving averages.

Month	January	February	March	April	May	June	July
Demand (units)	15	16	18	18	23	23	20

Section E

Solve the following :

1. The data about exports (in crore ₹) of ready-made garments of a textile manufacturer are shown below :

Year	2010	2011	2012	2013	2014	2015
Export (crore ₹)	22	25	23	26	20	25

Fit a linear trend to these data and estimate the trend for the export in the year 2017.

2. The following data are available for the number of passengers who travelled in the last 5 years by the aircrafts of an airline company. Estimate the trend for the year 2016 by fitting linear trend.

Year	2011	2012	2013	2014	2015
No. of passengers (thousands)	45	47	44	40	38

3. The data about closing prices of shares of a company registered in a stock exchange for different months is given in the following table. Find the trend using three monthly moving averages.

Month	2015 April	May	June	July	August	Sept.	Oct.	Nov.	Dec.	2016 January
Share price (₹)	76	73	65	68	67	60	63	67	65	66

4. The following data show the sales (in thousand ₹) of a commodity. Find the trend by graphical method.

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Sales (thousand ₹)	200	216	228	235	230	232	236	235	230	233

5. The quantity index numbers of consumption of edible oil in a state are given in the following table. Find the trend using five yearly moving averages.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Index No.	115	121	119	120	117	119	120	118	116	124	125

Section F

Solve the following:

1. Find a linear equation using the method of least squares for the trend of production from the following data about sugar production of a country recorded for the last 6 years. Find the trend estimates for the production of the year 2016-17 and 2017-18.

Year	2009 – 10	2010 – 11	2011 – 12	2012 – 13	2013 – 14	2014 – 15
Sugar production (crore tons)	29.2	34.2	35.4	36.4	33.6	37.7

2. The number of students studying in a college are shown in the following table. Find the trend by four yearly moving averages.

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
No. of students	332	317	357	392	402	405	410	427	405	438

3. The birth rates of a state in different years are given in the following table. Fit a linear trend for these data. Also find the estimates for birth rates in the year 2016 and 2017.

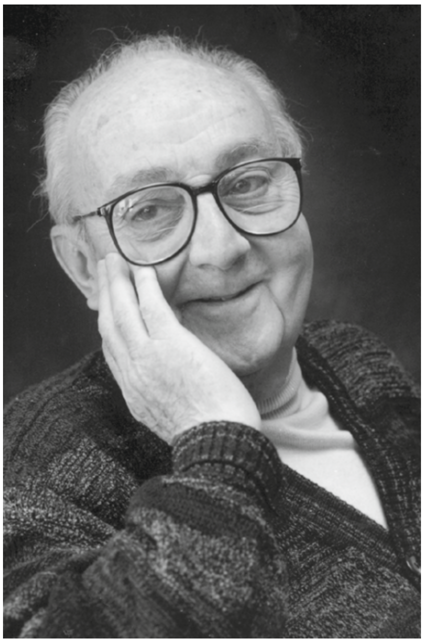
Year	2009	2010	2011	2012	2013	2014	2015
Birth rate	22.2	21.8	21.3	20.9	20.6	20.2	19.9

4. The data about goods transported in different years by a division of railways are given below. Find the estimates for each year by fitting a linear equation and represent it by a graph. Also find the estimate for the year 2016.

Year	2011	2012	2013	2014	2015
Goods transported (tons)	180	192	195	204	202

5. The data of weekly prices (in USD per barrel) of crude oil are given in the following table. Find the trend using four weekly moving averages.

Month	March 2016				April 2016				May 2016			
Week	1	2	3	4	1	2	3	4	1	2	3	4
Price of Crude oil	35.92	38.50	39.44	39.46	36.79	39.72	40.36	43.73	45.92	44.66	46.21	48.45



George Edward Pelham Box
(1919 -2013)

George E. P. Box worked in the areas of quality control, time series analysis, design of experiments and Bayesian inference. He has been called "one of the greatest statistical minds of the 20th century." He has been associated with University at Raleigh (now North Carolina State University), Princeton University, University of Wisconsin–Mandison. Box has published numerous articles and papers and he is an author of many books. He is a recipient of prestigious honours, medals and was the president of American Statistical Association in 1978 and of the Institute of Mathematical Statistics in 1979. His name is associated with results in statistics such as Box–Jenkins models, Box–Cox transformations, Box–Behnken designs, and others. Box was elected a member of the American Academy of Arts and Sciences in 1974 and a Fellow of the Royal Society (FRS) in 1985.

●

Answers

Exercise 1.1

1. (1) Fixed base index numbers : 100, 103.27, 105.09, 106.55, 108, 113.82, 119.27, 125.45

(2) Chain base index numbers : 100, 103.27, 101.76, 101.38, 101.37, 105.39, 104.79, 105.18

(3) Index numbers using average wage : 91.36, 94.35, 96.01, 97.34, 98.67, 103.99, 108.97, 114.62

2. (1) Fixed base index numbers : 100, 101.79, 105.36, 107.14, 110.71, 114.29,121.43, 128.57

(2) Chain base index numbers : 100, 101.79, 103.51, 101.69, 103.33, 103.23, 106.25, 105.88

(3) Index numbers using average price : 96.55, 98.28, 101.72, 103.45, 106.90, 110.34, 117.24, 124.14

3. (1) Fixed base index numbers : 100, 108.70, 112.78, 115.19, 119.44

(2) Chain base general price index numbers : 100, 108.70, 103.65,102.26, 103.71

4. General index number of *n* items : 126.45; Overall increase in the price of fuel items is 26.45 %

Exercise 1.2

1. Fixed base index numbers : 100, 110, 104.5, 112.86, 135.43, 143.56, 157.92
2. Chain base index numbers : 117.4, 100.51, 102.80, 103.13, 102.64, 102.49, 102.28
3. Chain base index numbers : 100, 99.63, 99.26, 100, 103.73, 101.80, 100, 103.53, 100, 102.05
4. Fixed base index numbers : 110, 123.2, 134.29, 145.03, 152.28, 169.03

Exercise 1.3

1. $I = 307$, prices have increased by 207 %.
2. $I = 123.80$, prices have increased by 23.80 %.
3. $I_L = 126.72,\ I_P = 126.85,\ I_F = 126.78$
4. $I_L = 141.13,\ I_P = 140.15,\ I_F = 140.64$ **5.** $I_F = 142.57$ **6.** $I_P = 115.2,\ I_F = 115.14$

Exercise 1.4

1. Index number by family budget method = 135.64 and total expenditure has increased by 35.64 %. Average monthly disposable income = ₹ 20,346.
2. Index number $I = 128.53$ and rise in total expenditure is 28.53 %.
3. Index number $I = 132.51$ and rise in total expenditure is 32.51 %.
4. Index number $I = 213.20$ and rise in total expenditure is 113.20 %.
5. Index number by family budget method = 129.64 and by total expenditure method $I = 129.64$ Thus, both index numbers are same.

Exercise 1

Section A

1.	(c)	**2.**	(a)	**3.**	(d)	**4.**	(c)	**5.**	(d)
6.	(d)	**7.**	(c)	**8.**	(c)	**9.**	(c)	**10.**	(c)
11.	(a)	**12.**	(c)						

Section B

12. The statement is false. Price index number of oil is 500.

Section C

7. Real wage ₹ 16,392.85 and loss to worker ₹ 1642.85 (Decrease in purchasing power)
8. Real wages ₹ 29166.67, 26666.67, 32307.69, 31250
9. Rate of inflation for year 2015 : 2.03 %
10. 449.55
11. Average monthly disposable income = ₹ 30,000
12. Index number of income = 125 **13.** Index number of production = 280 **14.** $I_p = 222.5$

Section D

7. 161.87

8. Fixed base index numbers = 100, 111.11, 133.33, 144.44, 166.67, 222.22, 263.89

9. Chain base index numbers = 100, 104, 100.96, 102.86, 100.93, 116.51

10. Fixed base index numbers = 120, 108, 151.20, 189

11. Chain base index numbers = 100, 112.5, 106.67, 114.58, 109.09, 116.67

12. Index number = 226.6

13. $I_L = 166.67$, $I_p = 150$, $I_F = 158.12$

14. $I_p = 167.71$

Section E

1. General index number = 122.32

2. Index number by total expenditure method = 149.41

3. Index number by total expenditure method = 115.69

4. Fixed base index numbers = 100, 118.75, 125, 131.25, 140.63, 187.5, 203.13;

   Index numbers using average price = 91.43, 108.57, 114.29, 120, 128.57, 171.43, 185.71

5. Index number of industrial production I = 379.19

6. Index number I = 126.79 and rise in price is 26.79 %.

7. Real wages = 12,500, 10,000, 9268.29, 9090.91, 9361.7, 9615.38 Purchasing power of money = ₹ 0.38

Section F

1. $I_L = 113.65$, $I_P = 113.94$, $I_F = 113.79$ and rise in price is 13.79 %.

2. $I_P = 191.53$, $I_F = 211.52$

3. $I_F = 84.84$

4. $I_L = 109.52$, $I_P = 110.29$, $I_F = 109.90$

5. Index number by family budget method = 118.58 and index number by total expenditure method = 118.58. Thus, both index numbers are same.

6. Index number for year 2014 $I_1 = 239.41$ and index number for year 2015 $I_2 = 253.44$. The rise in cost of living in the current year is 14.03 %. The percentage rise in the price index number is 5.86 % and rise in wage is 5 %. Hence, wage rise is 0.86 % less.

7. Index number $I = 231.44$ Income should be ₹ 13,886.40 to maintain earlier standard of living.

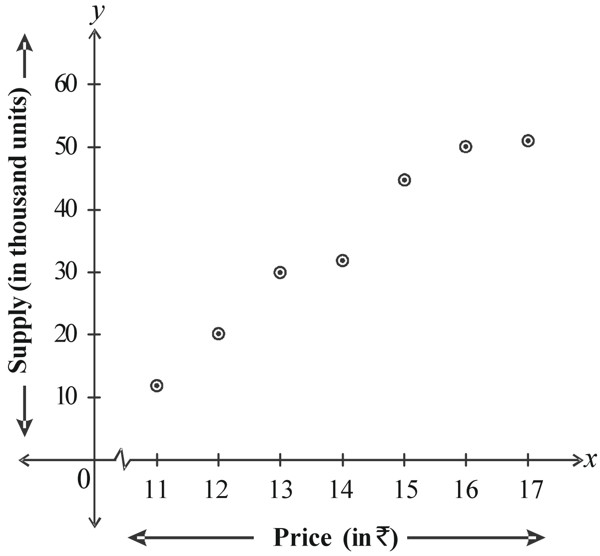
8. Index number of industrial production $= 100.10$, which indicates a rise of 0.10 % with respect to the base year.

9. Index number $I = 128.75$.

10. Cost of living index number $= 196.35$ and the rise is $(196.35 - 100) = 96.35$ % as compared to the base year.

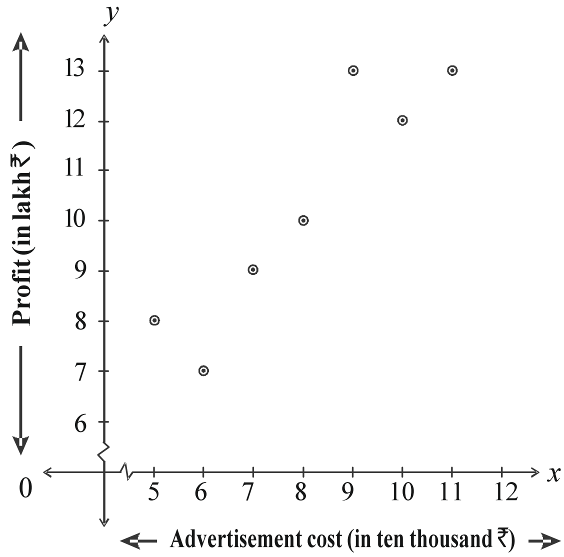
Exercise 2.1

1.



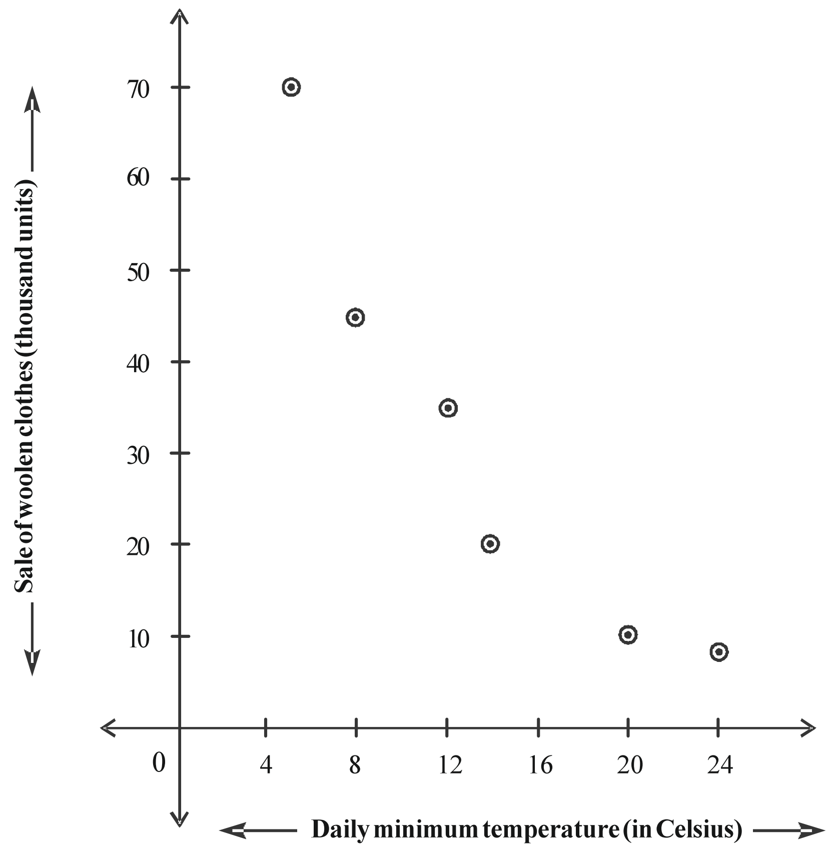
There is partial positive corelation between price and supply

2.



There is partial positive correlation between advertisement cost and profit

3.



There is partial negative correlation between daily minimum temperature and sale of woolen clothes

Exercise 2.2

1. $r = 0.81$ **2.** $r = -0.90$ **3.** $r = 0.90$ **4.** $r = 0.24$ **5.** $r = 0.82$

6. $r = -0.96$ **7.** $r = 0.67$ **8.** $r = -0.92$ **9.** $r = 0.99$ **10.** $r = 0.80$

11. $r = 0.84$ **12.** $r = 0.5$ **13.** $r = 0.8$ **14.** (1) $r = 0.94$ (2) $r = 0.96$

15. $r = -0.55$

Exercise 2.3

1. $r = 0.49$ **2.** $r = 0.78$ **3.** $r = 0.7$ **4.** $r = 0.82$ **5.** $r = 0.91$

6. $r = 0.90$ **7.** $r = -0.30$ **8.** Corrected $\Sigma d^2 = 82.5$, $r = 0.26$

Exercise 2

Section A

1. (c) **2.** (a) **3.** (d) **4.** (b) **5.** (c)

6. (d) **7.** (b) **8.** (b) **9.** (b) **10.** (c)

11. (a) **12.** (b) **13.** (c) **14.** (c) **15.** (a)

16. (b) **17.** (b) **18.** (a)

Section B

3. Positive **4.** Positive **5.** Negative **6.** Negative **7.** Nonsense correlation

8. r remain unchanged due to change of origion, so $r = 0.4$ **10.** $r = 0$ **11.** Negative

Section C

11. $r = 0.67$ **12.** $r = -0.54$ **13.** $r = 0.27$

Section D

10. $r = 0.75$ **11.** $r = 0$ **12.** $r = -0.5$ **13.** $r = 0.2$

Section E

1. $r = -0.81$ **2.** $r = 0.43$ **3.** $r = 0.79$ **4.** $r = 0.77$ **5.** $r = 0.54$

6. $r = 0.13$

Section F

1. r = 0.99
2. r = –0.96
3. r = 0.88
4. r = 0.81
5. r = 0.38
6. r = 0.79
7. r = 0
8. r = 0.6
9. r = 0.3
10. r = 0.79
11. Corrected $\Sigma d^2 = 78$; r = 0.53
12. r = 0.73

●

Exercise 3.1

1. $\hat{y} = 31.44 - 1.34\,x$ and for price x = 20 ₹, estimate of demand $\hat{y} = 4.64$ (hundred units)

2. $\hat{y} = 3.35 + 1.93\,x$ and for usage time of car x = 5 year, Estimate of annual maintenance cost $\hat{y} = 13$ (thousand ₹)

   $\therefore$ Error $e = y - \hat{y} = 13 - 13 = 0$ (Here for x = 5, the observed value of y given in the table is 13)

3. $\hat{y} = 64.27 + 0.83\,x$ and for average rain x = 35 cm, estimate of yield of crop $\hat{y} = 93.32$ (ton)

4. $\hat{y} = 69.7 + 1.13\,x$ and for experience of worker, x = 7 year, estimate of performance index $\hat{y} = 77.61$

Exercise 3.2

1. $\hat{y} = 54.84 + 2.52\,x$ and for 300 kg usage of fertilizer [$\therefore$ x = 30 (ten kg.)], estimate of crop of cotton $\hat{y} = 130.44$ (Quintal per Hectare)

2. $\hat{y} = 52.84 + 0.68\,x$ and for a father's height x = 170 cm, estimate of height of the son $\hat{y} = 168.44$ cm

3. $\hat{y} = 20.72 - 0.71\,x$ and for altitude x = 7 thousand feat, estimate of effective Oxygen $\hat{y} = 15.75$ %

4. $\hat{y} = -3495.7 + 327.73\,x$ and for carpet area x = 110 sq. meter estimated monthly rent $\hat{y} = 32554.6$ ₹

5. $\hat{y} = 0.53 + 0.02\,x$ and for x = 80 customers, estimated sales $\hat{y} = 2.13$ (thousand ₹)

6. $\hat{y} = 7.6 + 0.29\,x$; x = Profit (lakh ₹) and y = Administrative cost (lakh ₹)

7. $\hat{y} = 53.72 + 1.54\,x$ and for rainfall x = 60 cm, estimate of yield of corn $\hat{y} = 146.12$ Quintal

8. $\hat{y} = 8.74 + 1.02\,x$ and for price x = 16 ₹, estimated supply $\hat{y} = 25.06$ (hundred units)

9. $\hat{y} = -4.8 + 0.15\,x$ and for maximum daily temperature x = 42 celcius, estimate of sale of icecream $\hat{y} = 1.5$ (lakh ₹)

Exercise 3

Section A

1. (b)	**2.** (a)	**3.** (c)	**4.** (d)	**5.** (a)
6. (a)	**7.** (c)	**8.** (c)	**9.** (d)	**10.** (b)
11. (c)	**12.** (c)	**13.** (c)	**14.** (c)	**15.** (b)

Section B

8. Error = 0

9. Both variables are multiplied by 2 there fore $c_x = \frac{1}{2}$ and $c_y = \frac{1}{2}$. ∴ Regression coefficient will not change

10. $b_{yx} = 0.5 \times \frac{4}{2} = 1$ **11.** $\hat{y} = 50$ **12.** $r = 1$ **13.** $r = -1$

Section C

2. Error $e = 1$ **3.** $a = 2$ and $\hat{y} = 2 + 0.6\,x$

4. $b_{yx} = 5$ So, it can be said that because of increase of 1 unit in *x*, there is appropriate 5 units of increase in *y*.

5. $s_y = 3$ **6.** $R^2 = 1$ **7.** $s_x = 5$ **8.** 5 Units

9. $b_{yx} = 1.2$ and $a = 13$ **10.** $b_{vu} = b_{yx} \times \frac{c_x}{c_y} = 0.75 \times \frac{\frac{1}{6}}{\frac{1}{2}} = 0.25$

Section D

8. $\hat{y} = 4 + 0.75\,x$ **9.** $\hat{y} = -10 + 2\,x$

10. $R^2 = 0.81$; 81 % variation of the total variation in *y*, can be explained by the regression model.

11. $b_{yx} = 2.52$ so it can be said that because of increase of 1 unit in *x*, there is approximate 2.52 units of increase in y.

12. (i) $b_{vu} = 0.8$ (ii) $b_{vu} = 1.6$ (iii) $b_{vu} = 0.08$ **13.** $\hat{y} = 12 + 0.88\,x$

Section E

1. $\hat{y} = 2 + 0.75\,x$ **2.** $\hat{y} = 38.8 + 0.67\,x$ **3.** $\hat{y} = 58 + 3.2\,x$

4. $\hat{y} = 764.8 + 11.4\,x$ and for $x = 20$ cm, estimate of yield of crop is $\hat{y} = 992.8$ kg.

5. $\hat{y} = 18 + 0.8\,x$ and for $x =$ ₹ 45 lakh, estimate of market price is $\hat{y} = 54$ (thousand ₹)

Section F

1. $\hat{y} = 73.29 - 1.59\,x$ and for price x = 40 ₹, estimate of demand $\hat{y} = 9.69$ (hundred units)

2. $\hat{y} = 73.43 + 0.9\,x$ and for price x = 17 years, the estimate of performance rating $\hat{y} = 88.73$

3. $\hat{y} = 34.8 + 0.74\,x$ and for daily income x = 500 ₹, estimate of expenditure $\hat{y} = 404.8$ ₹

4. $\hat{y} = 3.73 + 0.13\,x$ and for advertisement cost $x = 50$ (ten thousand ₹), estimate of sales $\hat{y} = 10.23$ crore ₹

5. $\hat{y} = -122.94 + 91.67\,x$ and $R^2 = 0.97$ $\therefore$ Regression model is reliable.

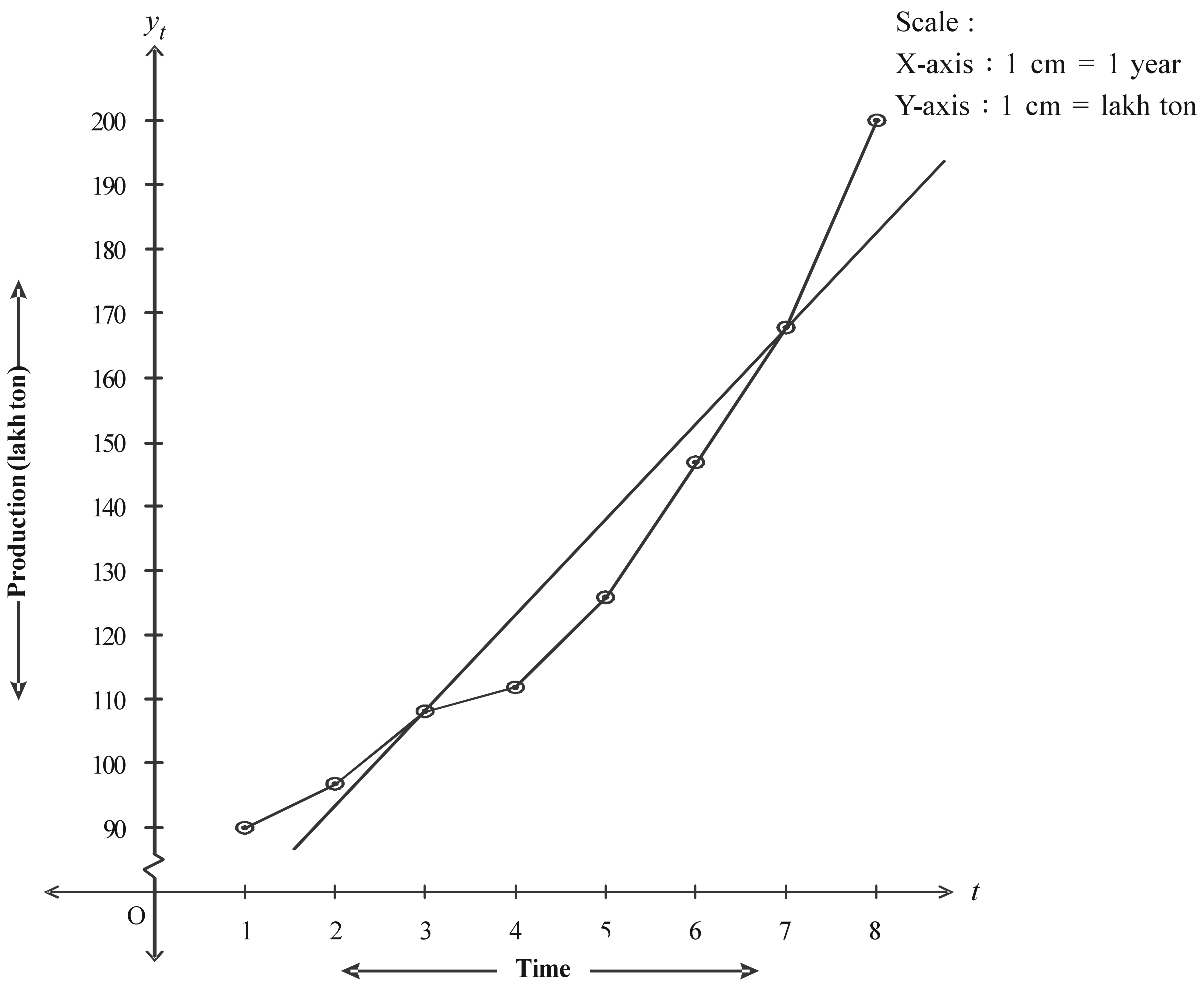
6. $\hat{y} = -10 + 1.6\,x$ and for x = 30 $\hat{y} = 38$

7. $\hat{y} = -0.44 + 0.7\,x$ and for x = 5, $\hat{y} = 3.06$

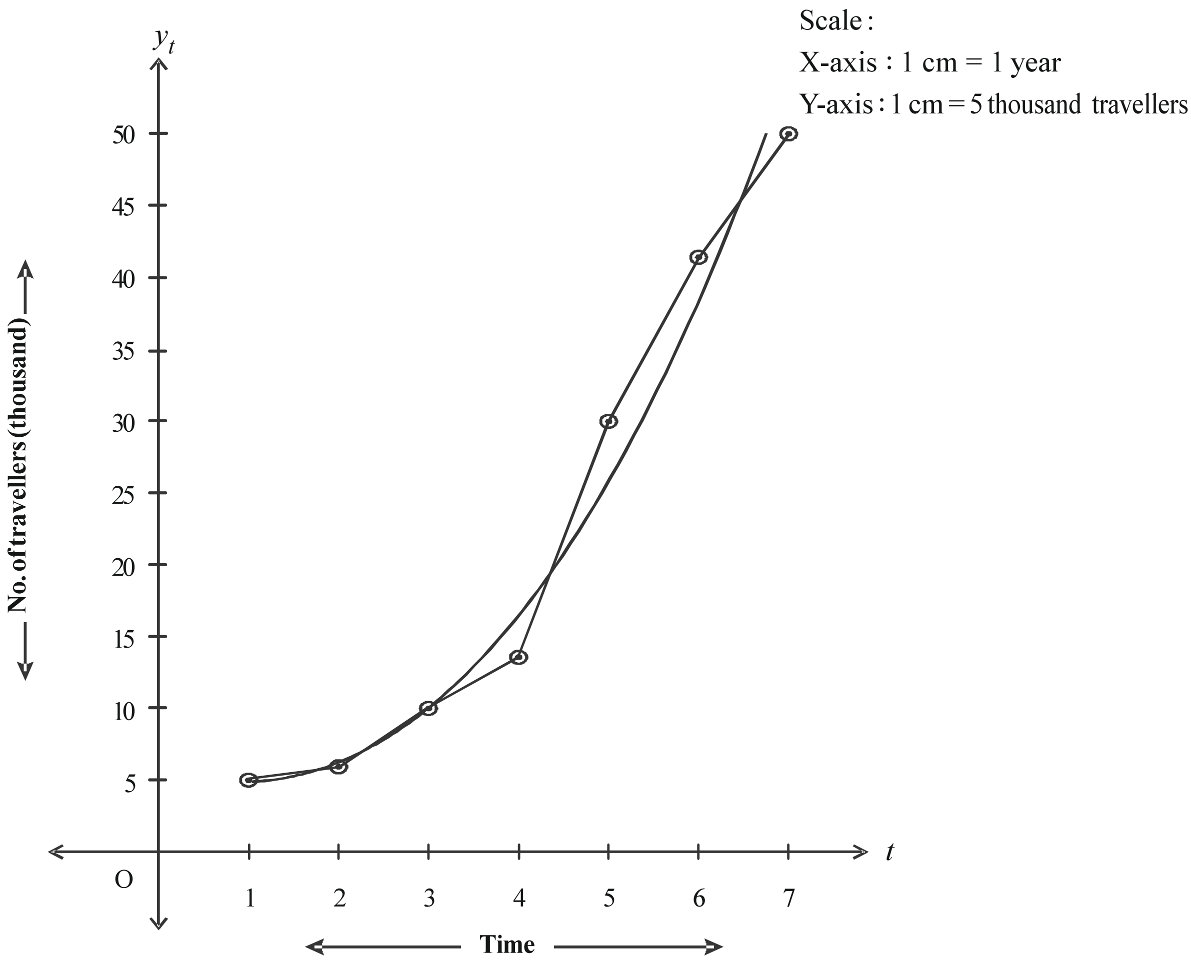
●

Exercise 4.1

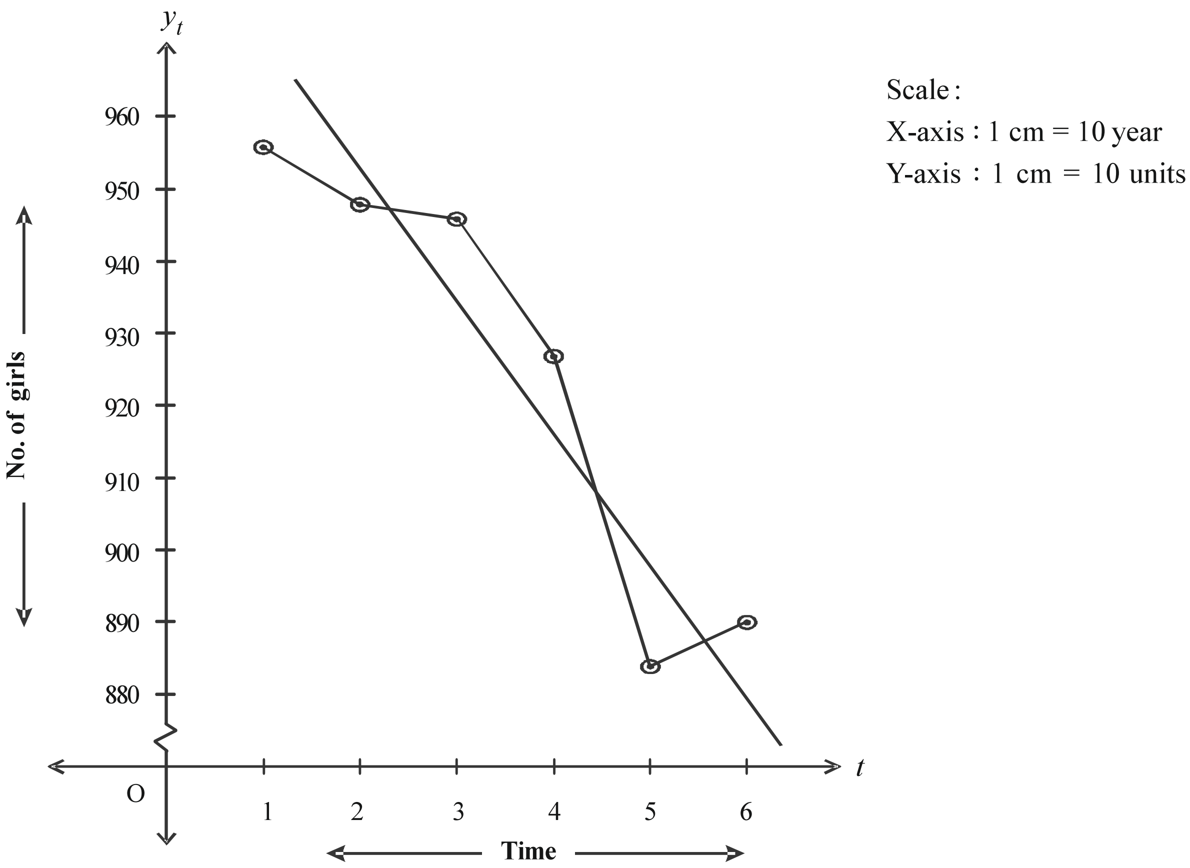
1.



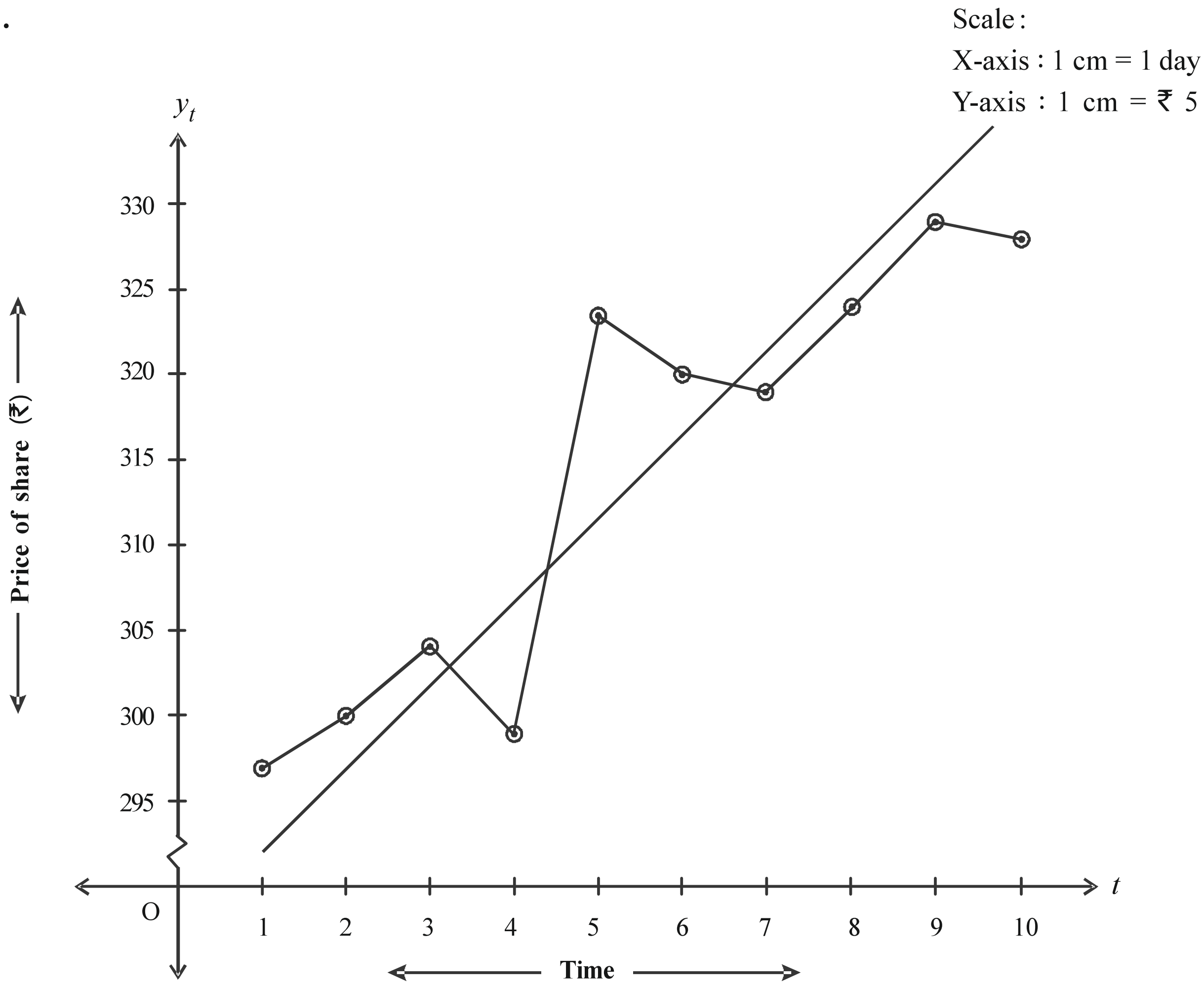
2.


Scale :
X-axis : 1 cm = 1 year
Y-axis : 1 cm = 5 thousand travellers
y_t
No. of travellers (thousand)
50
45
40
35
30
25
20
15
10
5
O
1
2
3
4
5
6
7
t
Time

3.


Scale :
X-axis : 1 cm = 10 year
Y-axis : 1 cm = 10 units
y_t
No. of girls
960
950
940
930
920
910
900
890
880
O
1
2
3
4
5
6
t
Time

4.



Exercise 4.2

1. $\hat{y} = 7.41 - 0.07\,t$, for year 2017 $\hat{y} = 6.78$

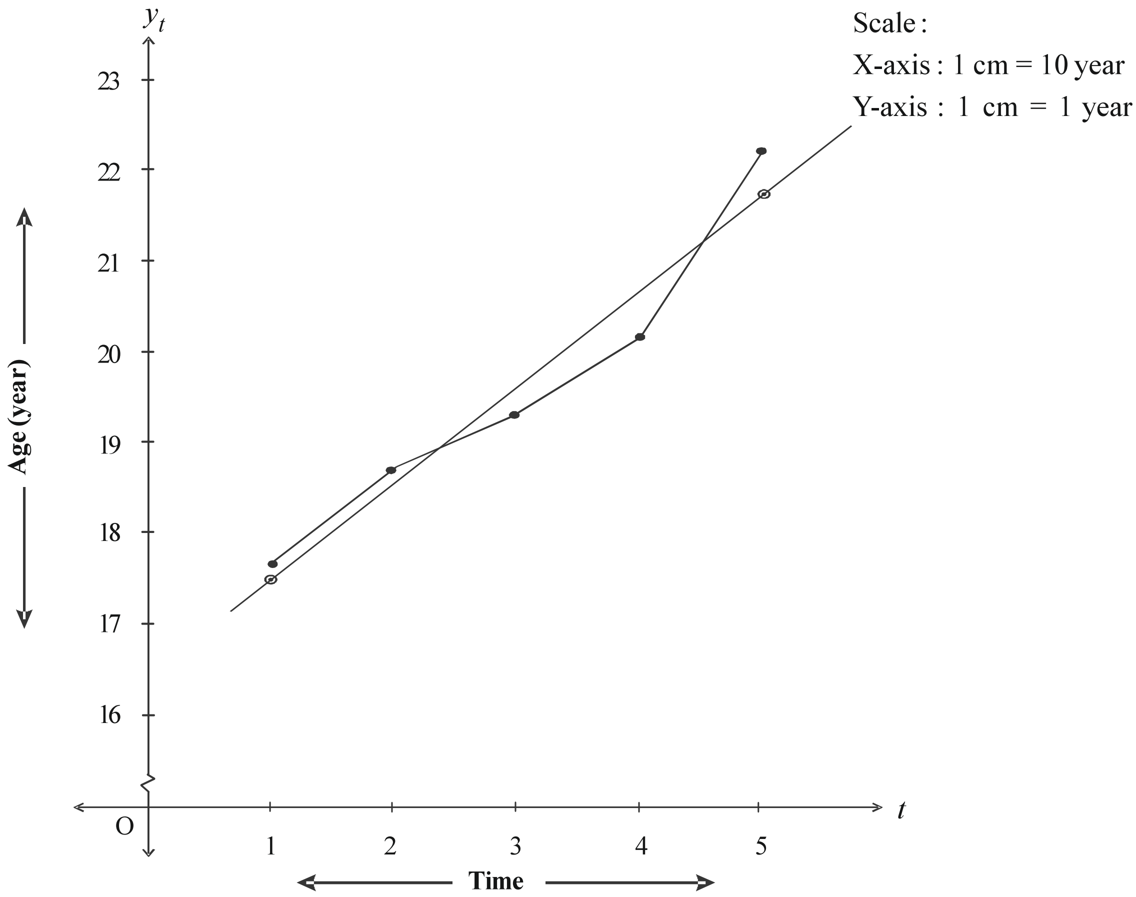
2. $\hat{y} = 447.2 + 69.4\,t$, for year 2015-16 $\hat{y} = 1071.8$

3. $\hat{y} = 57.12 + 9.06\,t$, $\hat{y} = 120.54$ thousand for year 2016

$\hat{y} = 129.6$ thousand for year 2017

Year	2010	2011	2012	2013	2014	2015
Estimated values of trend (thousand vehicles)	66.18	75.24	84.3	93.36	102.42	111.48

4. $\hat{y} = 16.47 + 1.05\,t$, $\hat{y} = 22.77$ years for year 2021



Exercise 4.3

1.

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Three yearly moving average	–	5	6	7	8	9	10	11	12	–

2.

Month	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
Four yearly moving average	–	–	274.88	280.38	273.88	263	265.38	269.25	272.63	275.88	–	–

3.

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015
Four yearly moving average	–	–	16.8	18.6	21.2	22.6	23.2	–	–

4.

Year	2013				2014				2015			
Quarter	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
Four Quarterly moving average	–	–	124.38	134	143	150.88	153.25	148	141.63	136.88	–	–

Exercise 4

Section A

1. (c) **2.** (a) **3.** (b) **4.** (b) **5.** (c)

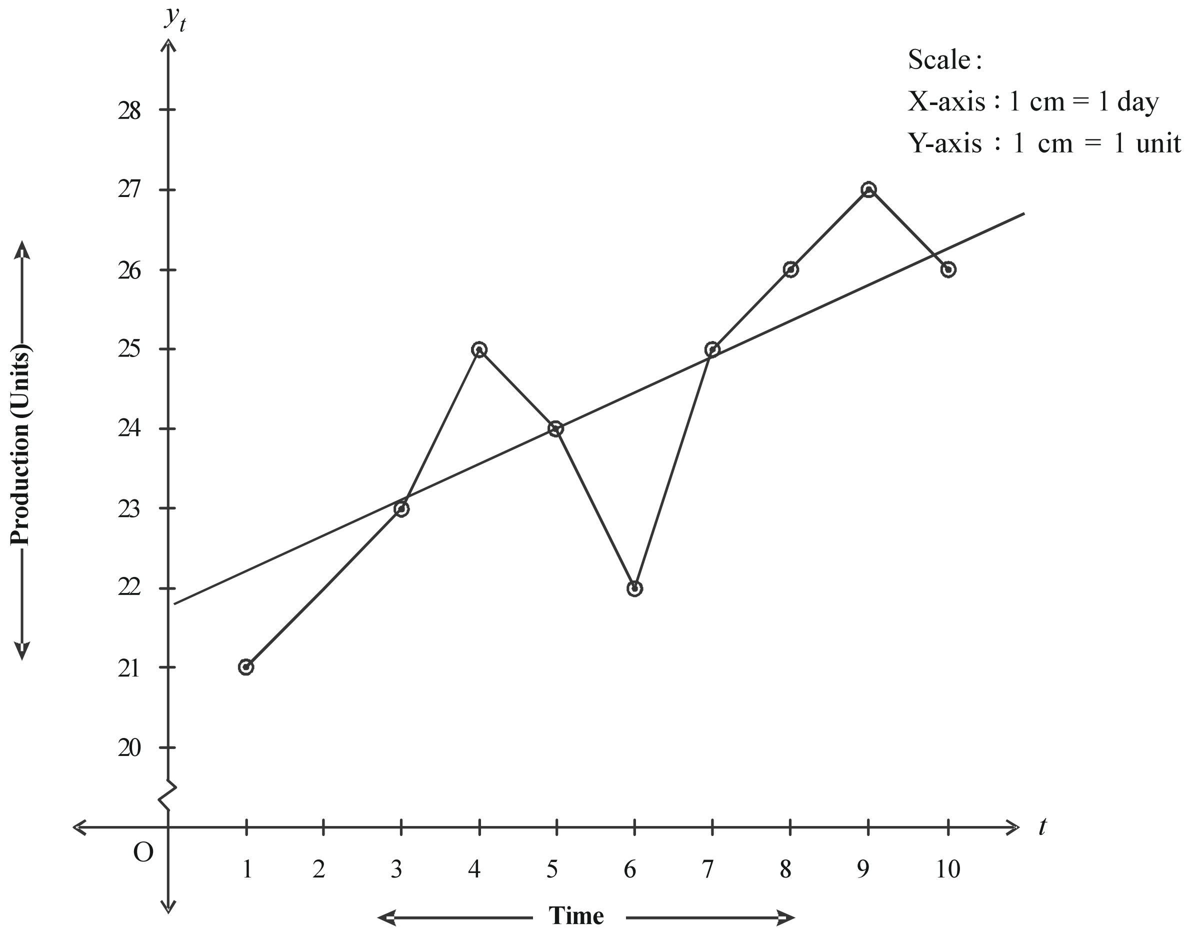
6. (a) **7.** (b) **8.** (c) **9.** (c) **10.** (d)

Section B

9. $\hat{y} = 14.7$ for eighth week

Section D

9.



10. $\hat{y} = 38 + 11.8\,t$

11.

Month	January	February	March	April	May	June	July
Three monthly moving average	–	16.33	17.33	19.67	21.33	22	–

Section E

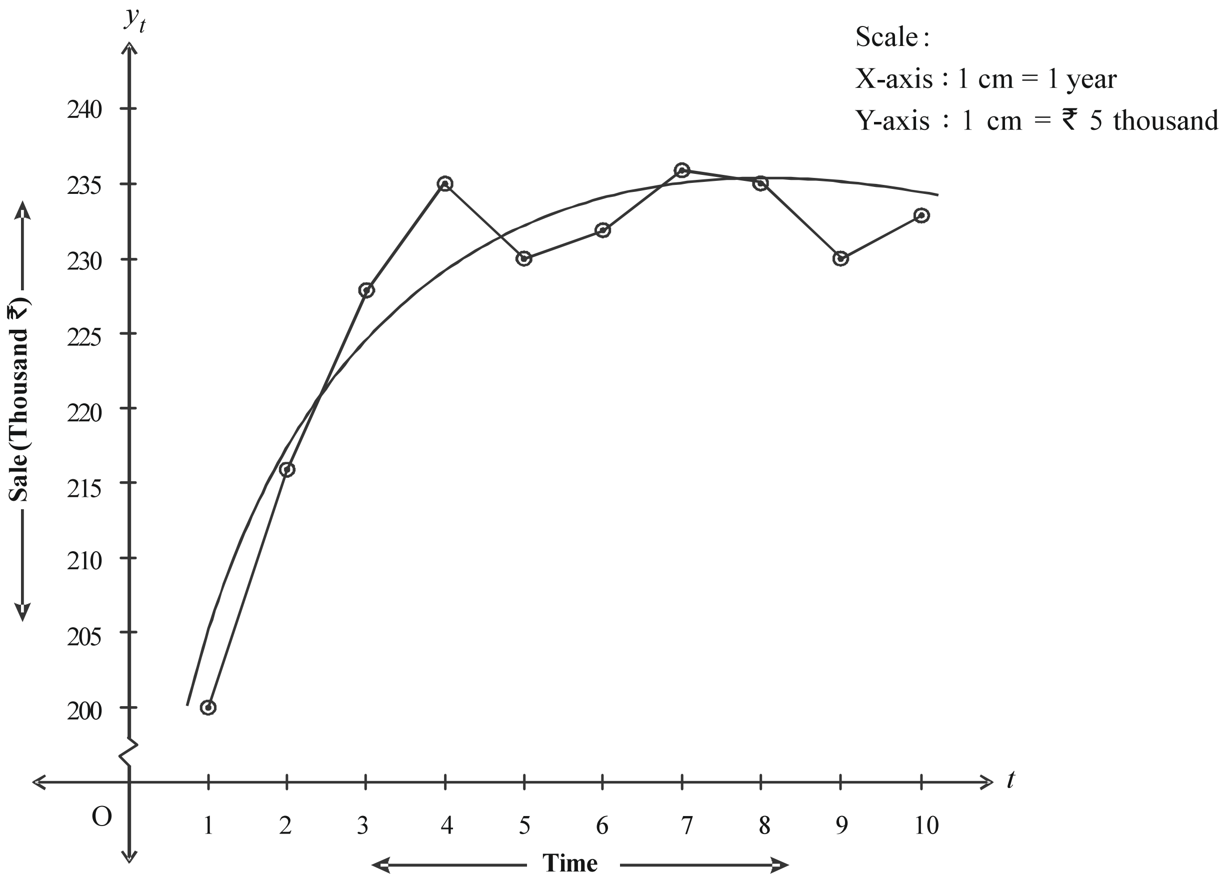
1. $\hat{y} = 23.18 + 0.09\,t$, $\hat{y} =$ 23.91 crore for year 2017

2. $\hat{y} = 49.1 - 2.1\,t$, $\hat{y} =$ 36.5 thousand for year 2016

3.

Month	April 2015	May	June	July	August	Sept.	Oct.	Nov.	Dec.	Jan. 2016
Three monthly moving average	–	71.33	68.67	66.67	65	63.33	63.33	65	66	–

4.



5.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Five yearly moving average	–	–	118.4	119.2	119	118.8	118	119.4	120.6	–	–

Section F

1. $\hat{y} = 30.26 + 1.19\,t$, $\hat{y} =$ 39.78 crore tons for year 2016-17

$\hat{y} =$ 40.97 crore tons for year 2017-18

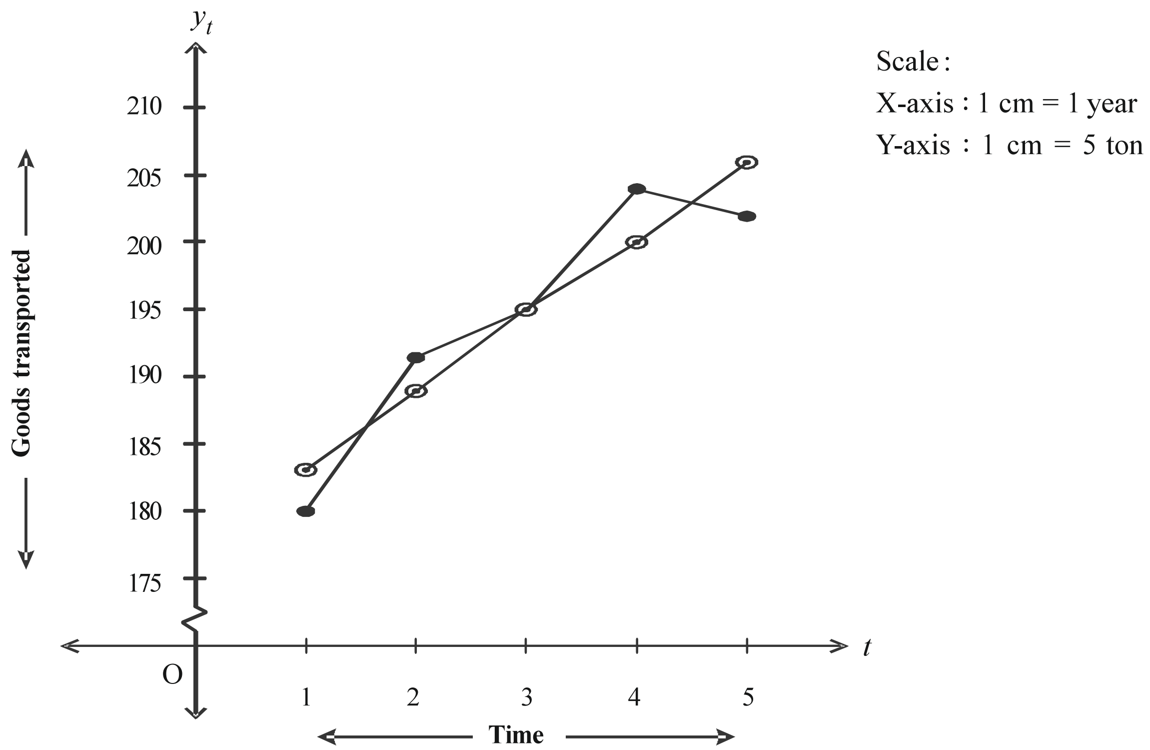
2.

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Four yearly moving average	–	–	358.25	378	395.63	406.63	411.38	415.88	–	–

3. $\hat{y} = 22.55 - 0.39\,t$, $\hat{y} =$ 19.43 for year 2016 $\hat{y} =$ 19.04 for year 2017

4. $\hat{y} = 177.8 + 5.6\,t$, $\hat{y} =$ 211.4 ton for year 2016

Year	2011	2012	2013	2014	2015
Estimated value of trend (tons)	183.4	189	194.6	200.2	205.8



5.

Month	March				April				May			
Week	1	2	3	4	1	2	3	4	1	2	3	4
Four weekly moving average	–	–	38.44	38.7	38.97	39.62	41.29	43.05	44.4	45.72	–	–

●●●

ગુજરાત રાજ્યના શિક્ષણવિભાગના પત્ર-ક્રમાંક
મશબ/1215/178/છ, તા. 24-11-2016-થી મંજૂર

STATISTICS

(Part 2)

Standard 12

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.

રાજ્ય સરકારની વિનામૂલ્યે યોજના હેઠળનું પુસ્તક

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

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
Shri Yatin A. Parikh

Translation

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

Reviewers

Shri Dilip S. Gandhi
Smt. Saroj M. Dave
Shri Unnikumaran. K. M.
Dr. Sejal A. Desai
Smt. Beena Joseph
Dr. Nirav B. Shah

Language Reviewer

Dr. Chetan Mewada

Artist

Media Graphics

Co-ordinator

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

Preparation and Planning

Dr. Kamlesh N. Parmar
(Dy. Director : Academic)

Lay-out and Planning

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

PREFACE

Gujarat State Board of School Textbooks has prepared new textbooks as per the new curricula developed by the Gujarat State Secondary and Higher Secondary Education Board and which has been sanctioned by the Education Department of the Government of Gujarat. A panel of experts from Universities/Colleges, Teachers Training Colleges and Schools have put lot of efforts in preparing the manuscript of the subject. It is then reviewed by another panel of experts to suggest changes and filter out the mistakes, if any. The suggestions of the reviewers are considered thoroughly and necessary changes are made in the manuscript. Thus, the Textbook Board takes sufficient care in preparing an error-free manuscript. The Board is vigilant even while printing the textbooks.

The Board expresses the pleasure to publish the Textbook of **Statistics (Part 2)** for **Std. 12** which is a translated version of Gujarati. The Textbook Board is thankful to all those who have helped in preparing this textbook. However, we welcome suggestions to enhance the quality of the textbook.

H. N. Chavda
Director
Date : 15-02-2017

Dr. Nitin Pethani
Executive President
Gandhinagar

First Edition : 2017

Published by : H. N. Chavda, Director, on behalf of Gujarat State Board of School Textbooks, 'Vidyayan', 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;

(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.

*Constitution of India : Section 51-A

CONTENTS

•

"Statistically, the probability of any one of us being here is so small that the mere fact of our existence should keep us all in a state of contented dazzlement."

– Lewis Thomas

1

Probability

Contents :

1.1 Introduction

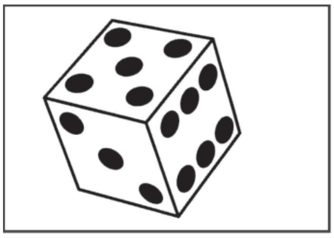
Many events occur in our day-to-day life. We can definitely say for many events that these events will certainly happen. For example, each person taking birth will die, a fruit freely falling from a tree will fall on ground, if the profit per item of a trader is ₹ 10 then he will earn a profit of ₹ 500 by selling 50 items, if a person invests ₹ 1,00,000 in a nationalised bank at an annual interest rate of 7.5 percent then the interest received will be ₹ 7,500, etc. These events are certain but some events are such that we can not be definitely say in advance whether they will happen. For example, getting head on the upper side after tossing a balanced coin, getting number 3 on the upper side of a die when a six faced unbiased die is thrown, the new baby to be born will be a boy, an item produced in a factory is non-defective, what will be the total rainfall in a certain region in the current year, what will be the wheat production in a state in the current year, what will be the result of a cricket match played between teams of two countries, etc. We cannot say with certainty that these events will definitely occur. It is not possible to give precise prediction about the occurrance of such events. We can intuitively get some idea about possibility of happening (or not happening) for these events but there is uncertainty regarding happening (or not happening) of these events. We accept that the occurrence (or non-occurrence) of these events depends upon an unknown element which is called chance. Such events which depend on chance are called random events. Probability is used to numerically express the possibility of these uncertain events. We shall study the theory of probability, the classical definition of probability, its statistical definition and the illustrations showing utility of probability. Now, let us see the explanation of certain terms which are useful to study probability.

1.2 Random Experiment and Sample Space

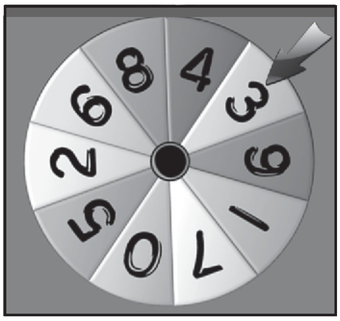
1.2.1 Random Experiment

Let us consider the following experiments :

Experiment 1 : Toss a balanced coin. Any one outcome is obtained out of two posible outcomes (i) Head-*H* (ii) Tail-*T* for this experiment. (We assume that the coin does not stand on its edge.) Thus, *'H'* and *'T'* are the only possible outcomes for the experiment of tossing a coin. But which of the outcomes will be obtained among these two outcomes cannot be said with certainty before conducting the experiment.



Experiment 2 : Throw a balanced die with six faces marked with numbers 1, 2, 3, 4, 5, 6 on it. Note the number appearing on its upper face. Any one outcome among the six possible outcomes 1, 2, 3, 4, 5, 6 will be obtained. There are only six possible outcomes 1, 2, 3, 4, 5, 6 for this experiment of throwing a die but which of the six outcomes will be obtained cannot be said with certainty before conducting the experiment.



Experiment 3 : Suppose there is a wheel marked with 10 numbers 0, 1, 2, ... , 9 and a pointer is kept against it. If this wheel is rotated with hand, it will spin and become stable after some time. When the wheel stops, any one of the numbers 0, 1, 2, ... , 9 will appear against the pointer. This number is the winning number. There are total ten possible outcomes 0, 1, 2, ... , 9 for this experiments. But which of the ten numbers will be obtained as a winning number cannot be said with certainty before conducting the experiment.

The experiments 1, 2, 3 shown above are called random experiments. A random experiment is defined as follows. The experiment which can be independently repeated under indentical conditions and all its possible outcomes are known but which of the outcomes will appear cannot be predicted with certainty before conducting the experiment is called a random experiment. The following characteristics of the random experiment can be deduced from its definition :

(1) A random experiment can be independently repeated under almost identical conditions.

(2) All possible outcomes of the random experiment are known but which of the outcomes will appear cannot be predicted before conducting the experiment.

(3) The random experiment results into a certain outcome.

1.2.2 Sample Space

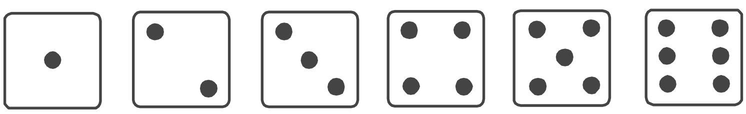
The set of all possible outcomes of a random experiment is called a sample space of that random experiment. The sample space is generally denoted by U or S. The elements of sample space are called sample points.

The sample space of the random experiment in the earlier discussion can be obtained as follows :

Experiment 1 : Toss a balanced coin. There are total two possible outcomes for this random experiment : H and T. Thus, the Sample Space can be written here as $U = \{H, T\}$ or $U = \{T, H\}$.

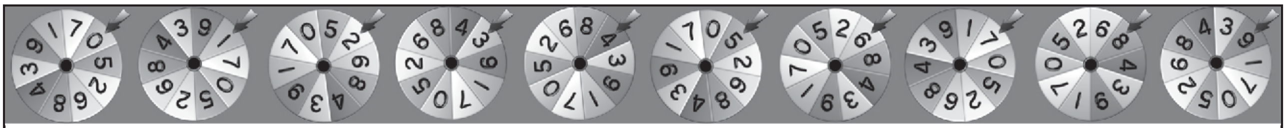
Head - H   Tail - T

Experiment 2 : Throw a balanced die with six faces with numbers 1, 2, 3, 4, 5, 6 on it. There are total six possible outcomes for this random experiment : 1, 2, 3, 4, 5, 6. Thus, the sample space is $U = \{1, 2, 3, 4, 5, 6\}$.



1 2 3 4 5 6

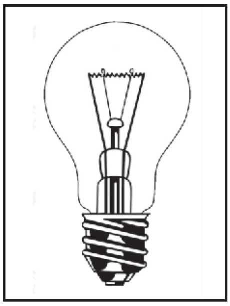
Experiment 3 : To decide the winning number by rotating a wheel marked with numbers 0, 1, 2, ... , 9. There are total ten possible outcomes for this random experiment. Thus, the sample space $U = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.



0 1 2 3 4 5 6 7 8 9

Finite Sample Space : If the total number of possible outcomes in the sample space is finite then it is called a finite sample space. For example, the sample spaces of all the three random experiments given above are finite sample spaces.

Infinite Sample Space : If the total number of possible outcomes in the sample space of a random experiment is infinite then it is called an infinite sample space. For example, if the life of electric bulbs (L) from a production is recorded in hours then it is a real number. The value of L will be 0 or more. Thus, there will be infinite possible outcomes for an experiment of measuring life of bulbs. The sample space will be $U = \{L \mid L \geq 0,\ L \in R\}$. If the maximum life of electric bulbs is assumed to be 700 hours, the sample space will be $U = \{L \mid 0 \leq L \leq 700;\ L \in R\}$ which is an Infinite sample space.



Now we shall see some more illustrations of sample space of a random experiment.

Illustration 1 : Two balanced coins are tossed simultaneously. Write the sample space of this random experiment.

We shall consider any one of the two coins here as the first coin and the other as the second coin. The outcome of this experiment will be as shown in the following diagram.

If we denote the head as H and the tail as T, the sample space will be as follows :

$U = \{HH, HT, TH, TT\}$

Any one of the outcomes out of H and T can be obtained on the first coin. Thus, this action can be done in two ways and the other coin can also show one of the outcomes H and T which can also be done in two ways. According to the fundamental principle of counting for multiplication, the total number of outcomes will be $2 \times 2 = 2^2 = 4$. It should be noted here that the sample space for the experiment of tossing one balanced coin two times will also be the same as above.

Illustration 2 : Two balanced dice are thrown where each die has numbers 1 to 6 on the six sides. Write the sample space of this experiment.

We shall consider any one die as the first die and the other will be called the second die. The number on the first die will be shown as i and the number on the second die will be shown as j. The following sample space will be obtained by denoting the pair of numbers on the two dice as (i, j) where $i, j = 1, 2, 3, 4, 5, 6$.

$$U = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),$$
$$(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),$$
$$(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),$$
$$(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),$$
$$(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),$$
$$(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$$

OR

$$U = \{(i, j);\ i, j = 1, 2, 3, 4, 5, 6\}$$

Any one of the integers 1 to 6 can be shown on the upper side of the first die which can occur in 6 ways and the second die can also show one of the integers among 1 to 6 which will also occur in 6 ways. The total number of outcomes will be $6 \times 6 = 6^2 = 36$ according to fundamental principle of counting for multiplication. Similarly, the sample space for the random experiment of throwing three balanced dice simultaneously will have $6^3 = 216$ total outcomes.

Illustration 3 : Write the sample space of the random experiment of finding the number of defective items while testing the quality of 1000 items produced in a factory.

If the defective items are found among 1000 items produced in the factory then the number of defective items in the production can be 0, 1, 2, ... , 1000. Thus, the sample sapce will be as follows :

$$U = \{0, 1, 2, ..., 1000\}$$

Illustration 4 : Write the sample space of random experiment of randomly selecting three numbers from the first four natural numbers.

If three numbers are selected simultaneously from the first four natural numbers 1, 2, 3, 4 then those three numbers can be $(1, 2, 3)$, $(1, 2, 4)$, $(1, 3, 4)$ or $(2, 3, 4)$. Thus, the sample space of the random experiment will be as follows :

$$U = \{(1, 2, 3), (1, 2, 4), (1, 3, 4), (2, 3, 4)\}$$

3 numbers are to be selected here from the 4 numbers which has ${}^4C_3 = 4$ combinations. Thus, the total number of outcomes for this random experiment is 4.

Illustration 5 : Write the sample space of a random experiment of randomly selecting any one number from the natural numbers.

The natural numbers are 1, 2, 3, ... If one number is randomly selected from these numbers then the sample space will be as follows :

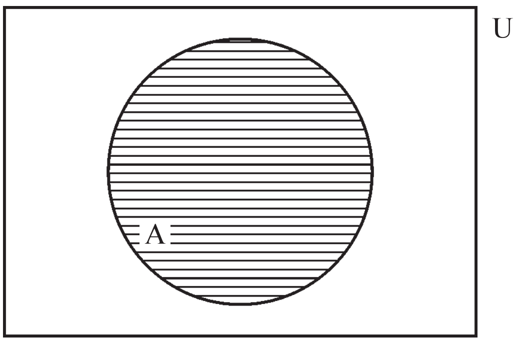
$$U = \{1, 2, 3, 4, ...\}$$

It should be noted here that this is an infinite sample space.

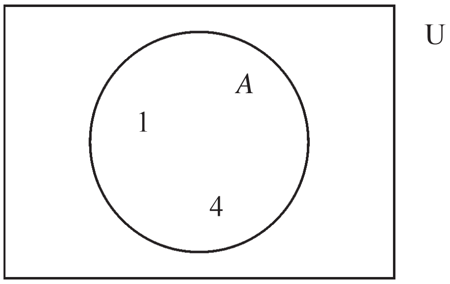
1.3 Events : Certain Event, Impossible Event, Special Events

We will study the different types of events by first understanding the meaning of an event.

(1) Event : A subset of the sample space of a random experiment is called an event. The events are generally denoted by letters *A, B, C,* ... or as A_1, A_2, A_3, ... The set formed by the sample points showing favourable outcomes of an event *A* will be a subset of the sample space *U*. Thus, any event *A* associated with the random experiment is the subset of sample space *U*. This is denoted as $A \subset U$.



For example, the sample space of a random experiment of throwing a balanced die is $U = \{1, 2, 3, 4, 5, 6\}$. If the event of obtaining a complete square as a number on the upper side of the die is denoted by A then event $A = \{1, 4\}$.



Now, we shall show that an event is a subset of the sample space by taking a few examples of events in the random experiment of throwing two balanced dice.

- A_1 = the sum of numbers on the dice is 6.

  $\therefore \ A_1 = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$

- A_2 = the numbers on the dice are same.

  $\therefore \ A_2 = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$

- A_3 = the sum of numbers on the dice is more than 9.

  $\therefore \ A_3 = \{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\}$

All these subsets are called events.

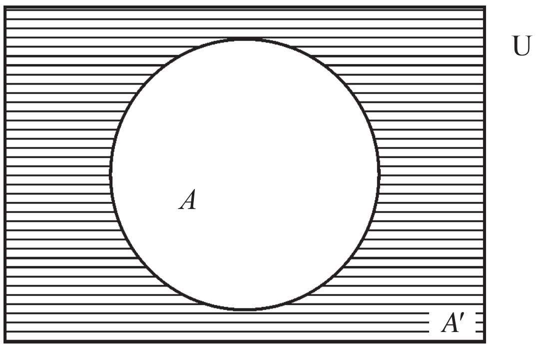
(2) Impossible Event : The special subset ϕ or $\{ \}$ of the sample space of a random experiment is called an impossible event. Impossible event is an event which never occurs. It is denoted by ϕ or $\{ \}$.

For example, the event of getting both head (H) and tail (T) on a balanced coin is an impossible event.

(3) Certain Event : The special subset U of the sample space of random experiment is called a certain event. The certain event is an event which always occurs. It is denoted by U.

For example, the day next to Saturday is Sunday, the number on the upper side die when a balanced die is thrown is less than 7, etc. are certain events.

(4) Complementary Event : Suppose U is a finite sample space and A is one of its events. The set of all the outcomes or elements of U which are not in the event A is called as complementary event of A. The complementary event of event A is denoted by A', $\overline{A}$, A^c. We will use the notation A' for complementary event of A.



A' = Complementary event of event A.

= Non-occurrence of event A.

= $U - A$.

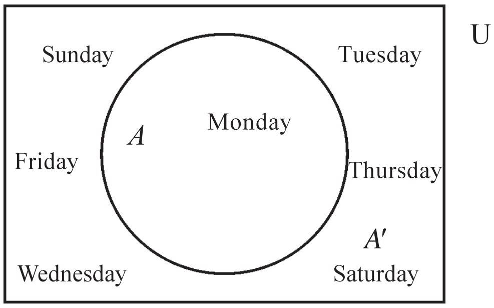
For example, the sample space of the random experiment of finding the day when a cargo ship will reach port Y after leaving from port X will be as follows.

U = {Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday}

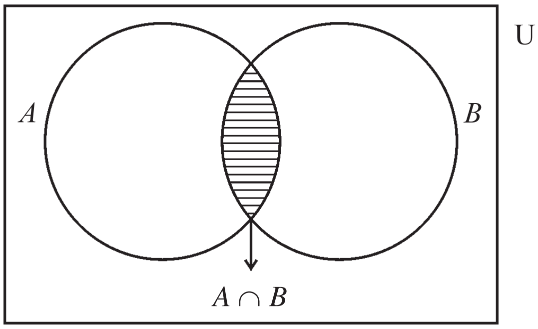
Suppose A denotes that this ship reaches port Y on Monday. Then the set of days except Monday will be the set of outcomes of event A'.

A = {Monday}

A' = $U - A$ = {Sunday, Tuesday, Wednesday, Thursday, Friday, Saturday}



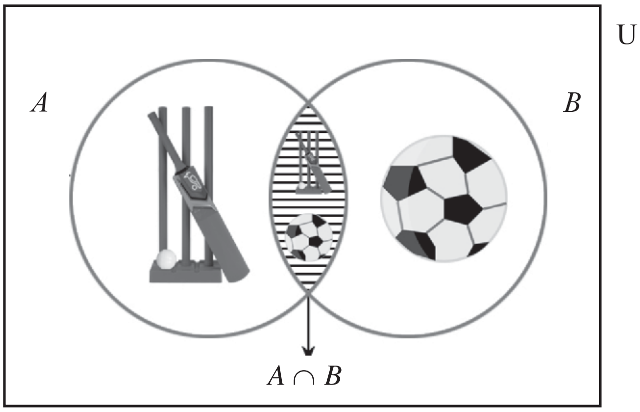
(5) Intersection of Events : Suppose A and B are two events of a finite sample space U. The event where events A and B occur simultaneously is called the intersection of two events A and B. It is denoted by $A \cap B$.



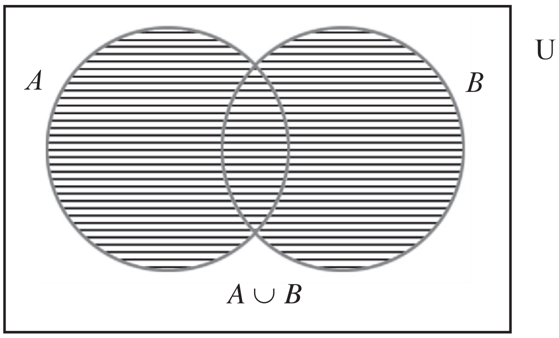
$A \cap B$ = Intersection of two events A and B

= Simultaneous occurrence of events A and B

For example, some of the students studying in a class of a school are the members of school cricket team and some students are members of school football team. Let us denote the event that a student is a member of cricket team by event A and the event that a student is a member of football team by B. If one student is randomly selected from this class then the event that the student is a member of school cricket and football team is called $A \cap B$, the intersection of events A and B.



(6) Union of Events : Suppose A and B are any two events of a finite sample space U. The event where the event A occurs or the event B occurs or both the events A and B occur is called the union of events A and B. It is denoted by $A \cup B$.

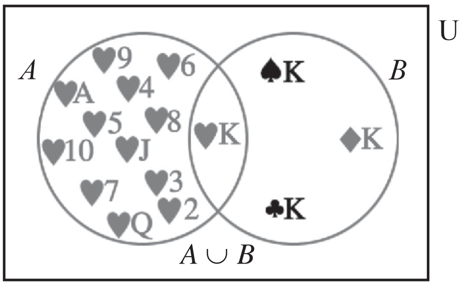


$A \cup B$ = Union of events A and B

= Event A occurs or event B occurs or both events A and B occur together

= At least one of the event A and B occurs.

For example, the event $A \cup B$ of getting heart card (say event A) or a king (say event B) when a card is drawn randomly from a pack of 52 cards will be as follows :



$$A = \{H_A, H_2, H_3, H_4, H_5, H_6, H_7, H_8, H_9, H_{10}, H_J, H_Q, H_K\}$$

$$B = \{S_K, D_K, C_K, H_K\}$$

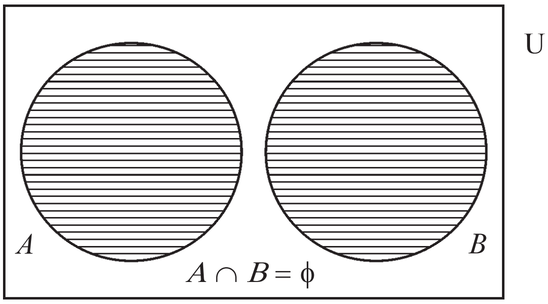
$$A \cup B = \{H_A, H_2, H_3, H_4, H_5, H_6, H_7, H_8, H_9, H_{10}, H_J, H_Q, H_K, S_K, D_K, C_K\}$$

Thus, the occurrence of $A \cup B$ is selecting any one of these 16 cards.

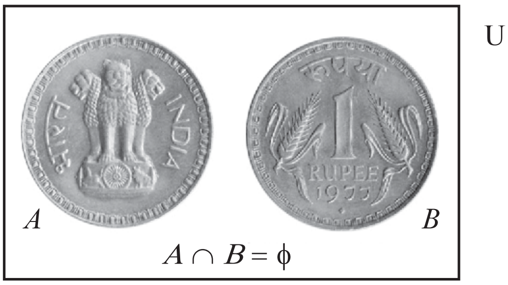
The suits and types of cards are shown as follows :

Spade - S	Diamond - D	Club - C	Heart - H
Ace - A	King - K	Queen - Q	Jack - J

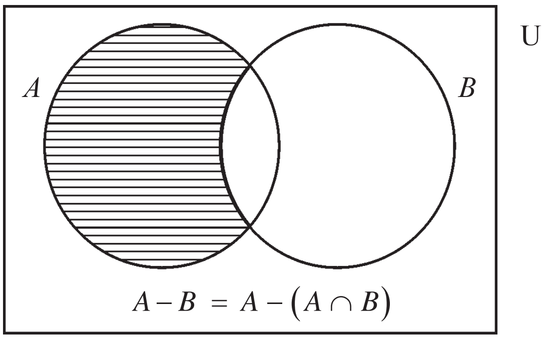
(7) Mutually Exclusive Events : Suppose A and B are any two events of a finite sample space U. Events A and B do not occur together which means $A \cap B = \phi$ or in other words, event B does not occur when event A occurs and event A does not occur when event B occurs then the events A and B are called mutually exclusive events.



For example, toss a balanced coin. Denote the outcome H on the coin as event A and the outcome T on the coin as B. We get $A = \{H\}$ and $B = \{T\}$. It is clear that $A \cap B = \phi$ because when we get H in a trial, it is not possible to get the outcome T in the same trial and vice versa, when we get T in a trial, it is not possible to get the outcome H in the same trial in the random experiment of tossing a balanced coin. Thus, these two events cannot occur simultaneously.



(8) Difference Event : Suppose A and B are any two events of a finite sample space U. The set of elements or outcomes where event A happens but event B does not happen is called the difference of events A and B. It is denoted by $A - B$. It is clear from the venn diagram given here that



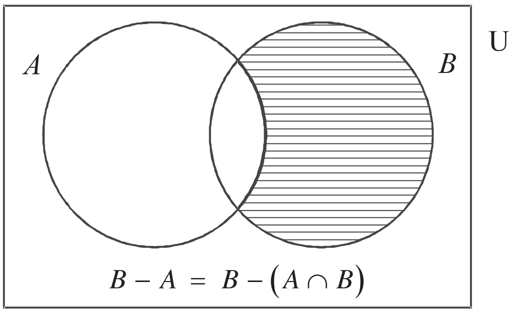
$$A - B = A \cap B' = A - (A \cap B) = (A \cup B) - B$$

$A - B$ = Difference of events A and B

= Event A happens but event B does not happen

= Only A happens out of events A and B.

Similarly, for two events A and B of a finite sample space U, the set of elements or outcomes where B happens but A does not happen is called as the difference of event B and event A. It is denoted by $B - A$. It is clear from the venn diagram given here that,



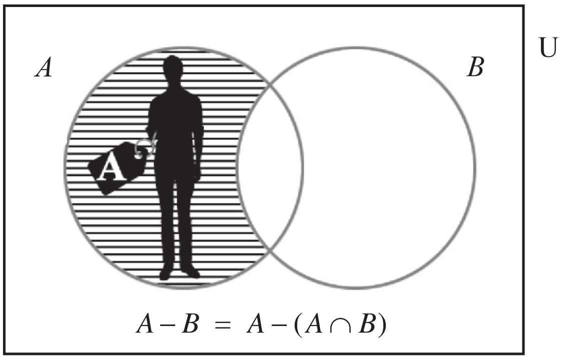
$$B - A = A' \cap B = B - (A \cap B) = (A \cup B) - A$$

$B - A$ = Difference of events B and A

= Event B happens but event A does not happen

= Only event B happens out of events A and B

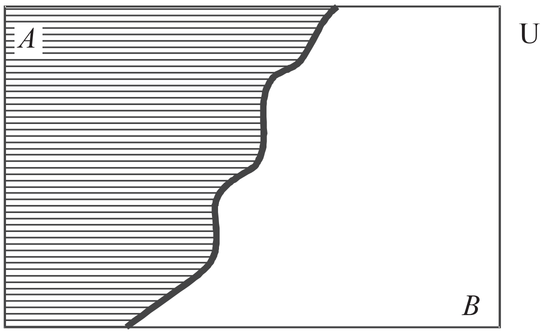
For example, two employees A and B among the employees working in an office are friends. Denote the presence of employee A in the office as event A and the presence of employee B in the office as event B. On a ctertain day, if it is said that only the employee A is present in the office out of employees A and B then it is clear that among two employees A and B, employee A is present but employee B is not present. Thus, it is called the difference of two events $A - B$ for events A and event B. Here,



$A - B$ = Only employee A is present in the office among the employees A and B

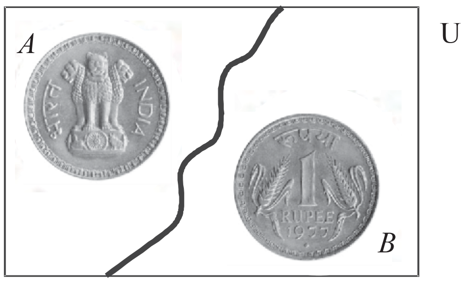
$B - A$ = Only employee B is present in the office among the employees A and B

(9) Exhaustive Events : If the group of favourable outcomes of events of random experiment is the sample space then the events are called exhaustive events. Suppose A and B are any two events of a sample space U. The events A and B are called the exhaustive events if the union $A \cup B$ of the two events A and B is the sample space U, that is $A \cup B = U$.



For example, denote the outcome H as event A and the outcome T as event B when a balanced coin is tossed. It is clear in this case that $A = \{H\}$, $B = \{T\}$ and $A \cup B = \{H, T\} = U$.

$\therefore$ A and B are exhaustive events.



(10) Mutually Exclusive and Exhaustive Events : Suppose A and B are two events of a finite sample space U. These two events A and B are called the mutually exclusive and exhaustive events if $A \cap B = \phi$ and $A \cup B = U$. It should be noted here that all the mutually exclusive events need not be exhaustive events and similarly, all the exhaustive events need not be the mutually exclusive events.

For example, consider the sample space $U = \{1, 2, 3, 4, 5, 6\}$ of the experiment of throwing a balanced die. Let the event A = getting odd number on the die $= \{1, 3, 5\}$ and event B = getting even number on the die $= \{2, 4, 6\}$. It is clear that $A \cap B = \phi$ and $A \cup B = U$. Thus, the events A and B are mutually exclusive and exhaustive.

(11) Elementary Events : The events formed by all the subsets of single elements of the sample space U of a random experiment are called the elementary events. The elementary events are mutually exclusive and exhaustive.

For example, consider the sample space $U = \{H, T\}$ for the random experiment of tossing a balanced coin. The events $A = \{H\}$ and $B = \{T\}$ having single elements are the elementary events. Since $A \cap B = \phi$ and $A \cup B = U$ in this case, it can be said that the elementary events are mutually exclusive and exhaustive.

Illustration 6 : There are 3 yellow and 2 pink flowers in a basket. One flower is randomly selected from this basket. Denote the selection of yellow flower as an event A and the selection of pink flower as the event B. Find the sets representing the following events and answer the given questions.

(1) U **(2)** A **(3)** B **(4)** A' **(5)** B' **(6)** $A \cap B$ **(7)** $A \cup B$ **(8)** $A \cap B'$ **(9)** $A' \cap B$

(10) State the elementary events of the sample space for this random experiment.

(11) Can it be said that the events A and B are mutually exclusive events ? Give reason.

(12) Can it be said that the events A and B are exhaustive events ? Give reason.

We will denote the 3 yellow flowers in the basket as Y_1, Y_2, Y_3 and the 2 pink flowers as P_1, P_2. The sets representing the required events will be as follows :

(1) $U = \{Y_1, Y_2, Y_3, P_1, P_2\}$

(2) $A = \{Y_1, Y_2, Y_3\}$

(3) $B = \{P_1, P_2\}$

(4) $A' = U - A = \{Y_1, Y_2, Y_3, P_1, P_2\} - \{Y_1, Y_2, Y_3\}$

$= \{P_1, P_2\}$

(5) $B' = U - B = \{Y_1, Y_2, Y_3, P_1, P_2\} - \{P_1, P_2\}$

$= \{Y_1, Y_2, Y_3\}$

(6) $A \cap B = \{Y_1, Y_2, Y_3\} \cap \{P_1, P_2\}$

$= \phi$

(7) $A \cup B = \{Y_1, Y_2, Y_3\} \cup \{P_1, P_2\}$

$= \{Y_1, Y_2, Y_3, P_1, P_2\}$

(8) $A \cap B' = \{Y_1, Y_2, Y_3\} \cap \{Y_1, Y_2, Y_3\}$

$= \{Y_1, Y_2, Y_3\}$

OR

$A \cap B' = A - (A \cap B)$

$= \{Y_1, Y_2, Y_3\} - \phi$

$= \{Y_1, Y_2, Y_3\}$

(9) $A' \cap B = \{P_1, P_2\} \cap \{P_1, P_2\}$

$= \{P_1, P_2\}$

OR

$A' \cap B = B - (A \cap B)$

$= \{P_1, P_2\} - \phi$

$= \{P_1, P_2\}$

(10) The elementary events are the subsets with one element. If we denote the different elementary events as $E_1, E_2, E_3, ...$ then

$E_1 = \{Y_1\}, \quad E_2 = \{Y_2\}, \quad E_3 = \{Y_3\}, \quad E_4 = \{P_1\}, \quad E_5 = \{P_2\}$

(11) The events A and B can be called mutually exclusive events because according to the definition of mutually exclusive events, the events A and B are called the mutually exclusive events if $A \cap B = \phi$. It can be seen from the answer to the question 6 that $A \cap B = \phi$.

(12) The events A and B can be called exhaustive events because according to the definition of exhaustive events, the A and B are called the exhaustive events if $A \cup B = U$. It can be seen from the answer to the question 7 that $A \cup B = U$.

Illustration 7 : The events A and B of a random experiment are as follows :

$$A = \{1, 2, 3, 4\}, \qquad B = \{-1, 0, 1\}$$

If the sample space $U = A \cup B$ then find the sets showing the following events.

(1) B' (2) $A' \cap B$ (3) $A - B$

Here, $A = \{1, 2, 3, 4\}$

$B = \{-1, 0, 1\}$

$U = A \cup B = \{1, 2, 3, 4\} \cup \{-1, 0, 1\}$

$= \{-1, 0, 1, 2, 3, 4\}$

(1) $B' = U - B$

$= \{-1, 0, 1, 2, 3, 4\} - \{-1, 0, 1\}$

$= \{2, 3, 4\}$

(2) $A' \cap B = B - (A \cap B)$

First we find $A \cap B$,

$A \cap B = \{1, 2, 3, 4\} \cap \{-1, 0, 1\}$
$= \{1\}$

Now, $A' \cap B = B - (A \cap B)$

$= \{-1, 0, 1\} - \{1\}$

$= \{-1, 0\}$

Alternate Method :

$A' = U - A = \{-1, 0, 1, 2, 3, 4\} - \{1, 2, 3, 4\} = \{-1, 0\}$

$\therefore A' \cap B = \{-1, 0\} \cap \{-1, 0, 1\} = \{-1, 0\}$

(3) $A - B = \{1, 2, 3, 4\} - \{-1, 0, 1\}$
$= \{2, 3, 4\}$

Illustration 8 : One number is randomly selected from the first 50 natural numbers. Find the sets showing the following events.

(1) The number selected is a multiple of 5 or 7.

(2) The number selected is a multiple of both 5 and 7.

(3) The number selected is a multiple of 5 but not a multiple of 7.

(4) The number selected is only a multiple of 7 out of 5 and 7.

If one number is selected from the first 50 natural numbers then the group of all possible outcomes of this experiment, which is the sample space U, is as follows :

$U = \{1, 2, 3, ..., 50\}$

Event A = Selected number is a multiple of 5

$= \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$

Event B = Selected number is a multiple of 7

$= \{7, 14, 21, 28, 35, 42, 49\}$

Now, the required events are as follows :

(1) The event of selecting a number which is a multiple of 5 or 7 = $A \cup B$

$\therefore A \cup B = \{5, 7, 10, 14, 15, 20, 21, 25, 28, 30, 35, 40, 42, 45, 49, 50\}$

(2) The event of selecting a number which is a multiple of both 5 and 7 = $A \cap B$

$\therefore A \cap B = \{35\}$

(3) The event of selecting a number which is a multiple of 5 but not of 7 = $A \cap B'$

$$\therefore A \cap B' = A - (A \cap B)$$
$$= \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\} - \{35\}$$
$$= \{5, 10, 15, 20, 25, 30, 40, 45, 50\}$$

(4) The event of selecting a number which is only a multiple of 7 out of 5 and 7 = $A' \cap B$

$$\therefore A' \cap B = B - (A \cap B)$$
$$= \{7, 14, 21, 28, 35, 42, 49\} - \{35\}$$
$$= \{7, 14, 21, 28, 42, 49\}$$

Illustration 9 : The events A_1 and A_2 of a random experiment are defined as follows. Find the sets showing union event $A_1 \cup A_2$ and intersection event $A_1 \cap A_2$.

$A_1 = \{x \mid x = -1, 0, 1\}$, $A_2 = \{x \mid x = 1, 2, 3\}$

It is given that $A_1 = \{-1, 1, 0\}$ and $A_2 = \{1, 2, 3\}$.

Union of events $A_1 \cup A_2 = \{-1, 0, 1, 2, 3\}$

Intersection of events $A_1 \cap A_2 = \{1\}$

Illustration 10 : A factory produces screws of different lengths. The length (in cm) of screw is denoted by x. The events A_1 and A_2 are defined as follows in the experiment of finding the length of selected screws. Find the events showing union event $A_1 \cup A_2$ and intersection event $A_1 \cap A_2$.

$A_1 = \{x \mid 0 < x < 1\}$, $A_2 = \{x \mid \frac{1}{2} \le x < 2\}$

If is given that $A_1 = \{x \mid 0 < x < 1\}$ and $A_2 = \{x \mid \frac{1}{2} \le x < 2\}$.

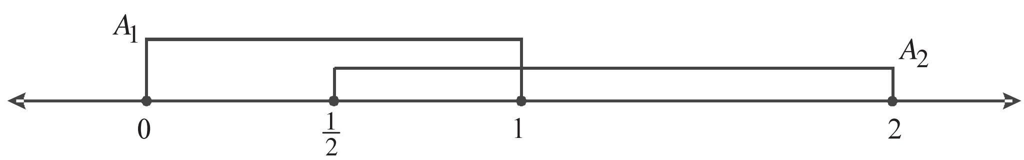
Union of events $A_1 \cup A_2 = \{x \mid 0 < x < 2\}$

$= (0, 2)$ (interval form)

Intersection of events $A_1 \cap A_2 = \{x \mid \frac{1}{2} \le x < 1\}$

$= \left[\frac{1}{2}, 1\right)$ (interval form)

See the following diagram carefully for better explanation of $A_1 \cup A_2$ and $A_1 \cap A_2$.



Exercise 1.1

1. State the sample space for the following random experiments :

(1) A balanced die is thrown three times.

(2) A balanced die with six sides and a balanced coin are tossed together.

(3) Two persons are to be selected from five persons *a, b, c, d, e*.

2. Write the sample space for the marks (in integers) scored by a student appearing for an examination of 100 marks and state the number of sample points in it.

3. Write the sample space for randomly selecting one minister and one deputy minister from four persons.

4. A balanced die in thrown in a random experiment till the first head is obtained. The experiment is terminated with a trial of first head. Write the sample space of this experiment and state whether it is finite or infinite.

5. Write the sample space for the experiment of randomly selecting three numbers from the first five natural numbers.

6. The sample space of a random experiment of selecting a number is $U = \{1, 2, 3, ... , 20\}$.

Write the sets showing the following events :

(1) The selected number is odd number

(2) The selected number is divisible by 3

(3) The selected number is divisible by 2 or 3.

7. One family is selected from the families having two children. The sex (male or female) of the children from this family is noted. State the sample space of this experiment and write the sets showing the following events :

(1) Event A_1 = One child is a female

(2) Event A_2 = At least one child is a female.

8. Two six faced balanced dice are thrown simultaneously. State the sample space of this random experiment and hence write the sets showing the following events :

(1) Event A_1 = The sum of numbers on the dice is 7

(2) Event A_2 = The sum of numbers on the dice is less than 4

(3) Event A_3 = The sum of numbers on the dice is divisible by 3

(4) Event A_4 = The sum of numbers on the dice is more than 12.

9. Two numbers are selected at random from the first five natural numbers. The sum of two selected numbers is at least 6 is denoted by event A and the sum of two selected numbers is even is denoted by event B. Write the sets showing the following events and answer the given questions :

(1) U (2) A (3) B (4) $A \cup B$ (5) $A \cap B$ (6) A' (7) $A - B$ (8) $A' \cap B$

(9) Can it be said that the events A and B are mutually exclusive ? Give reason.

(10) State the number of sample points in the sample space of this random experiment.

10. Three female employees and two male employees are working in an office. One employee is selected from the employees of this office for training. The event that the employee selected for the training is a female is denoted by A and the event that this employee is a male is denoted by B. Find the sets showing the following events and answer the given questions :

(1) U (2) A (3) B (4) $A \cup B$ (5) $A \cap B$ (6) $A' \cap B$

(7) Can it be said that the events A and B are mutually exclusive ? Give reason.

(8) Can it be said that the events A and B are exhaustive ? Give reason.

11. One card is randomly drawn from a pack of 52 cards. If drawing a spade card is denoted by event A and drawing a card from ace to ten (non-face card) is denoted by B then write the sets showing the following events :

(1) U (2) A (3) B (4) $A \cup B$ (5) $A \cap B$ (6) B'

12. The events A_1 and A_2 of a random experiment are as follows. Find the sets showing the union event $A_1 \cup A_2$ and intersection event $A_1 \cap A_2$.

$$A_1 = \{x \mid 0 < x < 5\}, \qquad A_2 = \{x \mid -1 < x < 3, x \text{ is an integer}\}$$

13. The events A_1 and A_2 of a random experiment are as follows. Find the sets showing the union event $A_1 \cup A_2$ and intersection event $A_1 \cap A_2$.

$$A_1 = \{x \mid 2 \le x < 6, \ x \in N\}, \qquad A_2 = \{x \mid 3 < x < 9, \ x \in N\}$$

14. The sample space U of a random experiment and its event A are defined as follows. Find the complementary event A' of A.

$$U = \{x \mid x = 0, 1, 2, ..., 10\}, \qquad A = \{x \mid x = 2, 4, 6\}$$

15. The sample space U of a random experiment and its event A are defined as follows. Find the complementary event A' of A.

$$U = \{x \mid 0 < x < 1\}, \qquad A = \{x \mid \tfrac{1}{2} \le x < 1\}$$

*

After getting aquainted with the random experiment, sample space and different events, we shall now study the probability. We shall begin with the definition of probability.

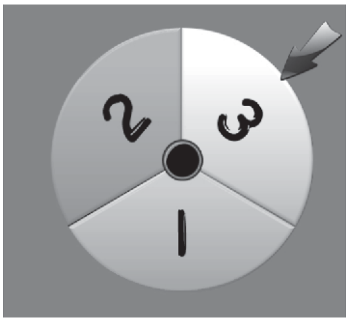
1.4 Mathematical Definition of Probability

To understand the mathematical definition of probability, we shall first understand the two important terms namely equiprobable events and favourable outcomes.

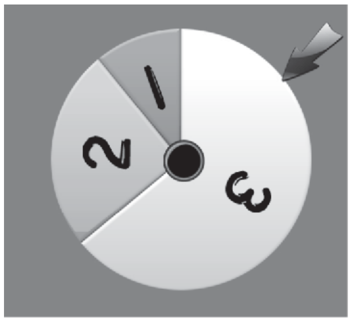
Equiprobable Events : If there is no apparent reason to believe that out of one or more events of a random experiment, any one event is more or less likely to occur than the other events then the events are called as equiprobable events.

For example, a manufacturer of a certain item has two machines M_1 and M_2 in his factory for the production of items. Both the machines produce the same number of items during a day. The lots are made of the produced goods by properly mixing the items produced on both the machines during the day. An item randomly selected from such a lot is made on machine M_1 or machine M_2 are the elementary events which are equiprobable.

Similarly, the wheels A and B marked with numbers 1, 2, 3 as shown in the following pictures are rotated by hand. The number against the pointer is noted down when the wheels stop rotating after some time. It is clear from the picture that all the three numbers on wheel A will come against the pointer are equiprobable events. But the numbers 1, 2 and 3 coming against the pointer for wheel B are not equiprobable events.



Wheel A



Wheel B

Favourable Outcomes : If some outcomes out of all the elementary outcomes in the sample space of random experiment indicate the occurrence of a certain event A then these outcomes are called the favourable outcomes of the event A. For example, a card is drawn from a pack of 52 cards. If event A denotes that the card drawn is a face card then the set of favourable outcomes is as follows :

$$A = \{S_K, D_K, C_K, H_K, S_Q, D_Q, C_Q, H_Q, S_J, D_J, C_J, H_J\}$$

Thus, 12 outcomes are favourable for event A.

Mathematical Definition of probability : Suppose there are total n outcomes in the finite sample space of a random experiment which are mutually exclusive, exhaustive and equiprobable. If m outcomes among them are favourable for an event A then the probability of the event A is $\frac{m}{n}$. The probability of event A is denoted by $P(A)$.

$$P(A) = \text{Probability of event } A$$

$$= \frac{\text{Favourable outcomes of event } A}{\text{Total number of mutually exclusive, exhaustive and equi-probable outcomes of sample space}}$$

$$= \frac{m}{n}$$

Both the numbers $m(\geq 0)$ and $n(> 0)$ are integers and $m \leq n$. It should be noted here that n can not be zero and infinity. The mathematical definition of probability is also called the classical definition.

The assumptions of the mathematical definition are as follows :

(1) The number of outcomes in the sample space of the random experiment is finite.

(2) The number of outcomes in the sample space of the random experiment is known.

(3) The outcomes in the sample space of the random experiment are equi-probable.

We will accept some of the following important results about probability without proof :

(1) The range for the value of probability $P(A)$ for any event A in the sample space U is 0 to 1. Thus, $0 \leq P(A) \leq 1$.

(2) The probability of an impossible event is zero. Earlier we have denoted an impossible event by ϕ. Hence, $P(\phi) = 0$.

(3) The probability of certain event is always 1. Earlier we have denoted a certain event by U. Hence, $P(U) = 1$.

(4) The probability of complementary event A' of event A in the sample space U is $P(A') = 1 - P(A)$.

(5) If $A \subset B$ for two events A and B in the sample space of a random experiment then

- $P(A) \leq P(B)$
- $P(B - A) = P(B) - P(A)$

(6) For two events A and B in the sample space of a random experiment,

- $P(A \cap B) \leq P(A)$ $\quad [\because A \cap B \subset A]$
- $P(A \cap B) \leq P(B)$ $\quad [\because A \cap B \subset B]$
- $P(A) \leq P(A \cup B)$ $\quad [\because A \subset A \cup B]$
- $P(B) \leq P(A \cup B)$ $\quad [\because B \subset A \cup B]$
- $P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$
- $P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B)$
- $P(A - B) = P(A \cap B') = P(A) - P(A \cap B)$
- $P(B - A) = P(A' \cap B) = P(B) - P(A \cap B)$
- $0 \leq P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$

We shall now consider illustrations of finding probability of different events using the mathematical definition.

Illustration 11 : If two balanced coins are tossed, then find the probability of (1) getting one head and one tail and (2) getting at least one head.

The sample space for the random experiment of tossing two balanced coins is as follows :

$$U = \{HH, HT, TH, TT\}$$

$\therefore$ No. of mutually exclusive, exhaustive and equi-probable outcomes $n = 4$.

(1) If A denotes the event of getting one head H and one tail T then HT and TH are two favourable outcomes of event A. Thus, $m = 2$.

From the mathematical definition of probability,

$$P(A) = \frac{m}{n}$$

$$= \frac{2}{4}$$

$$= \frac{1}{2}$$

Required probability $= \frac{1}{2}$

(2) If B denotes the event of getting at least one head then HT, TH, HH are the favourable outcomes of event B. Hence, the number of favourable outcomes $m = 3$ for even B.

From the mathematical definition of probability,

$$P(B) = \frac{m}{n}$$

$$= \frac{3}{4}$$

Required probability $= \frac{3}{4}$

Illustration 12 : Two balanced dice marked with numbers 1 to 6 are thrown simultaneously. Find the probability that (1) sum of numbers on both the dice is 7 (2) sum of numbers on both the dice is more than 10 (3) sum of number on both the dice is at the most 4 (4) both the dice show same numbers (5) sum of numbers on both the dice is 1 (6) sum of numbers on both the dice is 12 or less.

The sample space for throwing two balanced dice simultaneously is as follows :

$$U = \{(i,\ j)\ ;\ i,\ \ i, j = 1, 2, 3, 4, 5, 6\}$$

$\therefore$ Total number of outcomes $n = 36$.

(1) If A_1 denotes that sum of the numbers on the dice is 7 then there are total 6 outcomes $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$ favourable for this event A_1. Thus, the number of favourable outcomes $m = 6$ for event A_1. Probability of event A_1

$P(A_1) = \frac{m}{n}$

$= \frac{6}{36}$

$= \frac{1}{6}$

$\therefore$ Required probability $= \frac{1}{6}$

(2) If A_2 denotes the event that the sum of numbers on two dice is more than 10 then $(5, 6)$, $(6, 5)$, $(6, 6)$ are the favourable outcomes of event A_2. Thus, the number of favourable outcomes $m = 3$ for even A_2. Probability of A_2

$P(A_2) = \frac{m}{n}$

$= \frac{3}{36}$

$= \frac{1}{12}$

Required probability $= \frac{1}{12}$

(3) If A_3 denotes the event that the sum of numbers on two dice is at the most 4 then total 6 outcomes $(1, 1)$, $(1, 2)$, $(1, 3)$, $(2, 1)$, $(2, 2)$, $(3, 1)$ are favourable outcomes of event B. Thus, the number of favourable outcomes $m = 6$ for event A_3. Probability of event A_3

$P(A_3) = \frac{m}{n}$

$= \frac{6}{36}$

$= \frac{1}{6}$

Required probability $= \frac{1}{6}$

(4) Event A_4 = both the dice show the same numbers.

$\therefore$ Total 6 outcomes $(1, 1)$, $(2, 2)$, $(3, 3)$, $(4, 4)$, $(5, 5)$, $(6, 6)$ are favourable for the event A_4. Thus, the number of favourable outcomes $m = 6$ for event A_4. Probability of event A_4

$P(A_4) = \frac{m}{n}$

$= \frac{6}{36}$

$= \frac{1}{6}$

Required probability $= \frac{1}{6}$

(5) Let A_5 be the event that the sum of numbers on two dice is 1. It is obvious that not a single outcome in the sample space is favourable for A_5. Hence, the number of favourable outcomes $m = 0$ for event A_5. Probability of event A_5

$$P(A_5) = \frac{m}{n}$$

$$= \frac{0}{36}$$

$$= 0$$

Required probability $= 0$

(The probability of impossible event is always 0.)

(6) Let A_6 be the event that the sum of numbers on two dice is 12 or less. It is obvious that all the outcomes in the sample space are favourable for event A_6. Hence, the number of favourable outcomes $m = 36$ for the event A_6. Probability of event A_6

$$P(A_6) = \frac{m}{n}$$

$$= \frac{36}{36}$$

$$= 1$$

Required probability $= 1$

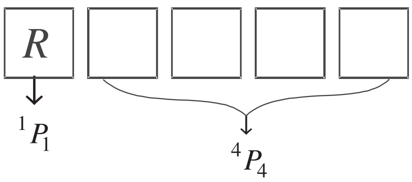
(The probability of certain event is always 1.)

Illustration 13 : Find the probability of getting *R* in the first place in all possible arrangements of each and every letter of the word *RUTVA*.

There are 5 letters *R, U, T, V, A* in the word *RUTVA*. These five letters can be arranged in ${}^5P_5 = 5! = 120$ different ways. Thus, total number of outcomes $n = 120$.

Event of getting *R* in the first place of the arrangement $= A$.

The favourable outcomes of event *A* are obtained as follows :



R can be arranged in the first place in ${}^1P_1$ ways and the remaining four letters *U, T, V, A* in the rest of the four places can be arranged in ${}^4P_4$ ways. According to the fundamental principle of multiplication, there will be ${}^1P_1 \times {}^4P_4$ arrangements of getting *R* in the first place. Hence, the number of favourable outcomes for event *A* will be

$$m = {}^1P_1 \times {}^4P_4 = 1! \times 4! = 1 \times 24 = 24.$$

Probability of event *A* $P(A) = \frac{m}{n}$

$$= \frac{24}{120}$$

$$= \frac{1}{5}$$

Required probability $= \frac{1}{5}$

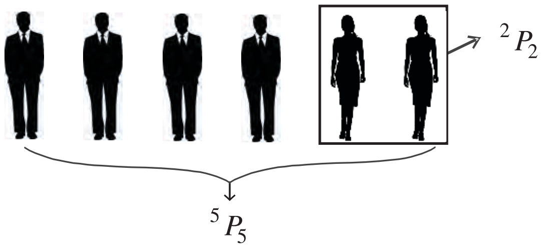
Downloaded from https:// www.studiestoday.com

Illustration 14 : Four male employees and two female employees working in a government department are sent one by one in turns to the training centre for training. Find the probability that the two female employees go successively for the training.

Total 6 persons, 4 males and 2 females can be sent for training at the training centre one by one in ${}^6P_6 = 6! = 720$ ways. Thus, total number of outcomes will be $n = 720$.

Event of two female employees go successively for training = A.

The favourable outcomes of event A can be obtained as follows :



Considering the two female employees going successively for the training as one person, total 5 persons can be arranged in ${}^5P_5$ ways and two female employees can be arranged among themselves in ${}^2P_2$ ways in each of these arrangements.

Thus, the number favourable outcomes of event A is $m = {}^5P_5 \times {}^2P_2$

$= 5! \times 2!$

$= 120 \times 2$

$= 240$

Probability of event A $P(A) = \frac{m}{n}$

$= \frac{240}{720}$

$= \frac{1}{3}$

Required probability $= \frac{1}{3}$

Illustration 15 : Find the probability of having 53 Thursdays in a leap year.

There are 366 days in a leap year where we have 52 complete weeks (52×7=364 days) and 2 additional days. Each day appears once in each week and thus each day will appear 52 times in 52 weeks. Now, the additional 2 days can be as follows which gives the sample space for this experiment.

U = {Sunday-Monday, Monday-Tuesday, Tuesday-Wednesday, Wednesday-Thursday, Thursday-Friday, Friday-Saturday, Saturday-Sunday}

Thus, total number of outcomes will be $n = 7$.

Event A = leap year has 53 Thursdays.

Wednesday-Thursday and Thursday-Friday are the 2 favourable outcomes of event A from the above 7 outcomes. Thus, $m = 2$.

Probability of event A $\quad P(A) = \frac{m}{n}$

$= \frac{2}{7}$

Required probability $= \frac{2}{7}$

Illustration 16 : There are 2 officers, 3 clerks and 2 peons among the 7 employees working in the cash department of a bank. A committee is formed by randomly selecting two employees from the employees of this department. Find the probability that there are

(1) two peons

(2) two clerks

(3) One officer and one clerk among the two employees selected in the committee.

There are 7 employees working in the cash department of the bank. If the employees are randomly selected from them then the total number of mutually exclusive, exhaustive and equi-probable outcomes will be $n = {}^7C_2 = \frac{7 \times 6}{2 \times 1} = 21$.

(1) Event of selecting two peons $= A$

Selecting 2 peons from the 2 peons and not selecting any employee from the remaining 5 employees will be the favourable outcomes of event A.

The number of such outcomes will be $m = {}^2C_2 \times {}^5C_0 = 1 \times 1 = 1$.

Probability of event A $\quad P(A) = \frac{m}{n}$

$= \frac{1}{21}$

Required probability $= \frac{1}{21}$

(2) Event of selecting two clerks $= B$

Selecting 2 clerks from the 3 clerks and not selecting any employee from the remaining four employees will be the favourable outcomes of event B.

The number of such outcomes will be $m = {}^3C_2 \times {}^4C_0 = 3 \times 1 = 3$.

Probability of event B $\quad P(B) = \frac{m}{n}$

$= \frac{3}{21}$

$= \frac{1}{7}$

Required probability $= \frac{1}{7}$

(3) Event of selecting one officer and one clerk = C

Selecting 1 officer from 2 officers, one clerk from three clerks and not selecting any peon from two peons will be the favourable outcomes of event C.

The number of such outcomes will be $m = {}^2C_1 \times {}^3C_1 \times {}^2C_0 = 2 \times 3 \times 1 = 6$.

Probability of event C $P(C) = \frac{m}{n}$

$= \frac{6}{21}$

$= \frac{2}{7}$

Required probability $= \frac{2}{7}$

Illustration 17 : A box contains 20 items and 10% of them are defective. Three items are randomly selected from this box. Find the probability that,

(1) two items are defective

(2) two items are non-defective

(3) all three items are non-defective among the three selected items.

There are 20 items wherein 10% that is 20 × 10 % = 2 items are defective and the rest 18 are non-defective. 3 items are selected from this box of 20 items at random. Hence, the total number of outcomes in the sample space will be $n = {}^{20}C_3 = \frac{20 \times 19 \times 18}{3 \times 2} = 1140$.

(1) Event of getting two defective items among three selected items = A

Selecting 2 items from 2 defective items and selecting 1 item from the 18 non-defective items will be the favourable outcomes for the event A.

The number of such outcomes $m = {}^2C_2 \times {}^{18}C_1 = 1 \times 18 = 18$.

Probability of event A $P(A) = \frac{m}{n}$

$= \frac{18}{1140}$

$= \frac{3}{190}$

Required probability $= \frac{3}{190}$

(2) Event of getting two non-defective items among three selected items = B

Selecting 2 items from 18 non-defective items and selecting one item from 2 defective items will be the favourable outcomes of the event B.

The number of such outcomes $m = {}^{18}C_2 \times {}^2C_1 = 153 \times 2 = 306$.

Probability of event B $\quad P(B) = \frac{m}{n}$

$= \frac{306}{1140}$

$= \frac{51}{190}$

Required probability $= \frac{51}{190}$

(3) Event of getting all three non-defective items $= C$

Selecting 3 items from 18 non-defective items and not selecting any item from the defective items will be the favourable outcomes of event C.

The number of such outcomes $m = {}^{18}C_3 \times {}^{2}C_0 = 816 \times 1 = 816.$

Probability of event C $\quad P(C) = \frac{m}{n}$

$= \frac{816}{1140}$

$= \frac{68}{95}$

Required probability $= \frac{68}{95}$

Illustration 18 : A box contains 10 chits of which 3 chits are eligible for a prize. A boy named Kathan randomly selects two chits from this box. Find the probability that Kathan gets the prize.

There are 10 chits of which 3 chits are eligible for a prize and 7 chits are not eligible for prize. If two chits are randomly selected from these 10 chits then the number of mutually exclusive, exhaustive and equiprobable outcomes in the sample space will be $n = {}^{10}C_2 = \frac{10 \times 9}{2} = 45.$

Event of Kathan getting prize $= A$

$\therefore$ Event that Kathan does not get prize $= A'$

The outcomes in which Kathan will draw 2 chits at random from the 7 chits which are not eligible for prize will be the favourable outcomes of the event A'.

The number of such outcomes $m = {}^{7}C_2 = 21.$

Probability of A' $\quad P(A') = \frac{m}{n}$

$= \frac{21}{45}$

$= \frac{7}{15}$

Now $P(A) = 1 - P(A')$

$= 1 - \frac{7}{15}$

$= \frac{8}{15}$

Thus, probability that Kathan gets prize $= \frac{8}{15}$

Limitations : The limitations of the mathematical definition of probability are as follows :

(1) The probability of an event cannot be found by this definition if there are infinite outcomes in the sample space of a random experiment.

(2) The probability of an event cannot be found by this definition if the total number of outcomes in the sample space of a random experiment are not known.

(3) The probability of an event cannot be found by this definition if the elementary outcomes in the sample space of a random experiment are not equi-probable.

(4) The word 'equi-probable' is mentioned in the mathematical definition of probability. Equi-probable events are the events with same probability. Thus, the word probability is used in the definition of probability.

Exercise 1.2

1. A balanced coin is tossed three times. Find the probability of the following events :

(1) Getting all three heads
(2) Not getting a single head
(3) Getting at least one head
(4) Getting more than one head
(5) Getting at the most one head
(6) Getting less than two heads
(7) Getting head and tail alternately
(8) Getting more number of heads than tails

2. Two balanced dice are thrown simultaneously. Find the probability of the following events :

(1) The sum of numbers on the dice in 6
(2) The sum of numbers on the dice is not more than 10
(3) The sum of numbers on the dice is a multiple of 3
(4) The product of numbers on the dice is 12

3. One family is randomly selected from the families having two children. Find the probability that

(1) One child is a girl and one child is a boy.
(2) At least one child is a girl among the two children of the selected family.
(**Note :** Assume that the chance of the child being a boy or girl is same.)

4. One number is selected at random from the first 100 natural numbers. Find the probability that this number is divisible by 7.

5. The sample space for a random experiment of selecting numbers is $U = \{1, 2, 3, ..., 120\}$ and all the outcomes in the sample space are equiprobable. Find the probability that the number selected is

(1) a multiple of 3
(2) not a multiple of 3
(3) a multiple of 4
(4) not a multiple of 4
(5) a multiple of both 3 and 4.

6. Find the probability of getting R in the first place and M in the last place when all the letters of the word *RANDOM* are arranged in all possible ways.

7. Find the probability of getting vowels in the first, third and sixth place when all the letters of the word *ORANGE* are arranged in all possible ways.

8. Five members of a family, husband, wife and three children, are randomly arranged in a row for a family photograph. Find the probability that the husband and wife are seated next to each other.

9. Seven speakers A, B, C, D, E, F, G are invited in a programme to deliver speech in random order. Find the probability that speaker B delivers speech immediately after speeker A.

10. Find the probability of having 5 Mondays in the month of February of a leap year.

11. Find the probability of having 53 Fridays in a year which is not a leap year.

12. Find the probability of having 5 Tuesdays in the month of August of any year.

13. 4 couples (husband-wife) attend a party. Two persons are randomly selected from these 8 persons. Find the probability that the selected persons are,

(1) husband and wife (2) one man and one woman

(3) one man and one woman who are not husband and wife.

14. 8 workers are employed in a factory and 3 of them are excellent in efficiency where as the rest of them are moderate in efficiency. 2 workers are randomly selected from these 8 workers. Find the probability that,

(1) both the workers have excellent efficiency

(2) both the workers have moderate efficiency

(3) one worker is excellent and one worker is moderate in efficiency.

15. Two cards are drawn from a well shuffled pack of 52 cards. Find the probability that,

(1) both the cards are of different colour (2) both the cards are face cards

(3) one of the two cards is a king.

16. 3 bulbs are defective in a box of 10 bulbs. 2 bulbs are randomly selected from this box. These bulbs are fixed in two bulb-holders installed in a room. Find the probability that the room will be lighted after starting the electric supply.

17. For two events A and B in the sample space of a random experiment, $P(A) = 0.6$, $P(B) = 0.5$ and $P(A \cap B) = 0.15$. Find

(1) $P(A')$ (2) $P(B - A)$ (3) $P(A \cap B')$ (4) $P(A' \cap B')$ (5) $P(A' \cup B')$

18. For two events A and B in the sample space of a random experiment, $P(A') = 2P(B') = 3P(A \cap B) = 0.6$. Find the probability of difference events $A - B$ and $B - A$.

*

Downloaded from https:// www.studiestoday.com

1.5 Law of Addition of Probability

The rule of obtaining the probability of the occurrence of at least one of the event *A* and *B* in the sample space of a random experiment is called the law of addition of probability. We have seen earlier that the occurrence of at least one of the events *A* and *B* is denoted by $A \cup B$, the union of events *A* and *B*. Hence we can say that the law of addition of probability is the rule of obtaining the probability of $A \cup B$, the union of events *A* and *B*. This rule is stated as follows and we will accept it without proof :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The law of addition of probability can also be used for obtaining the probability of union of more than two events. The law of addition of probability for $A \cup B \cup C$, the union of three events *A, B* and *C* is as follows :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Some of the important results obtained from this rule are as follows :

(1) If the events *A* and *B* in the sample space of a random experiment are mutually exclusive then $A \cap B = \phi$ and $P(A \cap B) = 0$. Hence,

$$P(A \cup B) = P(A) + P(B)$$

(2) If three events *A, B* and *C* in the sample space of a random experiment are mutually exclusive then,

$A \cap B = \phi$, $A \cap C = \phi$, $B \cap C = \phi$, $A \cap B \cap C = \phi$ and

$P(A \cap B) = P(A \cap C) = P(B \cap C) = P(A \cap B \cap C) = 0$. Hence,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

(3) If two events *A* and *B* in the sample space of a random experiment are mutually exclusive and exhaustive then $A \cap B = \phi$ and $A \cup B = U$. As $P(\phi) = 0$ and $P(U) = 1$, $P(A \cap B) = 0$ and $P(A \cup B) = 1$.

$$P(A \cup B) = P(A) + P(B) = 1$$

(4) If three events *A, B* and *C* in the sample space of a random experiment are mutually exclusive and exhaustive then,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) = 1$$

Illustration 19 : A number is randomly selected from the first 50 natural numbers. Find the probability that it is a multiple of 2 or 3.

If one number is randomly selected from the first 50 natural numbers then the number of mutually exclusive, exhaustive and equi-probable outcomes in the sample space of this random experiment will be $n = {}^{50}C_1 = 50$.

If event *A* denotes that the number selected is a multiple of 2 and event *B* denotes that the number selected is a multiple of 3 then the event that the selected number is a multiple of 2 or 3 will be denoted by $A \cup B$. (This event can also be denoted as $B \cup A$. According to set theory, $A \cup B = B \cup A$). To find the probability of $A \cup B$, the union of events *A* and *B* by the law of addition of probability, we will first find $P(A), P(B)$ and $P(A \cap B)$.

A = Event that the selected number is a multiple of 2
= {2, 4, 6,..., 50}

Hence, the number of favourable outcomes of event A will be m = 25.

Probability of event A $P(A) = \frac{m}{n}$

$= \frac{25}{50}$

B = Event that the selected number is a multiple of 3
= {3, 6, 9,..., 48}

Hence, the number of favourable outcomes of event B will be m = 16.

Probability of event B $P(B) = \frac{m}{n}$

$= \frac{16}{50}$

$A \cap B =$ Event that the selected number is a multiple of 2 and 3 that is multiple the LCM of 2 and 3 which is 6.

= {6, 12, 18,..., 48}

Hence, the number of favourable outcomes of event $A \cap B$ will be m = 8.

Probability of event $A \cap B$ $P(A \cap B) = \frac{m}{n}$

$= \frac{8}{50}$

From the law of addition of probability,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= \frac{25}{50} + \frac{16}{50} - \frac{8}{50}$$

$$= \frac{25 + 16 - 8}{50}$$

$$= \frac{33}{50}$$

Required probability $= \frac{33}{50}$

Illustration 20 : One card is randomly selected from a pack of 52 cards. Find the probability that the selected card is

(1) club or queen card

(2) neither a club nor a queen card.

If one card is randomly selected from a pack of 52 cards then the number of mutually exclusive, exhaustive and equi-probable outcomes in the sample space of this random experiment $n = {}^{52}C_1 = 52$.

Event that the selected card is a club card = A

Event that the selected card is a queen = B

(1) Event that the selected card is club or queen card = $A \cup B$

To find the probability of event $A \cup B$ by the law of addition of probability, we will first find $P(A), P(B)$ and $P(A \cap B)$.

A = Event that the selected card is club card.

There are 13 club cards in a pack of 52 cards. Thus, the number of favourable outcomes of event A is $m = 13$.

Probability of event A $\quad P(A) = \frac{m}{n}$

$$= \frac{13}{52}$$

$B =$ Event that the selected card is a queen card.

There are 4 queen cards in a pack of 52 cards. Thus, the number of favourable outcomes of event B is $m = 4$.

Probability of event B $\quad P(B) = \frac{m}{n}$

$$= \frac{4}{52}$$

$A \cap B =$ Event that the selected card is club and queen card that is a club queen.

There is only 1 card in the pack of 52 cards which is club queen. Hence, the number of favourable outcomes of $A \cap B$ is $m = 1$.

Probability of $A \cap B$ $\quad P(A \cap B) = \frac{m}{n}$

$$= \frac{1}{52}$$

From the law of addition of probability,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$$

$$= \frac{13 + 4 - 1}{52}$$

$$= \frac{16}{52}$$

$$= \frac{4}{13}$$

Required probability $= \frac{4}{13}$

Event $A \cup B$ can be easily explained by the following diagram :

Suit	**Type of Card**												
	A	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**J**	**Q**	**K**
♠	♠A	♠2	♠3	♠4	♠5	♠6	♠7	♠8	♠9	♠10	♠J	♠Q	♠K
♦	♦A	♦2	♦3	♦4	♦5	♦6	♦7	♦8	♦9	♦10	♦J	♦Q	♦K
♣	♣A	♣2	♣3	♣4	♣5	♣6	♣7	♣8	♣9	♣10	♣J	♣Q	♣K
♥	♥A	♥2	♥3	♥4	♥5	♥6	♥7	♥8	♥9	♥10	♥J	♥Q	♥K

(2) Event that the selected card is not of club = A'

Event that the selected card is not queen = B'

Hence, the event that the selected card is neither club nor queen is $A' \cap B'$

Thus, the probability of $A' \cap B'$

$$P(A' \cap B') = P(A \cup B)'$$

$$= 1 - P(A \cup B)$$

$$= 1 - \frac{4}{13}$$

$$= \frac{9}{13}$$

Required probability = $\frac{9}{13}$

Illustration 21 : 3 persons from medical profession and 5 persons from engineering profession offer services at a social organization. 2 persons are randomly selected from these persons with the purpose of forming a committee. Find the probability that both the persons selected belong to the same profession.

There are in all $3 + 5 = 8$ persons. Hence, 2 persons can be selected in ${}^{8}C_{2} = \frac{8 \times 7}{2 \times 1} = 28$ ways.

Thus, the total number of mutually exclusive, exhaustive and equi-probable outcomes in the sample space is $n = 28$.

Event that both the persons selected belong to medical profession = A

Even that both the persons selected belong to the engineering profession = B

Event that both the persons selected belong to the same profession = $A \cup B$

The two events A and B can not occur together that is $A \cap B = \phi$

Thus, the events A and B are mutually exclusive. Hence, from the law of addition of probability,

$$P(A \cup B) = P(A) + P(B)$$

For which we first find $P(A)$ and $P(B)$.

A = Event that both the persons selected belong to medical profession.

The number of favourable outcomes of A is $m = {}^{3}C_{2} = 3$.

Probability of event A $P(A) = \frac{m}{n}$

$$= \frac{3}{28}$$

B = Event that both the persons selected belong to engineering profession.

The number of favourable outcomes of B is $m = {}^{5}C_{2} = 10$.

Probability of event B $P(B) = \frac{m}{n}$

$$= \frac{10}{28}$$

Now,

$$P(A \cup B) = P(A) + P(B)$$

$$= \frac{3}{28} + \frac{10}{28}$$

$$= \frac{3+10}{28}$$

$$= \frac{13}{28}$$

Required probability $= \frac{13}{28}$

Illustration 22 : The probability that a person from a group reads newspaper X is 0.55, the probability that he read newspaper Y is 0.69 and the probability that he reads both the newspaper X and Y is 0.27. Find the probability that a person selected at random from this group.

(1) reads at least one of the newspapers X and Y.

(2) does not read any of the newspapers X and Y.

(3) reads only one of the newspapers X and Y.

If the event that a person from the group reads newspaper X is denoted by event A and reads newspaper Y by event B then the given information can be shown as follows :

$P(A) = 0.55,\ P(B) = 0.69,\ P(A \cap B) = 0.27$

(1) Event that the selected person reads at least one of the newspapers $= A \cup B$

From the law of addition of probability,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= 0.55 + 0.69 - 0.27$$

$$= 0.97$$

Required probability = 0.97

(2) Event that the selected person does not read newspaper $A = A'$

Event that the selected person does not read newspaper $B = B'$

Hence, event that the selected person does not read any of the newspaper X and $Y = A' \cap B'$.

Probability of $A' \cap B'$

$$P(A' \cap B') = P(A \cup B)'$$

$$= 1 - P(A \cup B)$$

$$= 1 - 0.97$$

$$= 0.03$$

Required probability = 0.03

(3) If the event that the selected person reads only one of the newspapers X and Y is denoted by C then the event C can occur as follows :

The person reads newspaper X (event A) and does not read newspaper Y (event B')

OR

The person does not read newspaper X (event A') and reads newspaper Y (event B)

Thus $C = (A \cap B') \cup (A' \cap B)$

Since the events $A \cap B'$ and $A' \cap B$ are mutually exclusive,

$$P(C) = P(A \cap B') + P(A' \cap B)$$
$$= [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)]$$
$$= [0.55 - 0.27] + [0.69 - 0.27]$$
$$= 0.28 + 0.42$$
$$= 0.7$$

Required probability $= 0.7$

Illustration 23 : For two events A and B in the sample space of a random experiment $P(A) = 2P(B) = 4P(A \cap B) = 0.6$. Find the probability of the following events :

(1) $A' \cap B'$ (2) $A' \cup B'$ (3) $A - B$ (4) $B - A$

It is given that $P(A) = 2P(B) = 4P(A \cap B) = 0.6$. Hence,

$P(A) = 0.6$	$2P(B) = 0.6$	$4P(A \cap B) = 0.6$
	$\therefore P(B) = 0.3$	$\therefore P(A \cap B) = 0.15$

(1) Probability of event $A' \cap B'$ $= P(A' \cap B')$

$$= P(A \cup B)'$$
$$= 1 - P(A \cup B)$$
$$= 1 - [P(A) + P(B) - P(A \cap B)]$$
$$= 1 - [0.6 + 0.3 - 0.15]$$
$$= 1 - 0.75$$
$$= 0.25$$

Required probability $= 0.25$

(2) Probability of event $A' \cup B'$ $= P(A' \cup B')$

$$= P(A \cap B)'$$
$$= 1 - P(A \cap B)$$
$$= 1 - 0.15$$
$$= 0.85$$

Required probability $= 0.85$

(3) Probability of event $A-B$ $= P(A-B)$

$= P(A) - P(A \cap B)$

$= 0.6 - 0.15$

$= 0.45$

Required probability $= 0.45$

(4) Probability of event $B-A$ $= P(B-A)$

$= P(B) - P(A \cap B)$

$= 0.3 - 0.15$

$= 0.15$

Required probability $= 0.15$

Illustration 24 : For two events A and B in the sample space of a random experiment $P(A')=0.3$, $P(B)=0.6$ and $P(A \cup B)=0.83$. Find $P(A \cap B')$ and $P(A' \cap B)$.

Here, $P(A')=0.3 \therefore P(A)=1-P(A')=1-0.3=0.7$

$P(B)=0.6$ and $P(A \cup B)=0.83$

First we will find $P(A \cap B)$.

$P(A \cup B)=P(A)+P(B)-P(A \cap B)$

$\therefore\ 0.83=0.7+0.6-P(A \cap B)$

$\therefore\ P(A \cap B)=0.7+0.6-0.83$

$\therefore\ P(A \cap B)=0.47$

Now,

$P(A \cap B')=P(A)-P(A \cap B)$

$= 0.7 - 0.47$

$= 0.23$

Required probability $= 0.23$

$P(A' \cap B)=P(B)-P(A \cap B)$

$= 0.6 - 0.47$

$= 0.13$

Required probability $= 0.13$

Illustration 25 : Two events A and B in the sample space of a random experiment are mutually exclusive. If $3P(A)=4P(B)=1$ then find $P(A\cup B)$.

Since $3P(A)=4P(B)=1$

$3P(A)=1$	$4P(B)=1$
$\therefore\ P(A)=\frac{1}{3}$	$\therefore\ P(B)=\frac{1}{4}$

As the events A and B are mutually exclusive $(A\cap B=\phi)$,

$$P(A\cup B)=P(A)+P(B)$$

$$=\frac{1}{3}+\frac{1}{4}$$

$$=\frac{7}{12}$$

Required probability $=\frac{7}{12}$

Illustration 26 : For three mutually exclusive and exhaustive events A, B and C in the sample space of a random experiment $2P(A)=3P(B)=4P(C)$. Find $P(A\cup B)$ and $P(B\cup C)$.

Taking $2P(A)=3P(B)=4P(C)=x$,

$2P(A)=x$	$3P(B)=x$	$4P(C)=x$
$\therefore\ P(A)=\frac{x}{2}$	$\therefore\ P(B)=\frac{x}{3}$	$\therefore\ P(C)=\frac{x}{4}$

Since A, B and C are mutually exclusive and exhaustive events,

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)=1$$

$$\therefore\ \frac{x}{2}+\frac{x}{3}+\frac{x}{4}=1$$

$$\therefore\ \frac{6x+4x+3x}{12}=1$$

$$\therefore\ 13x=12$$

$$\therefore\ x=\frac{12}{13}$$

Thus,

$$P(A)=\frac{x}{2}=\frac{\frac{12}{13}}{2}=\frac{6}{13}$$

$$P(B)=\frac{x}{3}=\frac{\frac{12}{13}}{3}=\frac{4}{13}$$

$$P(C)=\frac{x}{4}=\frac{\frac{12}{13}}{4}=\frac{3}{13}$$

Now, the probability of required events,

$$P(A \cup B) = P(A) + P(B)$$

$$= \frac{6}{13} + \frac{4}{13}$$

$$= \frac{10}{13}$$

Required probability $= \frac{10}{13}$

$$P(B \cup C) = P(B) + P(C)$$

$$= \frac{4}{13} + \frac{3}{13}$$

$$= \frac{7}{13}$$

Required probability $= \frac{7}{13}$

Exercise 1.3

1. 2 cards are drawn from a pack of 52 cards. Find the probability that both the cards drawn are

(1) of the same suit

(2) of the same colour.

2. 3 books of Statistics and 4 of Mathematics are arranged on a shelf. Two books are randomly selected from these books. Find the probability that both the books selected are of the same subject.

3. One card is randomly drawn from a pack of 52 cards. Find the probability that it is

(1) Spade card or ace (2) Neighter spade nor ace.

4. A number is selected from the natural number 1 to 100. Find the probability of the event that the selected number is a multiple of 3 or 5.

5. Two balanced dice are thrown simultaneously. Find the probability that the sum of numbers on two dice is a multiple of 2 or 3.

6. The probability that the price of potato rises in the vegetable market during festive days in 0.8. The probability that the price of onion rises is 0.7. The probability of rise in price of both potato and onion is 0.6. Find the probability of rise in price of at least one of the two, potato and onion.

7. Two aircrafts drop bomb to destroy a bridge. The probability that a bomb dropped from the first aircraft hits the target is 0.9 and the probability that a bomb from the second aircraft hits the target is 0.7. The probability of one bomb dropped from both the aircrafts hitting the target is 0.63. The bridge is destroyed even if one bomb drops on it. Find the probability that the bridge is destroyed.

8. The probability that a teenager coming to a restaurant for dinner orders pizza is 0.63. The probability of ordering cold-drink is 0.54. The probability that the teenager orders at least one out of pizza and cold-drink is 0.88. Find the probability that the teenager coming for dinner on a certain day orders only one of the two items from pizza and cold-drink.

9. If A and B are mutually exclusive and exhaustive events in a sample space U and $P(A)=2P(B)$ then find $P(A)$.

10. Three events A, B and C in a sample space are mutually exclusive and exhaustive. If $4P(A)=5P(B)=3P(C)$ then find $P(A\cup C)$ and $P(B\cup C)$.

11. Find $P(A\cup B\cup C)$ using the following information about three events A, B and C in a sample space.

$P(A)=0.65$, $P(B)=0.45$, $P(C)=0.25$, $P(A\cap B)=0.25$, $P(A\cap C)=0.15$, $P(B\cap C)=0.2$, $P(A\cap B\cap C)=0.05$

12. Three events A, B and C in a sample space are mutually exclusive and exhaustive. If $P(C')=0.8$ and $3P(B)=2P(A')$ then find $P(A)$ and $P(B)$.

*

1.6 Conditional Probability and Law of Multiplication of Probability

1.6.1 Conditional Probability

Suppose U is a finite sample space and A and B are any two events in it. The probability of occurrence of event B under the condition that A occurs is called the conditional probability. If the occurrence of event B under the condition that event A occurs is denoted by B/A then the probability $P(B/A)$ of the conditional event B/A is called the conditional probability. This probability is obtained using the following formula :

$$P(B/A) = \frac{P(A\cap B)}{P(A)}; \quad P(A)\neq 0$$

Similarly, if the occurrence of event A under the condition that event B occurs is denoted by A/B then the probability $P(A/B)$ of the conditional event A/B is obtained using the following formula :

$$P(A/B) = \frac{P(A\cap B)}{P(B)}; \quad P(B)\neq 0$$

Suppose a company produces a certain type of item in its two different factories A_1 and A_2. One item is randomly selected from a store selling the items produced by this company. Let us denote the event that the selected item is defective as D.

- If the selected item is produced at factory A_1 then the event that it is defective is denoted by D/A_1.
- If the selected item is produced at factory A_2 then the event that it is defective is denoted by D/A_2.

Thus,

$P(D/A_1)$ = Probability of occurrence of D under the condition that event A_1 has occurred and

$P(D/A_2)$ = Probability of occurrence of D under the condition that event A_2 has occurred

1.6.2 Independent Events

Suppose A and B are any two events in a finite sample space U. If the probability of occurrence of event A does not change due to occurrence (or non-occurrence) of event B then the events A and B are called the independent events.

Thus, if $P(A) = P(A/B) = P(A/B')$ and $P(B) = P(B/A) = P(B/A')$ the events A and B are called independent events.

For example, Event A = First throw of a balanced die shows number 1.

Event B = Second throw of a balanced die shows an even number.

It can be said here that the probability of getting an even number in the second throw of the die does not change because the first throw had shown the number 1. This fact can be easily understood by the following calculation :

The total number of outcomes by throwing the dice two times is $n = 6 \times 6 = 36$.

A = Event that the first throw of a balanced die shows the number 1.

The number of favourable outcomes of A is $m = 6$.

Probability of event A $P(A) = \frac{m}{n}$

$= \frac{6}{36}$

B = Event that the second throw of balanced die shows an even number.

The number of favourable outcomes of B is $m = 18$

Probability of event B $P(B) = \frac{m}{n}$

$= \frac{18}{36}$

$= \frac{1}{2}$

$A \cap B$ = Event that the first throw of a balanced die shows the number 1 and the second throw shows even number.

The number of favourable outcomes of $A \cap B$ is $m = 3$.

Probability of event $A \cap B$ $P(A \cap B) = \frac{m}{n}$

$= \frac{3}{36}$

Now, if the first throw of the die shows number 1 then the probability $P(B/A)$ for the event B/A of getting an even number in the second throw can be obtained as follows :

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{\frac{3}{36}}{\frac{6}{36}}$$

$$= \frac{1}{2}$$

Since we get $P(B) = P(B/A)$, we say that the events A and B are independent.

1.6.3 Law of Multiplication of Probability

If A and B are the two events in a sample space U then the rule of obtaining the probability of simultaneous occurrence of events A and B is called the law of multiplication of probability.

For example, Event A = Getting head when a coin is tossed for the first time.

Event B = Getting head when a coin is tossed for the second time.

If the coin is tossed two times then the probability of getting head both the times that is event $A \cap B$ can be obtained by the law of multiplication of probability. The law of multiplication of probability is as follows :

$P(A \cap B) = P(A) \times P(B/A); \; P(A) \neq 0$

$P(A \cap B) = P(B) \times P(A/B); \; P(B) \neq 0$

Some of the important results deduced from this rule are as follows which will be accepted without proof.

(1) If A and B are independent events then $P(A \cap B) = P(A) \times P(B)$

(2) If A and B are independent events then

(i) The events A' and B' are also independent. Hence, $P(A' \cap B') = P(A') \times P(B')$

(ii) The events A and B' are also independent. Hence, $P(A \cap B') = P(A) \times P(B')$

(iii) The events A' and B are also independent. Hence, $P(A' \cap B) = P(A') \times P(B)$

1.6.4 Selection with Replacement and without Replacement

When the units are to be randomly selected one by one from the population, the selection can be done in two ways :

(1) **Selection with replacement :** If the selection of a unit from the population in any trial is done by replacing the unit selected in the previous trial back to the population then that selection is called the selection with replacement.

(2) **Selection without Replacement :** If the selection of a unit from the population in any trial is done by not replacing the unit selected in the previous trial back to the population then that selection is called the selection with replacement.

Illustration 27 : A balanced coin is tossed twice. If the first toss of the coin shows head then find the probability of getting head in both the tosses.

The sample space of the random experiment of tossing a balanced coin twice is $U = \{HH, HT, TH, TT\}$, where the first symbol shows the outcome of the first toss of the coin and the second symbol shows the outcome of the second toss of the coin. The total number of outcomes in this sample space is n = 4.

If A denote the event of getting head in the first toss of the coin and B denotes the event that both the tosses result in head then we have to find $P(B/A)$, probability of B/A.

Event A = First toss shows head

= $\{HH, HT\}$

Hence, the number of favourable outcomes of A is $m = 2$.

Probability of event A $P(A) = \frac{m}{n}$

$= \frac{2}{4}$

Event B = Head is shown in both the tosses

= $\{HH\}$

Hence, the number of favourable outcomes of B is $m = 1$.

Probability of event B $P(B) = \frac{m}{n}$

$= \frac{1}{4}$

Event $A \cap B$ = Getting head in the first toss and getting head in both the tosses of the coin (we have $B \subset A$.)

= $\{HH\}$

Hence, the number of favourable outcomes of $A \cap B$ is $m = 1$.

Probability of $A \cap B$ $P(A \cap B) = \frac{m}{n}$

$= \frac{1}{4}$

Now,

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{\frac{1}{4}}{\frac{2}{4}}$$

$$= \frac{1}{2}$$

Required probability $= \frac{1}{2}$

Illustration 28 : A factory has received an order to prepare 50,000 units of an item in a certain time period. The probability of completing this work in the given time is 0.75 and the probability that the workers will not declare strike during that time period is 0.8. The probability that this work will be completed during the given period and the workers will not declare strike is 0.7. Find the probability that

(1) The work will be completed as per schedule under the condition that the workers have not declared strike.

(2) Find the probability that the workers do not declare strike in the given period knowing that the work is completed as per schedule.

If we denote event A that the work will be completed as per schedule and event B that the workers will not declare strike then the given information can be written as follows :

$P(A) = 0.75,\ P(B) = 0.8,\ P(A \cap B) = 0.7$

(1) Event that the work will be completed in the given period under the condition that the workers do not declare strike = A/B

Probability of A/B from the definition of condition probability,

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{0.7}{0.8}$$

$$= \frac{7}{8}$$

Required probability = $\frac{7}{8}$

(2) If it is given that the work is completed as per schedule then the event that the workers do not declare strike = B/A

Probability of B/A from the definition of conditional probability,

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{0.7}{0.75}$$

$$= \frac{14}{15}$$

Required probability = $\frac{14}{15}$

Illustration 29 : If $P(A') = \frac{7}{25}$, $P(B/A) = \frac{5}{12}$ and $P(A/B) = \frac{1}{2}$ for two events A and B in the sample space of a random experiment then find $P(A \cap B)$ and $P(B)$.

It is given that $P(A') = \frac{7}{25}$, $P(B/A) = \frac{5}{12}$ and $P(A/B) = \frac{1}{2}$.

$$P(A) = 1 - P(A')$$

$$= 1 - \frac{7}{25}$$

$$= \frac{18}{25}$$

We will find $P(A \cap B)$ from the formula of $P(B/A)$.

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$$\therefore \quad \frac{5}{12} = \frac{P(A \cap B)}{\frac{18}{25}}$$

$$\therefore \quad \frac{5}{12} \times \frac{18}{25} = P(A \cap B)$$

$$\therefore \quad P(A \cap B) = \frac{3}{10}$$

Required probability $= \frac{3}{10}$

Now, we will find $P(B)$ by substituting $P(A/B)$ and $P(A \cap B)$ in the formula of $P(A/B)$.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$\therefore \quad \frac{1}{2} = \frac{\frac{3}{10}}{P(B)}$$

$$\therefore \quad P(B) = \frac{\frac{3}{10}}{\frac{1}{2}}$$

$$= \frac{3 \times 2}{10 \times 1}$$

$$= \frac{3}{5}$$

Required probability $= \frac{3}{5}$

Illustration 30 : A medicine is tested on a group of rabbits and mice to know its effect. It was observed that 7 rabbits show the effect of medicine in a group of 10 rabbits who were given the medicine and 5 mice show the effect of medicine in a group of 7 mice who were given the medicine. One animal is selected at random from each group. Find the probability that (1) both the selected animals show the effect of medicine and (2) one of the two selected animals shows the effect of medicine and the other animal does not show the effect of medicine.

The given information will be shown as follows :

Animals affected by Medicine	**Animals not affected by Medicine**
Rabbits 7	Rabbits 3
Mice 5	Mice 4
Total 12	**Total 7**

(1) Event that a rabbit shows effect of medicine $= A$

Event that a mouse shows effect of medicine $= B$

$\therefore$ Event that both the animals show the effect of medicine $= A \cap B$

The events A and B are independent. Whether the mice show the effect of medicine is not affected by the effect of medicine on rabbits. Hence,

$$P(A \cap B) = P(A) \times P(B)$$

From the total number of outcomes $n = 10$, $m = 7$ outcomes are favourable for event A.

$\therefore$ Probability of event A $P(A) = \frac{m}{n}$

$= \frac{7}{10}$

From the total number of outcomes $n = 9$, $m = 5$ outcomes are favourable for event B.

$\therefore$ Probability of event B $P(B) = \frac{m}{n}$

$= \frac{5}{9}$

$\therefore P(A \cap B) = \frac{7}{10} \times \frac{5}{9}$

$= \frac{7}{18}$

Required probability $= \frac{7}{18}$

(2) Let C denote the event that one animal is affected by the medicine and the other animal is not affected by the medicine. The event C can occur as follows :

Rabbit is affected by the medicine (event A) and mouse is not affected by the medicine (event B')

OR

Rabbit is not affected by the medicine (event A') and mouse is affected by the medicine (event B)

Thus, event $C = (A \cap B') \cup (A' \cap B)$

Since the events $A \cap B'$ and $A' \cap B$ are mutually exclusive,

$P(C) = P(A \cap B') + P(A' \cap B)$

$= [P(A) \times P(B')] + [P(A') \times P(B)]$ ($\because$ A and B are independent events)

Here, $P(A') = 1 - P(A)$

$= 1 - \frac{7}{10}$

$= \frac{3}{10}$

$P(B') = 1 - P(B)$

$= 1 - \frac{5}{9}$

$= \frac{4}{9}$

$\therefore$ $P(C) = [\frac{7}{10} \times \frac{4}{9}] + [\frac{3}{10} \times \frac{5}{9}]$

$= \frac{28}{90} + \frac{15}{90}$

$= \frac{43}{90}$

Required probability $= \frac{43}{90}$

Downloaded from https:// www.studiestoday.com

Illustration 31 : A company produces a certain type of item in its two different factories A_1 and A_2 in the proportion 60% and 40% respectively. The proportions of defectives in the production of these factories are 2% and 3% respectively. One item is randomly selected after mixing the items produced in the two factories. Find the probability that this item is defective.

Event that the selected item is produced in factory A_1 = A_1

$$\therefore \quad P(A_1) = \frac{60}{100}$$

$$= \frac{3}{5}$$

Event that the selected item is produced in factory A_2 = A_2

$$\therefore \quad P(A_2) = \frac{40}{100}$$

$$= \frac{2}{5}$$

Let D denote the event that the item selected from the total production is defective.

Event that the selected item is defective when it is produced in factory A_1 = D/A_1

$$\therefore \quad P(D/A_1) = \frac{2}{100}$$

$$= \frac{1}{50}$$

Event that the selected item is defective when it is produced in factory A_2 = D/A_2

$$\therefore \quad P(D/A_2) = \frac{3}{100}$$

Event D can occur as follows.

The selected item is produced in factory A_1 and it is defective.

OR

The selected item is produced in factory A_2 and it is defective.

Thus event $D = (A_1 \cap D) \cup (A_2 \cap D)$

Since the events $A_1 \cap D$ and $A_2 \cap D$ are mutually exclusive,

$$P(D) = P(A_1 \cap D) + P(A_2 \cap D)$$

$$= \left[P(A_1) \times P(D/A_1)\right] + \left[P(A_2) \times P(D/A_2)\right]$$

$$= \left[\frac{3}{5} \times \frac{1}{50}\right] + \left[\frac{2}{5} \times \frac{3}{100}\right]$$

$= \frac{3}{250} + \frac{6}{500}$

$= \frac{12}{500}$

$= \frac{3}{125}$

Required probability $= \frac{3}{125}$

Illustration 32 : There are 12 screws in a box of which 4 screws are defective. Two screws are randomly selected one by one without replacement from this box. Find the probability that both the screws selected are defective.

4 screws are defective in the box having 12 screws. Hence, the number of non-defective screws will be 8.

Total number of mutually exclusive, exhaustive and equiprobable outcomes for selecting the first screw are $n = {}^{12}C_1 = 12$.

If A denotes the event that the first screw selected is defective then the number of favourable outcomes of A is $m = {}^{4}C_1 = 4$.

Probability of event A $P(A) = \frac{m}{n} = \frac{4}{12}$

The screws are selected without replacement which means that the first screw is not kept back into the box. Hence, the total number of mutually exclusive, exhaustive and equi-probable outcomes for selecting the second screw is $n = {}^{11}C_1 = 11$.

Let B denote the event that the second screw selected is defective.

The event B occurs under the condition that the event A has occurred. This is the occurrence of event B/A.

Since event A has occurred earlier, there are 3 defective screws in the box.

Hence, the number of favourable outcomes for event B/A is $m = {}^{3}C_1 = 3$.

Probability of B/A $P(B/A) = \frac{m}{n}$

$= \frac{3}{11}$

Now, $A \cap B$ = Event that both the screws are defective

From the law of multiplication of probability,

$P(A \cap B) = P(A) \times P(B/A)$

$= \frac{4}{12} \times \frac{3}{11}$

$= \frac{1}{11}$

Required probability $= \frac{1}{11}$

Illustration 33 : There are 3 boys and 2 girls in a friend-circle. Two persons are randomly selected from this friend-circle one by one with replacement to sing a song. Find the probability that the first person is a boy and the second person is a girl in the two persons selected to sing a song.

The friend-circle consists of 3 boys and 2 girls that is total 5 persons. Two persons are selected one by one with replacement. This means that the person selected first is sent back to the group before selecting the second person. Hence, the events of selecting two persons one by one are independent events. The total number of mutually exclusive, exhaustive and equi-probable outcomes for selecting the first person is $n = {}^5C_1 = 5$.

Event that the first person selected to sing a song is a boy $= A$

The number of favourable outcomes for event A is $m = {}^3C_1 = 3$

Probability of event A $P(A) = \frac{m}{n}$

$= \frac{3}{5}$

The selection is with replacement here. This means that the total number of mutually exclusive, exhaustive and equi-probable outcomes for selecting the second person is $n = {}^5C_1$.

Event that the second person selected to sing a song is a girl $= B$

The number of favourable outcomes of B is $m = {}^2C_1 = 2$

Probability of event B $P(B) = \frac{m}{n}$

$= \frac{2}{5}$

Now, $A \cap B$ = Event that the first boy and the second girl are the two person selected to sing a song. Since the events A and B are independent,

$P(A \cap B) = P(A) \times P(B)$

$= \frac{3}{5} \times \frac{2}{5}$

$= \frac{6}{25}$

Required probability $= \frac{6}{25}$

Illustration 34 : Two balanced dice are thrown simultaneously. Find the probability that at least one of the two dice shows the number 3.

Event that the first die shows number 3 $= A$

Event that the second die shows number 3 $= B$

Event that at least one die shows number 3 $= A \cup B$

The number of favourable outcome for event A is $m=1$

Probability of event A $P(A)=\frac{m}{n}$

$=\frac{1}{6}$

The number of favourable outcomes for event B is $m=1$

Probability of event B $P(B)=\frac{m}{n}$

$=\frac{1}{6}$

Since the events A and B are independent, the events A' and B' are also independent. Moreover, $P(A')=1-P(A)=1-\frac{1}{6}=\frac{5}{6}$ and $P(B')=1-P(B)=1-\frac{1}{6}=\frac{5}{6}$.

Probability of the event that at least one die shows number 3 $=P(A\cup B)$

$=1-P(A'\cap B')$

$=1-\left[P(A')\times P(B')\right]$

$=1-\left[\frac{5}{6}\times\frac{5}{6}\right]$

$=1-\frac{25}{36}$

$=\frac{11}{36}$

Required probability $=\frac{11}{36}$

Illustration 35 : Two cities A and B of different states have rains on 60% and 75% days respectively during the monsoon. For the cities A and B, find the probability that on a certain monsoon day,

(1) both the cities have rains

(2) at least one city has rains

(3) only one city has rains.

Note : The events of rains on a day in these two cities are independent.

Let event A denote that it rains in city A and event B denote that it rains in city B. The given information can be stated as follows :

$P(A)=\frac{60}{100}=\frac{3}{5}$ $\quad\therefore P(A')=1-P(A)=1-\frac{3}{5}=\frac{2}{5}$

$P(B)=\frac{75}{100}=\frac{3}{4}$ $\quad\therefore P(B')=1-P(B)=1-\frac{3}{4}=\frac{1}{4}$

(1) Event that both the cities A and B have rains $A\cap B$

Since the events A and B are independent,

Probability of event $A\cap B$ $\quad P(A\cap B)=P(A)\times P(B)$

$=\frac{3}{5}\times\frac{3}{4}$

$=\frac{9}{20}$

Required probability $=\frac{9}{20}$

(2) Event that at least one of the cities A and B has rains = $A \cup B$

Probability of $A \cup B$ $\quad P(A \cup B) = 1 - P(A' \cap B')$

$$= 1 - [P(A') \times P(B')]$$

$$= 1 - \left[\frac{2}{5} \times \frac{1}{4}\right]$$

$$= 1 - \frac{1}{10}$$

$$= \frac{9}{10}$$

Required probability = $\frac{9}{10}$

(3) Event that only one of cities A and B has rains $= (A \cap B') \cup (A' \cap B)$

If the events A and B are independent then events A and B' as well as A' and B are also independent.

Probability of $(A \cap B') \cup (A' \cap B)$ $\quad = P(A \cap B') + P(A' \cap B)$

$$= [P(A) \times P(B')] + [P(A') \times P(B)]$$

$$= \left[\frac{3}{5} \times \frac{1}{4}\right] + \left[\frac{2}{5} \times \frac{3}{4}\right]$$

$$= \frac{3}{20} + \frac{6}{20}$$

$$= \frac{9}{20}$$

Required probability = $\frac{9}{20}$

Exercise 1.4

1. There are two children in a family. If the first child is a girl then find the probability that both the children in the family are girls.

2. Two six-faced balanced dice are thrown simultaneously. If the sum of numbers on both the dice is more than 7 then find the probability that both the dice show same numbers.

3. Among the various vehicle-owners visiting a petrol pump, 80% vehicle-owners visit to fill petrol in their vehicle and 60% vehicle-owners visit to fill air in their vehicles. 50% vehicle-owners visit to fill air and petrol in their vehicle. Find the probability for the following events :

(1) If a vehicle-owner has come to fill petrol in his vehicle then that vehicle-owner will fill air in his vehicle.

(2) If a vehicle-owner has come to fill air in his vehicle then that vehicle-owner will fill petrol in his vehicle.

4. 80% customers hold saving account and 50% customers hold current account of a nationalised bank. 90% of the customers hold at least one of the saving account and the current account. If one of the account holders randomly selected from this bank holds a current account, find the probability that he holds a saving account.

5. If $P(A)=\frac{2}{3}$, $P(B)=\frac{3}{5}$ and $P(B/A) = \frac{3}{4}$ for two events in the sample space of a random experiment then find $P(A/B)$.

6. If $P(M)=P(F)=\frac{1}{2}$, $P(A/M)= \frac{1}{10}$ and $P(A/F)=\frac{1}{2}$ for events A, M and F then find $P(A\cap M)$ and $P(A\cap F)$.

7. There are 2 gold-coins and 4 silver-coins in a box. The other box contains 3 gold and 5 silver coins. One coin is selected from each box. Find the probability that one of the selected coins is a gold coin and the other is a silver coin.

8. One joint family has 3 sons and 2 daughters whereas the other joint family has 2 sons and 4 daughters. One joint family is selected from two joint families and a child is randomly selected from that family. Find the probability that the selected child is a girl.

9. There are 10 icecream cones in a box of which 3 cones weigh less than the specification and the rest of the 7 cones have the specified weight. Two cones are randomly selected one by one with replacement. Find the probability that both the cones selected weigh less than the specified weight.

10. There are 10 CDs in a CD rack in which 6 are action film CDs and 4 are drama film CDs. Two CDs are randomly selected one by one without replacement from this box. Find the probability that the first selected CD is of action film and the second CD is of drama film.

11. If two balanced dice are thrown then find the probability that

(1) at least one die shows number 5

(2) the first die shows the number 5 or 6 and the other die shows an even number.

12. A problem in Mathematics is given to Tania, Kathan and Kirti to solve. The probabilities of them solving the problem correctly are $\frac{2}{3}, \frac{3}{4}$ and $\frac{1}{2}$ repectively. Find the probability that the problem is solved correctly.

13. Person A can hit the target in 3 out of 5 attempts whereas person B can hit the target in 5 out of 6 attempts. If both of them attempt simultaneously, find the probability that the target is hit.

14. Person A speaks truth in 90% cases whereas person B speaks truth in 80% cases. Find the probability that persons A and B differ in stating the same fact.

15. If three events A, B and C of a random experiment are independent events and $P(A)=0.2$, $P(B)=0.3$ and $P(C)=0.5$ then find $P(A\cup B\cup C)$.

*

1.7 Statistical Definition of Probability

We have seen the mathematical definition of probability earlier. This definition can help to find the probability only in the cases where the outcomes of the sample space of a random experiment are equi-probable and their number is known. But we find several cases in practice where the outcomes of the sample space are infinite and unknown. For example, there are many fish of different types in a huge lake. We have to find the probability of catching a certain type of fish when a fisherman throws net in the lake to catch fish. The mathematical definition of probability can not be used here as the total number of fish in the lake is unknown. Moreover, we come across many cases in practice where the outcomes of the random experiment are not equi-probable. For example, a trader transports certain goods from his godown to his sales centre. The event that these goods safely reach the sales centre and event that it does not safely reach the sales centre are not equi-probable events. It is not possible to evaluate probability using the mathematical definition of probability in such cases. Let us consider another definition of probability, called the statistical definition of probability, which is generally more useful in such situations.

Let us start with an illustration. We have to find the probability that a customer will purchase while visiting a showroom selling ready made garments for a long time. To know this, we should obtain the data about the customers purchasing from this show-room. These data can be obtained by sample inquiry. As the size of the sample increases, we can say that the information from the sample inquiry is more close to the true (population) information. Suppose it is found that 79 customers purchase out of 100 customers in the sample inquiry. When the number of customers in the sample inquiry was 500 then it was found that 403 customers purchased. The data obtained by increasing the sample size (n) are as follows :

Size of the sample (No. of customers visiting the show-room)	No. of customers purchasing r (Frequency)	Proportion of customers purchasing $\frac{r}{n}$ (Expected Frequency)
100	79	0.79
500	403	0.806
1000	799	0.799
5000	3991	0.7982
10,000	8014	0.8014

It can be seen from the above data that as the size of the sample n increases, the proportion or expected frequency of customers purchasing the ready-made garments takes values close to 0.8. We accept this value as the probability of the event that the customer visiting the show-room will purchase. Thus, the probability is obtained in the form of relative frequency. The definition of porbability based on the relative frequency is called the statistical definition of probability. It is also called the empirical definition. The definition is as stated below :

Suppose a random experiment is repeated n times under identical conditions. If an event A occurs in m trials out of n trials then the relative frequency $\frac{m}{n}$ of event A gives the estimate of the probability of event A, $P(A)$. When the larger and larger value of n is taken, that is when n tends to infinity, the limiting value of $\frac{m}{n}$ is called the probability of event A.

In notation,

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

The limiting value of the ratio $\frac{m}{n}$ when n tends to infinite value is denoted by $\lim_{n \to \infty} \frac{m}{n}$. In practice, the relative frequency $\frac{m}{n}$ itself is taken as the probability of event A. Now we shall consider the examples showing the use of the statistical definition of probability.

Illustration 36 : The sample data obtained about marks scored by a large group of candidates appearing for a public examination of 100 marks are given in the following table.

Marks	20 or less	21–40	41–60	61–80	81–100
No. of Candidates	83	162	496	326	124

One candidate is randomly selected from those appearing for the public examination. Find the probability that this candidate has scored :

(1) less than 41 marks

(2) More than 60 marks

(3) Marks from 21 to 80.

The number of candidates selected in the sample is $n = 83 + 162 + 496 + 326 + 124 = 1191$.

(1) Event A = The selected candidate scores less than 41 marks.

$P(A)$ = Relative frequency for the candidates scoring less than 41 marks.

$$= \frac{\text{No. of candidates scoring less than 41 marks}}{\text{Total number of candidates in the sample}} = \frac{m}{n}$$

m = No. of candidates scoring less than 41 marks

$= 83 + 162$

$= 245$

Now, $P(A) = \frac{m}{n}$

$= \frac{245}{1191}$

Required probability $= \frac{245}{1191}$

(2) Event B = The selected candidate scores more than 60 marks

$P(B)$ = relative frequency for candidates scoring more than 60 marks.

$$= \frac{\text{No. of candidates scoring more than 60 marks}}{\text{Total number of candidates in the sample}} = \frac{m}{n}$$

m = No. of candidates scoring more than 60 marks

$= 326 + 124$

$= 450$

Now, $P(B) = \frac{m}{n}$

$= \frac{450}{1191}$

$= \frac{150}{397}$

Required probability $= \frac{150}{397}$

(3) Event C = The selected candidate scores from 21 to 80 marks

$P(C)$ = relative frequency for candidates scoring from 21 to 80 marks.

$$= \frac{m}{n} = \frac{\text{No. of candidates scoring from 21 to 80 marks}}{\text{Total number of candidates in the sample}}$$

m = No. of candidates scoring from 21 to 80 marks

$= 162 + 496 + 326$

$= 984$

Now, $P(C) = \frac{m}{n}$

$= \frac{984}{1191}$

$= \frac{328}{1191}$

Required probability $= \frac{328}{1191}$

Illustration 37 : A factory runs in two shifts. The sample data about the quality of items produced in these shifts are shown in the following table :

Quality	Shift		Total
	I	II	
Defective items	24	46	70
Non-defective items	2176	2754	4930
Total	**2200**	**2800**	**5000**

One item is randomly selected from the production of the factory.

(1) If the item is taken from the production of the first shift then find the probability that it is defective.

(2) If the item is defective then find the probability that it is taken from the production of the first shift.

The total number of units in the sample = 5000

We shall define the events as follows :

Event A = The selected item is from the production of first shift

$$P(A) = \frac{\text{No. of items produced in the first shift}}{\text{Total number of items in the sample}} = \frac{m}{n}$$

m = No. of items produced in the first shift

= 2200

Now, $P(A) = \frac{m}{n}$

$= \frac{2200}{5000}$

Event D = The selected item is defective

$P(D)$ = relative frequency for defective items

$$= \frac{\text{No. of defective items}}{\text{Total number of items in the sample}} = \frac{m}{n}$$

m = No. of defective items

= 70

Now, $P(D) = \frac{m}{n}$

$= \frac{70}{5000}$

Event $A \cap D$ = The selected item is produced in the first shift and it is defective

$P(A \cap D)$ = relative frequency for event $A \cap D$

$$= \frac{\text{No. of items in event } A \cap D}{\text{Total number of items in the sample}} = \frac{m}{n}$$

m = No. of items in event $A \cap D$

= 24

Now, $P(A \cap D) = \frac{m}{n} = \frac{24}{5000}$

(1) The event that the item is defective when it is taken from the first shift = D/A

Probability of D/A using the formula of conditional event

$$P(D/A) = \frac{P(A \cap D)}{P(A)}$$

$$= \frac{\frac{24}{5000}}{\frac{2200}{5000}}$$

$$= \frac{24}{2200}$$

$$= \frac{3}{275}$$

Required probability $= \frac{3}{275}$

(This probability can be directly obtained as relative frequency $\frac{24}{2200}$ of the event D/A.)

(2) The event that the item is taken from the first shift when it is defective = A/D

Probability of A/D using the formula of condition probability

$$P(A/D) = \frac{P(A \cap D)}{P(D)}$$

$$= \frac{\frac{24}{5000}}{\frac{70}{5000}}$$

$$= \frac{24}{70}$$

$$= \frac{12}{35}$$

Required probability $= \frac{12}{35}$

(This probability can be directly obtained as relative frequency $\frac{24}{70}$ of the event A/D.)

Limitations : The limitations of the statistical definition of probability are as follows :

(1) The value of probability can be obtained by the statistical definition of probability only if $n \rightarrow \infty$ that is if n tends to infinity. But the infinite value of n can not be taken in practice.

(2) The probability obtained by this definition is an estimated value. The exact value of probability cannot be known using this definition.

Exercise 1.5

1. The sample data about monthly travel expense (in ₹) of a large group of travellers of local bus in a megacity are given in the following table :

Monthly travel expense (₹)	501–600	601–700	701–800	801–900	901 or more
No. of travellers	318	432	639	579	174

One person from this megacity travelling by local bus is randomly selected. Find the probability that the monthly travel expense of this person will be (1) more than ₹ 900 (2) at the most ₹ 700 (3) ₹ 601 or more but ₹ 900 or less.

2. The details of a sample inquiry of 4979 voters of constituency are as follows :

Details	**Males**	**Females**
Supporters of party *A*	1319	1118
Supporters of party *B*	1217	1325

One voter is randomly selected from this constituency.

(1) If this voter is a male, find the probability that he is a supporter of Party A.

(2) If this voter is a supporter of Party A, find the probability that he is a male.

*

Summary

- The events based on chance are called random events.
- The experiment which can be independently repeated under identical conditions and all its posible outcomes are known but which of the outcomes will appear can not be predicted with certainty before conducting the experiment is called a random experiment.
- The set of all possible outcomes of a random experiment is called a sample space of that experiment.
- A subset of the sample space of random experiment is called an event of that random experiment.
- U is a finite sample space and A and B are two events in it. If events A and B can never occur together that is if $A \cap B = \phi$ then the events A and B are called mutually exclusive events.
- If the group of favourable outcomes of events of a random experiment is the sample space then the events are called exhaustive events.
- The elementary events are mutually exclusive and exhaustive.
- If there is no apparant reason to believe that out of one or more events of a random experiment, any one event is more or less likely to occur than the other events then the events are called equi-probable events.
- The number of mutually exclusive, exhaustive and equi-probable outcomes in the sample space U of a random experiment is n. If m outcomes among them are favourable for the event A then probability of event A is $\frac{m}{n}$.
- The range of values of $P(A)$, the probability of any event A of a sample space U, is 0 to 1. That is $0 \le P(A) \le 1$.
- A and B are any two events in a finite sample space U. If the probability of occurrence of event A does not change due to occurrence (or non-occurrence) of event B then A and B are independent events.
- Suppose a random experiment is repeated n times under identical conditions. If an event A occurs in m trials out of n trials then the relative frequency $\frac{m}{n}$ of event A gives the estimate of the probability of event A, $P(A)$. When the larger and larger value of n is taken that is when n tends to infinity, the limiting value of $\frac{m}{n}$ is called the probability of event A.

List of Formulae

(1) Complementary event of A $A' = U - A$

(2) Difference event of A and B $A - B = A \cap B' = A - (A \cap B)$ (only event A occurrs.)

(3) Difference event of B and A $B - A = A' \cap B = B - (A \cap B)$ (only event B occurrs.)

(4) The probability of an event A of the sample space of a random experiment is $P(A) = \frac{m}{n}$.

(5) Law of addition of probability

For two events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

For any three events A, B and C,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

If two events A and B are mutually exclusive,

$$P(A \cup B) = P(A) + P(B)$$

If three events A, B and C are mutually exclusive,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

If two events A and B are mutually exclusive and exhaustive,

$$P(A \cup B) = P(A) + P(B) = 1$$

If three events A, B and C are mutually exclusive and exhaustive,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) = 1$$

(6) Conditional probability

Event B occurs under the condition that event A occurs

$$P(B/A) = \frac{P(A \cap B)}{P(A)}; \quad P(A) \neq 0$$

Event A occurs under the condition that event B occurs

$$P(A/B) = \frac{P(A \cap B)}{P(B)}; \quad P(B) \neq 0$$

(7) Law of multiplication of probability

For any two events A and B,

$P(A \cap B) = P(A) \times P(B/A);\ P(A) \neq 0$

$P(A \cap B) = P(B) \times P(A/B);\ P(B) \neq 0$

- For independent events A and B,

$P(A \cap B) = P(A) \times P(B)$

$P(A' \cap B') = P(A') \times P(B')$

$P(A' \cap B) = P(A') \times P(B)$

$P(A \cap B') = P(A) \times P(B')$

(8) According to statistical definition of probability,

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

Exercise 1

Section A

Find the correct option for the following multiple choice questions :

1. Which event is given by a special subset ϕ of the sample space U ?

(a) Certain event (b) Complementary event of ϕ

(C) Union of events U and ϕ (d) Impossible event

2. What is the value of $P(A \cap A')$ for events A and A' ?

(a) 1 (b) 0 (c) 0.5 (d) between 0 and 1

3. Which of the following options is true for any event of the sample space ?

(a) $P(A) < 0$ (b) $0 \leq P(A) \geq 1$ (c) $0 \leq P(A) \leq 1$ (d) $P(A) > 1$

4. Which of the following options is not true for any two events A and B in the sample space U where $A \subset B$?

(a) $P(A \cap B) = P(B)$ (b) $P(A \cap B) = P(A)$

(c) $P(A \cup B) \geq P(A)$ (d) $P(B - A) = P(B) - P(A)$

5. What is the other name of the classical definition of probability ?

(a) Mathematical definition (b) Axiomatic definition

(c) Statistical definition (d) Geometric definition

6. Which of the following statement for probability of elementary events H and T of random experiment of tossing a balanced coin is not true ?

(a) $P(T)=0.5$ (b) $P(H)+P(T)=1$ (c) $P(H \cap T)=0.5$ (d) $P(H)=0.5$

7. Which random experiment from the following random experiments has an infinite sample space ?

(a) Throwing two dice (b) Selecting two employees from an office

(c) To measure the life of electric bulb (d) Select a card from 52 cards

8. If $A \cup A' = U$ then what type of events are A and A' ?

(a) Independent events (b) Complementary events

(c) Certain events (d) Impossible events

9. If $P(A/B) = P(A)$ and $P(B/A) = P(B)$ then what type of events are A and B ?

(a) Independent events (b) Complementary events

(c) Certain events (d) Impossible events

10. Two events A and B of a sample space are mutually exclusive. Which of the following will be equal to $P(B-A)$?

(a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$

11. What is the total number of sample points in the sample space formed by throwing three six-faced balanced dice simultaneously ?

(a) 6^2 (b) 3^6 (c) 6×3 (d) 6^3

12. If one number is randomly selected between 1 and 20, what is the probability that the number is a multiple of 5 ?

(a) $\frac{1}{2}$ (b) $\frac{1}{6}$ (c) $\frac{1}{5}$ (d) $\frac{1}{3}$

13. If events A and B are independent, which of the following options is true ?

(a) $P(A \cap B) = P(A) \times P(B)$ (b) $P(A \cup B) = P(A) + P(B)$

(c) $P(A \cup B) = P(A) \times P(B)$ (d) $P(A \cap B) = P(A) + P(B)$

14. What is the probability of having 5 Thursdays in the month of February in a year which is not a leap year ?

(a) 0 (b) $\frac{1}{7}$ (c) $\frac{2}{7}$ (d) $\frac{3}{7}$

15. If $P(A) = 0.4$ and $P(B') = 0.3$ for two independent events A and B of a sample space then state the value of $P(A \cap B)$.

(a) 0.12 (b) 0.42 (c) 0.28 (d) 0.18

16. For two events A and B of a samples space, state the event $(A \cap B) \cup (A \cap B')$.

(a) ϕ (b) B (c) A (d) U

17. According to the mathematical definition of probability, what is the probability of each outcome among the n outcomes of a random experiment ?

(a) 0 (b) $\frac{1}{n}$ (c) 1 (d) can not say

Section B

Answer the following questions in one sentence :

1. Give two examples of random experiment.
2. Draw the Venn diagram for A-B, the difference event of A and B.
3. Define an event.
4. Write the sample space of a random experiment of throwing one balanced die and a balanced coin simultaneously.
5. Define conditional probability.
6. State the formula for the probability of occurrence of at least one event out of three events A, B and C.
7. Define independent events.
8. Write the law of multiplication of probability for two independent events A and B in a sample space.
9. Interpret $P(A/B)$ and $P(B/A)$.
10. When can we say that three events A, B and C in a sample space are exhaustive ?
11. Arrange $P(A \cup B)$, $P(A)$, $P(A \cap B)$, 0, $P(A) + P(B)$ in the ascending order.
12. Define :

(1) Random Experiment (2) Sample Space

(3) Equi-probable Events (4) Favourable Outcomes

(5) Probability (Mathematical definition) (6) Probability (Statistical definition)

(7) Impossible Event (8) Certain Event

13. For two events A and B in a sample space, $A \cap B = \phi$ and $A \cup B = U$. State the values of $P(A \cap B)$ and $P(A \cup B)$.

14. If two events A and B in a sample space are independent then state the formula for $P(A \cup B)$.

15. If $A = \{x \mid 0 < x < 1\}$ and $B = \{x \mid \frac{1}{4} \le x \le 3\}$ then find $A \cap B$.

16. For two independent events A and B, $P(A) = 0.5$ and $P(B) = 0.7$. Find $P(A' \cap B')$.

17. If $P(A) = 0.8$ and $P(A \cap B) = 0.25$, find $P(A - B)$.

18. If $P(A) = 0.3$ and $P(A \cap B) = 0.03$, find $P(B/A)$.

19. If $P(A) = P(B) = K$ for two mutually exclusive events A and B, find $P(A \cup B)$.

20. If $P(A' \cap B) = 0.45$ and $A \cap B = \phi$, find $P(B)$.

21. Two events A and B in a sample space are mutually exclusive and exhaustive. If $P(A) = \frac{1}{3}$, find $P(B)$.

22. 2% items in a lot are defective. What is the probability that an item randomly selected from this lot is non-defective ?

23. State the number of sample points in the random experiment of tossing five balanced coins.

24. State the number of sample points in the random experiment of tossing one balanced coin and two balanced dice simultaneously.

25. Is it possible that $P(A) = 0.7$ and $P(A \cup B) = 0.45$ for two events A and B in a sample space ?

26. Two cards are selected one by one with replacement from 52 cards. State the number of elements in the sample space of this random experiment.

27. For two independent events A and B, $P(B/A) = \frac{1}{2}$ and $P(A \cap B) = \frac{1}{5}$. Find $P(A)$.

28. 1998 tickets out of 2000 tickets do not have a prize. If a person randomly selects one ticket from 2000 tickets then what is the probability that the ticket selected is eligible for prize ?

Section C

Answer the following questions :

1. Define the following events and draw their venn diagram :

(1) Mutually exclusive events (2) Union of events

(3) Intersection of events (4) Difference event

(5) Exhaustive events (6) Complementary event

2. Give the illustrations of finite and infinite sample space.

3. Give the illustrations of impossible and certain event.

4. State the characteristics of random experiment.

5. State the assumptions of mathematical definition of probability.

6. State the limitations of mathematical definition of probability.

7. State the limitations of statistical definition of probability.

8. Explain the equiprobable events with illustration.

9. State the law of addition of probability for two events A and B. Write the law of addition of probability if these two events are mutually exclusive.

10. State the law of multiplication of probability for two events A and B. Write the law of multiplication of probability if these two events are independent.

11. State the following results for two independent events A and B :

(1) $P(A \cap B)$ (2) $P(A' \cap B')$

(3) $P(A \cap B')$ (4) $P(A' \cap B)$

12. If $P(A) = \frac{1}{3}$, $P(B) = \frac{2}{3}$ and $P(A \cap B) = \frac{1}{6}$ then find $P(A' \cap B')$.

13. If $P(B) = 2P(A/B) = 0.4$ then find $P(A \cap B)$.

14. If the events A and B are independent and $3P(A) = 2P(B) = 0.12$ then find $P(A \cap B)$.

15. If $5P(A) = 3P(B) = 2\,P(A \cup B) = \frac{3}{2}$ for two events A and B then find $P(A' \cup B')$.

16. If $P(A \cap B) = 0.12$ and $P(B) = 0.3$ for two independent events A and B then find $P(A \cup B)$.

17. If $A = \{x \mid 1 < x < 3\}$ and $B = \{x \mid \frac{1}{2} \le x < 2\}$ then find $A \cup B$ and $A \cap B$.

18. The probability of occurrence of at least one of the two events A and B is $\frac{1}{4}$. The probability that event A occurs but event B does not occur is $\frac{1}{5}$. Find the probability of event B.

19. If $P(B)=\frac{3}{5}$ and $P(A'\cap B)=\frac{1}{2}$, for two events A and B, find $P(A/B)$.

20. 6 persons have a passport in a group of 10 persons. If 3 persons are randomly selected from this group, find the probability that

(1) all the three persons have a passport

(2) two persons among them do not have a passport.

21. The probability that the tax-limit for income of males increases in the budget of a year is 0.66 and the probability that the tax-limit increases for income of females is 0.72. The probability that the tax-limit increases for income of both the males and females is 0.47. Find the probability that

(1) the tax-limit increases for income of only one of the two, males and females.

(2) the tax-limit does not increase for income of males as well as females in the budget of that year.

22. The price of petrol rises in 80% of the cases and the price of diesel rises in 77% of the cases after the rise in price of crude oil. The price of petrol and diesel rises in 68% cases. Find the probability that the price of diesel rises under the condition that there is a rise in the price of petrol.

23. As per the prediction of weather bureau, the probabilities for rains on three days, Thursday, Friday and Saturday in the next week are 0.8, 0.7 and 0.6 respectively. Find the probability that it rains on at least one of the three days in the next week.

(**Note :** The events of rains on three days, Thursday, Friday and Saturday of a week are independent.)

Section D

Answer the following questions :

1. 6 *LED* televisions and 4 *LCD* televisions are displayed in digital store A whereas 5 *LED* televisions and 3 *LCD* televisions are displayed in digital store B. One of the two stores is randomly selected and one television is selected from that store. Find the probability that it is an *LCD* television.

2. One number is randomly selected from the natural number 1 to 100. Find the probability that the number selected is either a single digit number or a perfect square.

3. A balanced coin is tossed thrice. If the first two tosses have resulted in tail, find the probability that tail appears on the coin in all the three trials.

4. If events A, B and C are independent events and $P(A) = P(B) = P(C) = p$ then find the value of $P(A \cup B \cup C)$ in terms of p.

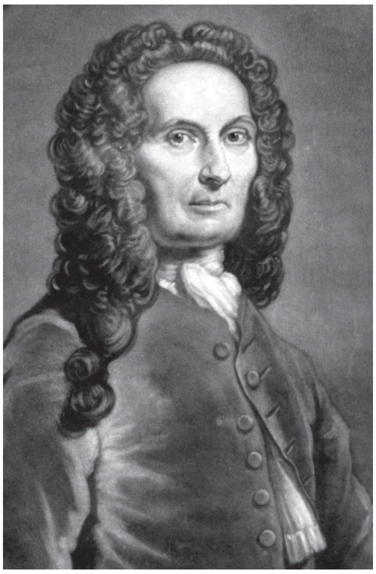
5. The genderwise data of a sample of 6000 employees selected from class 3 and class 4 employees in the government jobs of a state are shown in the following table :

Class of Employees	Gender		Total
	Males	Females	
Class 3	3600	900	4500
Class 4	400	1100	1500
Total	4000	2000	6000

One employee is randomly selected from all the class 3 and class 4 employees in government jobs of this state.

(1) If the selected employee is a male, find the probability that he belongs to class 3.

(2) If it is given that the selected employee belongs to class 3, find the probability that he is a male.



Abraham de Moivre
(1667 -1754)

Abraham de Moivre was a French mathematician known for de Moivre's formula, one of those that link complex numbers and trigonometry, and for his work on the normal distribution and probability theory. De Moivre wrote a book on probability theory, The Doctrine of Chances. De Moivre first discovered Binet's formula, the closed-form expression for Fibonacci numbers linking the n^{th} power of the golden ratio ϕ to the n^{th} Fibonacci number. He also was the first to postulate the Central Limit Theorem, a cornerstone of probability theory. In the later editions of his book, de Moivre included his unpublished result of 1733, which is the first statement of an approximation to the binomial distribution in terms of what we now call the normal or Gaussian function.

De Moivre continued studying the fields of probability and mathematics until his death and several additional papers were published after his death.

●

2

Random Variable and Discrete Probability Distribution

Contents :

2.1 Random Variable

We have studied about random experiment, sample space and probability in the chapter of probability. In this chapter, we shall study random variable and discrete probability distribution.

First of all, we shall define random variable and then we shall understand it by illustration.

Random Variable : Let U be a sample space of a random experiment. Every element of U need not always be a number. However, we wish to assign a specific number to each outcome.

A function associating a real number with each outcome of U is called a random variable. It is denoted by X. That is, a random variable based on a sample space U is denoted by $X : U \to R$.

For example,

(i) The number of heads (H) in tossing an unbiased coin three times

(ii) The number of accidents during a week in a city

(iii) The weight of a person (in kilogram)

(iv) The maximum temperature of a day at a particular place (in Celsius)

Now, let us understand the concept of random variable by some illustrations.

(1) A balanced die is tossed once. If the number observed on the die is denoted by 'u' then the elements of the sample space U of this experiment can be shown in the notation of a set as follows :

$$U = \{u \mid u = 1, 2, 3, 4, 5, 6\}$$

That is $U = \{1, 2, 3, 4, 5, 6\}$

If we associate a real number X with element u of sample space by

$X(u)$ = the number obtained on the die then we can write

$$X(u) = u, u = 1, 2, 3, 4, 5, 6$$

Thus, variable X will be a random variable assuming values 1, 2, 3, 4, 5 and 6.

In the above illustration, the element of U are numeric. Now, we consider an illustration in which the elements of U are non-numeric.

(2) Suppose a box contains four balls : one red, one blue, one yellow and one white ball. We donote the red ball by R, the blue ball by B, the yellow ball by Y and the white ball by W. A person draws three balls at a time at random from the box. The sample space associated with this experiment is

$$U = \{RBY, RBW, BYW, WYR\}$$

Suppose for the element u of U,

$X(u) =$ the number of white ball in u then $X(RBY) = 0$, $X(RBW) = 1$, $X(BYW) = 1$, $X(WYR) = 1$.

Thus, random variable X assumes the values in the set $\{0, 1\}$. The outcomes of this sample space are not in numbers but we associate them with real numbers by a random variable.

(3) Suppose the heights of students in a class lie between 120 cm and 180 cm. If we measure the height of a student of this class then it will assume any value between 120 cm and 180 cm.

Here, the sample space is $U = \{u \mid 120 \le u \le 180\}$.

If we denote the height of a selected student by X then $X(u) = u =$ the height (in cm) of a selected student. Thus, X becomes a random variable which will be denoted as $X = x$, $120 \le x \le 180$.

In the above example (1) and example (2), random variable X assumes particular countable values whereas in example (3), random variable X can assume any value in the interval [120, 180]. This random variable differs from the random variables in earlier two examples.

Now, we shall understand the difference between these random variables in the following section.

2.1.1 Discrete Random Variable

A random variable X which can assume a finite or countable infinite number of values in the set R of real numbers is called a discrete random variable.

For example (i) Birth year of a randomly selected student.

(ii) Number of broken eggs in a box of 6 eggs.

Now, we shall understand about the discrete random variable by some specific examples.

(1) Suppose there is one black and two white balls in a box. Suppose the black ball is denoted by B and two white balls by W_1 and W_2. A person can play the following game by paying ₹ 15.

The person playing a game is asked to select two balls randomly with replacement from the box. He is paid an amount according to the colour of the balls selected by him as per the following conditions :

If a white ball is selected then ₹ 5 are paid for each selected white ball and if a black ball is selected then ₹ 15 are paid per black ball.

If we denote the net amount earned (amount received – amount paid for the game) by the player corresponding to each outcome of the experiment by X then X becomes a discrete random variable. The values assumed by the variable X are denoted in the following table :

Outcome of the experiment (Event)	**The amount received by the person by playing the game**	**The amount paid to play the game**	**The value of X (in ₹)**
W_1W_1	$5 + 5 = 10$	15	$X(W_1W_1) = 10 - 15 = -5$
W_1W_2	$5 + 5 = 10$	15	$X(W_1W_2) = 10 - 15 = -5$
W_1B_1	$5 + 15 = 20$	15	$X(W_1B_1) = 20 - 15 = 5$
W_2W_1	$5 + 5 = 10$	15	$X(W_2W_1) = 10 - 15 = -5$
W_2W_2	$5 + 5 = 10$	15	$X(W_2W_2) = 10 - 15 = -5$
W_2B_1	$5 + 15 = 20$	15	$X(W_2B_1) = 20 - 15 = 5$
B_1W_1	$15 + 5 = 20$	15	$X(B_1W_1) = 20 - 15 = 5$
B_1W_2	$15 + 5 = 20$	15	$X(B_1W_2) = 20 - 15 = 5$
B_1B_1	$15 + 15 = 30$	15	$X(B_1B_1) = 30 - 15 = 15$

Thus, the random variable X assumes the values -5, 5 and 15 only. That is the total number of values of X is finite.

(2) Suppose a coin is tossed until either a tail (T) or four heads (H) occur. Let X denote the number of tosses required.

The sample space associated with this random experiment is

$U = \{T, HT, HHT, HHHT, HHHH\}$

The random variable X denotes the number of tosses required for the coin associated with the experiment and it assumes any one value out of 1, 2, 3 and 4 for the sample points of the sample space.

$$X(T) = 1,\ X(HT) = 2,\ X(HHT) = 3$$

$$X(HHHT) = 4,\ X(HHHH) = 4$$

The discrete random variable X assumes the finite number of values.

(3) Consider the random variable X denoting the number of tails before getting the first head in the experiment of tossing a coin till the first head is obtained.

In this experiment, head will appear either in the first trial or in the second trial or in the third trial and so on... Similarly, the first head may be obtained after tossing a coin infinite times. Hence, the sample space associated with random experiment becomes

$$U = \{H, TH, TTH, TTTH, TTTTH, ...\}$$

Thus, the number of tails before getting the first head will be $0, 1, 2, 3, 4....$

Thus, the random variable X assumes any one value from the countable infinite number of values $0, 1, 2, 3, 4....$

2.1.2 Continuous Random Variable

A random variable X which can assume any value in R, the set of real numbers or in any interval of R is called a continuous random variable.

For example (i) The actual amount of coffee in a coffee mug having a capacity of 250 millilitre.

(ii) Waiting time for a lift on any one floor of a high-rise office building.

Now, we shall understand more about the continuous random variable by the following examples.

(1) Denote the time taken by a student to finish a test of 3 hours duration by random variable X. The sample space here is

$$U = \{u \mid 0 \le u \le 3\}.$$

Since the time taken by any student for the exam takes any real value from 0 to 3 and the random variable X, the actual time taken by a student to complete the exam, will also be any real value from 0 to 3.

Thus,

$$X(u) = u, \ 0 \le u \le 3.$$

That means $X = x, 0 \le x \le 3$

The random variable X assumes any real value from 0 to 3, which is a subset of R and hence X is a continuous random variable.

(2) Suppose there are two stations A and B on an express highway. The distance of station B from station A is 200 km. Let us consider an experiment to know the place of accident between two stations A and B. For the sake of simplicity, let us fix the position of station A at 0 km and of station B at 200 km. The sample space of this experiment is any real value between 0 to 200. So, we can write the sample space for this experiment as

$$U = \{u \mid 0 \le u \le 200\}$$

Suppose the random variable X denotes the distance (in kilometer) of the place of the accident between two stations A and B from the station A. Then the random variable X is defined as below :

$X(u)$ = distance of the place of accident from the station A.

In short, we can define the random variable X as $X = x, 0 \le x \le 200$.

The random variable X assumes any real value from 0 to 200, which is subset of R, the set of real numbers. So, X is a continuous random variable.

2.2 Discrete Probability Distribution

Suppose $X : U \rightarrow R$ is a random variable which assumes all the values of the subset $\{x_1, x_2, ..., x_n\}$ of R. Further, suppose X assumes a value x_i with probability $P(X = x_i) = p(x_i)$. If $p(x_i) > 0, \ i = 1, 2, ..., n$ and $\Sigma \, p(x_i) = 1$ then the set of real values $\{x_1, x_2, ..., x_n\}$ and $\{p(x_1), p(x_2), ..., p(x_n)\}$ is called the discrete probability distribution of a random variable X. The discrete probability distribution of a random variable X is expressed in a tabular form as follow :

$X = x$	x_1	x_2	....	x_i	...	x_n
$p(x)$	$p(x_1)$	$p(x_2)$	....	$p(x_i)$	...	$p(x_n)$

Here, $0 < p(x_i) < 1, i = 1, 2, ..., n$ and $\Sigma \, p(x_i) = 1$

2.2.1 Illustrations for Probability Distribution of Discrete Variable

Illustration 1 : Determine whether the values given below are appropriate as the values of a probability distribution of a discrete random variable X, which assumes the values 1, 2, 3 and 4 only.

(i) $p(1) = 0.25, p(2) = 0.75, p(3) = 0.25, p(4) = -0.25$

(ii) $p(1) = 0.15, p(2) = 0.27, p(3) = 0.29, p(4) = 0.29$

(iii) $p(1) = \frac{1}{19}, \quad p(2) = \frac{9}{19}, \quad p(3) = \frac{3}{19}, \quad p(4) = \frac{4}{19}$

(i) The value of $P(4)$ is -0.25, which is negative. It does not satisfy the condition $p(x_i) > 0, i = 1, 2, 3, 4$ of discrete probability distribution. So, given values are not suitable for the probability distribution of a discrete variable. Thus, the given distribution cannot be called a probability distribution of a discrete variable.

(ii) For every value 1, 2, 3 and 4 of X, $p(x) > 0$, and $p(1) + p(2) + p(3) + p(4) = 1$. Thus, both the conditions of probability distribution of discrete variable are satisfied. So, the given values are appropriate and the given distribution is probability distribution of a discrete variable.

(iii) Here $p(x_i) > 0, i = 1, 2, 3, 4$ but, sum of probabilities

i.e. $p(1) + p(2) + p(3) + p(4) = \frac{17}{19}$, is not 1. So, the given values are not appropriate for the probability distribution. So, the given distribution cannot be called a probability distribution of discrete variable.

Illustration 2 : Determine when the following distribution is a probability distribution of discrete variable. Hence obtain the probability for $x = 2$:

$$p(x) = c\left(\frac{1}{4}\right)^x, \quad x = 1, 2, 3, 4$$

Here, $p(1) = c\left(\frac{1}{4}\right)$, $p(2) = c\left(\frac{1}{4}\right)^2 = c\left(\frac{1}{16}\right)$, $p(3) = c\left(\frac{1}{4}\right)^3 = c\left(\frac{1}{64}\right)$, $p(4) = c\left(\frac{1}{4}\right)^4 = c\left(\frac{1}{256}\right)$

Now, total probability should be 1 for a discrete probability distribution.

$p(1)+p(2)+p(3)+p(4)=1$

$\therefore c\left(\frac{1}{4}\right)+c\left(\frac{1}{16}\right)+c\left(\frac{1}{64}\right)+c\left(\frac{1}{256}\right)=1$

$\therefore c\left[\frac{1}{4}+\frac{1}{16}+\frac{1}{64}+\frac{1}{256}\right]=1$

$\therefore c\left[\frac{85}{256}\right]=1$

$\therefore c=\frac{256}{85}$

Thus, when $c=\frac{256}{85}$, the given distribution becomes probability distribution of a discrete variable.

Now, $P(X=2)=c\left(\frac{1}{4}\right)^2$

$=\frac{256}{85}\times\frac{1}{16}$

$=\frac{16}{85}$

$\therefore$ The probability of $X=2$ is $\frac{16}{85}$.

Illustration 3 : A random variable X denotes the number of accidents per year in a factory and the probability distribution of X is given below :

$X=x$	0	1	2	3	4
$p(x)$	$4K$	$15K$	$25K$	$5K$	K

(i) Find the constant K and rewrite the probability distribution.

(ii) Find the probability of the event that one or two accidents will occur in this factory during the year.

(iii) Find the probability that no accidents will take place during the year in the factory.

(i) By the definition of discrete probability distribution, we must have

$p(0)+p(1)+p(2)+p(3)+p(4)=1$

That is $4K+15K+25K+5K+K=1$

$\therefore 50K=1$

$\therefore K=\frac{1}{50}$

$=0.02$

Thus, when $K=0.02$, the given distribution becomes a probability distribution of a discrete variable, which is given below :

$X=x$	0	1	2	3	4	Total
$p(x)$	0.08	0.30	0.50	0.10	0.02	1

(ii) Probability of occurrence of one or two accidents

$= P(X=1)+P(X=2)$

$= 0.30 + 0.50$

$= 0.80$

(iii) Probability that accidents do not occur :

$= P(X=0)$

$= 0.08$

Illustration 4 : In a factory, packets of produced blades are prepared having 50 blades in each packet. A quality control engineer randomly selects a packet from these packets and examines all the blades of the selected packet. If 4 or more defective blades are observed in the selected packet then the packet is rejected. The probability distribution of the defective blades in the packet is given below :

Number of defective blades in the packet	0	1	2	3	4	5	6 or more
Probability	$9K$	$3K$	$3K$	$2K$	$2K$	$K-0.02$	0.02

From the given probability distribution,

(i) Find constant *K*.

(ii) Find the probability that the randomly selected packet is accepted by the quality control engineer.

(i) Let X = number of defective blades found during the inspection of the packet.

By definition of discrete probability distribution

$$p(0)+p(1)+p(2)+p(3)+p(4)+p(5)+p(6 \text{ or more})=1.$$

$\therefore\ 9K + 3K + 3K + 2K + 2K + K - 0.02 + 0.02 = 1$

$\therefore\ 20K = 1$

$\therefore\ K = \frac{1}{20} = 0.05$

(ii) The randomly selected packet is accepted by the quality control engineer only when 3 or less defective blades are found in the packet.

$\therefore\ P(X \le 3)$

$= p(0)+p(1)+p(2)+p(3)$

$= 9K + 3K + 3K + 2K$

$= 17K$

$= 17(0.05)$

$= 0.85 \quad (\because K = 0.05)$

Illustration 5 : There are 4 red and 2 white balls in a box. 2 balls are drawn at random from the box without replacement. Obtain probability distribution of number of white balls in the selected balls.

Suppose X denotes the number of white balls in the selected two balls. X may assume the values 0, 1 and 2.

$X = 0$ means there will not be any white balls in the selected two balls that means both the selected balls are red.

$$\therefore P(X=0) = P\text{ (2 red balls)} = \frac{{}^4C_2}{{}^6C_2} = \frac{6}{15}$$

Now, $x = 1$ means there will be one white ball and one red ball in the two selected balls.

$$\therefore P(X=1) = P\text{ (1 White ball, 1 Red ball)}$$

$$= \frac{{}^2C_1 \times {}^4C_1}{{}^6C_2}$$

$$= \frac{2 \times 4}{15} = \frac{8}{15}$$

And $X = 2$ means both the selected ball will be white.

$$\therefore P(X=2) = P\text{ (2 White balls)}$$

$$= \frac{{}^2C_2}{{}^6C_2}$$

$$= \frac{1}{15}$$

Thus, probability distribution of random variable X can be written as follows :

$X = x$	0	1	2
$p(x)$	$\frac{6}{15}$	$\frac{8}{15}$	$\frac{1}{15}$

$p(x) > 0$ and $\Sigma p(x) = 1$

2.2.2 Mean and Variance

Now, we will discuss two important results based on the probability distribution of discrete random variable. One of them is expected value (mean) of the random variable and the other is variance of the random variable.

Let X be a discrete random variable which assumes one of the values $x_1, x_2, ..., x_n$ only and its probability distribution is as follows :

$X = x$	x_1	x_2	....	x_i	...	x_n
$p(x)$	$p(x_1)$	$p(x_2)$	....	$p(x_i)$	...	$p(x_n)$

Where $0 < p(x_i) < 1, i = 1, 2, ..., n$ and $\Sigma\, p(x_i) = 1$

The mean of discrete random variable is denoted by μ or $E(X)$. It is defined as follows :

$$\mu = E(X) = \Sigma x_i \, p(x_i)$$

This value is also called expected value of discrete variable X.

The variance of discrete random variable X is denoted by σ^2 or $V(X)$, which is defined as follows :

$$\sigma^2 = V(X) = E(X-\mu)^2$$

$$= E(X^2) - (\mu)^2$$

$$= E(X^2) - (E(X))^2$$

Where $E(X^2) = \Sigma x_i^2 p(x_i)$

Note : (i) We will use the following notations for the sake of simplicity.

$\Sigma x \, p(x)$ instead of $\Sigma x_i \, p(x_i)$

and

$\Sigma x^2 p(x)$ instead of $\Sigma x_i^2 p(x_i)$

(ii) The mean and variance of variable X are also called mean and variance of the distribution of X respectively.

(iii) The value of the variance of variable X is always positive.

We consider the following examples to find mean and variance of the discrete probability distribution.

Illustration 6 : Find constant C for the following discrete probability distribution. Hence obtain mean and variance of this distribution.

$$p(x) = C \cdot {}^4P_x, \; x = 0, 1, 2, 3, 4$$

From the property of discrete probability distribution we must have

$$p(0) + p(1) + p(2) + p(3) + p(4) = 1$$

$$\therefore \; C \cdot {}^4P_0 + C \cdot {}^4P_1 + C \cdot {}^4P_2 + C \cdot {}^4P_3 + C \cdot {}^4P_4 = 1$$

$$\therefore C\left[\frac{4!}{4!} + \frac{4!}{3!} + \frac{4!}{2!} + \frac{4!}{1!} + \frac{4!}{0!}\right] = 1$$

$$\therefore C[1 + 4 + 12 + 24 + 24] = 1$$

$$\therefore C[65] = 1$$

$$\therefore C = \frac{1}{65}$$

Thus, the probability distribution can be written in the tabular form as follow :

$X = x$	0	1	2	3	4	Total
$p(x)$	$\frac{1}{65}$	$\frac{4}{65}$	$\frac{12}{65}$	$\frac{24}{65}$	$\frac{24}{65}$	1

Now, mean of the distribution $= \mu = \Sigma xp(x)$

$$= 0\left(\frac{1}{65}\right) + 1\left(\frac{4}{65}\right) + 2\left(\frac{12}{65}\right) + 3\left(\frac{24}{65}\right) + 4\left(\frac{24}{65}\right)$$

$$= \frac{0+4+24+72+96}{65}$$

$$= \frac{196}{65}$$

Now, we obtain $E\left(X^2\right)$.

$$E\left(X^2\right) = \Sigma\, x^2 p(x)$$

$$= (0)^2\left(\frac{1}{65}\right) + (1)^2\left(\frac{4}{65}\right) + (2)^2\left(\frac{12}{65}\right) + (3)^2\left(\frac{24}{65}\right) + (4)^2\left(\frac{24}{65}\right)$$

$$= 0 + \frac{4}{65} + \frac{48}{65} + \frac{216}{65} + \frac{384}{65}$$

$$= \frac{652}{65}$$

Hence, variance of the distribution $= V(X)$

$$= E\left(X^2\right) - \left(E(X)\right)^2$$

$$= \frac{652}{65} - \left(\frac{196}{65}\right)^2$$

$$= \frac{42380 - 38416}{4225} = \frac{3964}{4225}$$

Illustration 7 : There are two red and one green balls in a box. Two balls are drawn at random with replacement from the box. Obtain probability distribution of number of red balls in the two balls drawn and find its mean and variance.

Let us denote the number of red balls in the selected two balls by X. Then we obtain the probability distribution of X as follow.

Let us denote the two red balls of the box by R_1 and R_2 and green ball by G.

The number of red balls in the selected balls and its probability can be obtained as in the following table.

Selected two balls (Event)	**Probability of the event**	$X = x$
R_1R_1	$\frac{1}{9}$	2
R_1R_2	$\frac{1}{9}$	2
R_1G	$\frac{1}{9}$	1
R_2R_1	$\frac{1}{9}$	2
R_2R_2	$\frac{1}{9}$	2
R_2G	$\frac{1}{9}$	1
GR_1	$\frac{1}{9}$	1
GR_2	$\frac{1}{9}$	1
GG	$\frac{1}{9}$	0

From the above table we can say that :

(i) Probability of getting 0 red ball

$= P(X = 0)$

$= \frac{1}{9}$

(ii) Probability of getting 1 red ball

$= P(X = 1)$

$= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9}$

$= \frac{4}{9}$

(iii) Probability of getting 2 red balls

$= P(X = 2)$

$= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9}$

$= \frac{4}{9}$

Thus, the probability distribution of X can be written in the tabular from as follow :

$X = x$	0	1	2	Total
$p(x)$	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$	1

Now, mean of the distribution $= \mu = E(X)$

$$= \Sigma x\, p(x)$$

$$= 0\left(\tfrac{1}{9}\right) + 1\left(\tfrac{4}{9}\right) + 2\left(\tfrac{4}{9}\right)$$

$$= \frac{0 + 4 + 8}{9}$$

$$= \frac{12}{9}$$

Now, we first find $E\left(X^2\right)$ to obtain variance of the distribution.

$$E\left(X^2\right) = \Sigma x^2 p(x)$$

$$= 0^2\left(\tfrac{1}{9}\right) + 1^2\left(\tfrac{4}{9}\right) + 2^2\left(\tfrac{4}{9}\right)$$

$$= \frac{0+4+16}{9}$$

$$= \frac{20}{9}$$

So, using the formula $V(X) = E\left(X^2\right) - \left(E(X)\right)^2$,

$$V(X) = \frac{20}{9} - \left(\frac{12}{9}\right)^2$$

$$= \frac{20}{9} - \frac{144}{81}$$

$$= \frac{180 - 144}{81}$$

$$= \frac{36}{81}$$

Illustration 8 : There are 2 black and 2 white balls in a box. Two balls are drawn without replacement from it. Obtain probability distribution of the number of white balls in the selected balls. Hence find its mean and variance.

Suppose X = number of white balls in the selected two balls then by the formula of probability

(i) Probability of $X = 0$

$$= P(X = 0) = P \text{ (0 white balls)} = \frac{{}^2C_0}{{}^4C_2} = \frac{1}{6}$$

(ii) Probability of $X = 1$

$= P(X=1) = P$ (1 white ball and 1 black ball)

$= \dfrac{{}^2C_1 \times {}^2C_1}{{}^4C_2}$

$= \dfrac{2 \times 2}{6}$

$= \dfrac{4}{6}$

(iii) Probability of $X = 2$

$= P(X=2) = P$ (2 white balls)

$= \dfrac{{}^2C_2}{{}^4C_2}$

$= \dfrac{1}{6}$

Thus, the probability distribution of random variable X can be written in the tabular form as,

$X = x$	0	1	2	Total
$p(x)$	$\frac{1}{6}$	$\frac{4}{6}$	$\frac{1}{6}$	1

Now, mean of the probability distribution $= E(X)$

$= \Sigma\, x\, p(x)$

$= 0\left(\frac{1}{6}\right) + 1\left(\frac{4}{6}\right) + 2\left(\frac{1}{6}\right)$

$= \dfrac{0+4+2}{6}$

$= 1$

Now, to obtain variance of the probability distribution, we first find $E\left(X^2\right)$.

$E\left(X^2\right) = \Sigma\, x^2 p(x)$

$= 0^2\left(\frac{1}{6}\right) + 1^2\left(\frac{4}{6}\right) + 2^2\left(\frac{1}{6}\right)$

$= \dfrac{0+4+4}{6}$

$= \dfrac{8}{6}$

$\therefore V(X) = E\left(X^2\right) - \left(E(X)\right)^2$

$= \frac{8}{6} - (1)^2 \qquad (\because E(X) = 1)$

$= \dfrac{8-6}{6}$

$= \dfrac{2}{6}$

$= \dfrac{1}{3}$

Illustration 9 : Let X denote the maximum integer among the outcomes of tossing two dice simultaneously. Obtain the probability distribution of variable X and find its mean and variance.

By tossing two dice simultaneously, we have 36 events in the sample space U and the maximum integer of the outcomes will be one of the numbers 1, 2, 3, 4, 5 or 6. The following table gives the possible outcomes for variable X and the corresponding probability :

Event u of U	**Maximum integer** $X(u)=x$	$P(X=x)$
(1, 1)	1	$\frac{1}{36}$
(1, 2), (2, 1), (2, 2)	2	$\frac{3}{36}$
(1, 3), (2, 3), (3, 3) (3, 2), (3, 1)	3	$\frac{5}{36}$
(1, 4), (2, 4), (3, 4), (4, 4) (4, 3), (4, 2), (4, 1)	4	$\frac{7}{36}$
(1, 5), (2, 5), (3, 5), (4, 5), (5, 5) (5, 4), (5, 3), (5, 2), (5, 1)	5	$\frac{9}{36}$
(1, 6), (2, 6), (3, 6), (4, 6), (5, 6) (6, 6), (6, 5), (6, 4), (6, 3), (6, 2) (6, 1)	6	$\frac{11}{36}$
		Total 1

Now, mean of $X = E(X)$

$$= \Sigma x\, p(x)$$

$$= 1\left(\frac{1}{36}\right)+2\left(\frac{3}{36}\right)+3\left(\frac{5}{36}\right)+4\left(\frac{7}{36}\right)+5\left(\frac{9}{36}\right)+6\left(\frac{11}{36}\right)$$

$$= \frac{161}{36}$$

Now, $E\left(X^2\right)=\Sigma x^2 p(x)$

$$=1^2\left(\frac{1}{36}\right)+2^2\left(\frac{3}{36}\right)+3^2\left(\frac{5}{36}\right)+4^2\left(\frac{7}{36}\right)+5^2\left(\frac{9}{36}\right)+6^2\left(\frac{11}{36}\right)$$

$$=\frac{791}{36}$$

Variance of X $= V(X)$

$$= E\left(X^2\right)-\left(E(X)\right)^2$$

$$= \frac{791}{36} - \left(\frac{161}{36}\right)^2$$

$$= \frac{791}{36} - \frac{25921}{1296}$$

$$= \frac{791 \times 36 - 25921}{1296}$$

$$= \frac{28476 - 25921}{1296}$$

$$= \frac{2555}{1296}$$

Illustration 10 : It is observed from the life table that the probability that a 40 years old man will live one more year is 0.95. Life insurance company wishes to sell one year life insurance policy of Rs. 10,000 to such a man. What should be the minimum premium of the policy so that expected gain of the company would be positive ?

Let X be the company's gain and yearly premium of the policy be ₹ K, $K > 0$. Then gain of the company is $X = K$ if 40 year old man will live for one year and gain of the company is $X = K - 10,000$ if 40 year old man will die within a year.

Thus, the probability distribution of the gain of the company is as follow :

$X = x$	K	$K - 10000$
$p(x)$	0.95	0.05

Hence, expected gain of the company

$= E(X)$

$= \Sigma x\, p(x)$

$= K(0.95) + (K - 10000)(0.05)$

$= K(0.95) + K(0.05) - 500$

$= K(0.95 + 0.05) - 500$

$= K - 500$

Now, for positive expected gain, we must have

$K - 500 > 0$

$\therefore K > 500$

So, the company should fix the premium more than ₹ 500 so that the expected gain of the company be will positive.

EXERCISE 2.1

1. Examine whether the following distribution is a probability distribution of a discrete random variable X :

$$p(x) = \frac{x+2}{25}, \quad x = 1, 2, 3, 4, 5$$

2. If the following distribution is a probability distribution of variable X then find constant K.

$$p(x) = \frac{6-|x-7|}{K}, \quad x = 4, 5, 6, 7, 8, 9, 10$$

3. The probability distribution of a random variable X is defined as follows :

$$p(x) = \frac{K}{(x+1)!}, \quad x = 1, 2, 3; K = \text{constant}$$

Hence find (i) constant K (ii) $P(1 < X < 4)$

4. The probability distribution of a random variable X is as follows :

$X = x$	-2	-1	0	1	2
$p(x)$	$\frac{K}{3}$	$\frac{K}{3}$	$\frac{K}{3}$	$2K$	$4K^2$

Then (i) determine acceptable value of constant K. (ii) Find the variance of the distribution.

5. The probability distribution of a random variable X is $P(x)$. Variable X can assume the values $x_1 = -2, x_2 = -1, x_3 = 1$ and $x_4 = 2$ and if $4P(x_1) = 2P(x_2) = 3P(x_3) = 4P(x_4)$ then obtain mean and variance of this probability distribution.

6. A die is randomly tossed two times. Determine the probability distribution of the sum of the numbers appearing both the times on the die and obtain expected value of the sum.

7. A box contains 4 red and 2 blue balls. Three balls are simultaneously drawn at random. If X denotes the number of red balls in the selected balls, find the probability distribution of X and find the expected number of red balls in the selected balls.

8. A coin is tossed till either a head or 5 tails are obtained. If a random variable X denotes the necessary number of trials of tossing the coin then obtain probability distribution of the random variable X and calculate its mean and variance.

9. A shopkeeper has 6 tickets in a box. 2 tickets among them are worth a prize of ₹ 10 and the remaining tickets are worth a prize of ₹ 5. If a ticket is drawn at random from the box, find the expected value of the prize.

*

2.3 Binomial Probability Distribution

In the earlier sections we considered continuous and discrete random variable and probability distribution of a discrete random variable. Now, we shall study an important probability distribution of a discrete random variable.

In some random experiments, there are only two outcomes. We call such outcomes as success and failure. These outcomes are mutually exclusive. We call such experiments as dichotomous experiments. The illustrations of some of these situations are given in the table below :

Experiment		Possible outcomes	
		Success	**Failure**
(i)	To know the effect of advertisement given to increase the sale of produced units	sale increased	sale did not increase
(ii)	To find the error in a letter typed by a type-writer	Error observed	Error not observed
(iii)	To know the effect of a drug on blood pressure given to the patients of high blood pressure	Blood pressure decreased	Blood pressure did not decrease
(iv)	To inspect whether produced item is defective	Item is defective	Item is not defective

If we denote the success by S and failure by F for such types of dichotomous experiment and the probabilities of such outcomes by p and q respectively then

$P(S) = p$ and $P(F) = q, \ 0 < p < 1, \ 0 < q < 1, \ p + q = 1$

Since there are only two outcomes of such an experiment and both are mutually exclusive, we have $p + q = 1$ and hence $q = 1 - p$.

If it is possible to repeat such a dichotomous random experiment n times and each repetition is done under identical conditions then the probability of success p remains constant in each trial. We call such experiments as Bernoulli Trials. Its actual definition can be given as follows :

Bernoulli Trials : Suppose dichotomous random experiment has two outcomes, success (S) and failure (F). If this experiment is repeated n times under identical conditions and the probability $p(0 < p < 1)$ of getting a success at each trial is constant then such trials are called Bernoulli Trials.

Properties of Bernoulli Trials

(1) The probability of getting a success at each Bernoulli trial remains constant.

(2) Bernoulli trials are mutually independent. That means getting success or failure at any trial does not depend on getting success or failure at the previous trial.

(3) Success and failure are mutually exclusive and exhustive events. Therefore $q = 1 - p$.

Binomial Probability Distribution

Suppose X denotes the number of successes in a sequence of success *(S)* and failure (F) obtained in n Bernoulli trials, then X is called a binomial random variables and X assumes any value in the finite set $\{0, 1, 2, ..., n\}$. The probability distribution of the binomial random variable X is defined by the following formula :

$$P(X = x) = p(x) = {}^nC_x p^x q^{n-x}, \ x = 0, 1, 2, ..., n, 0 < p < 1, q = 1 - p$$

This probability distribution is called binomial Probability Distribution. We shall call such a distribution in short as binomial distribution.

If positive integer n and probability of success p are known here, the whole probability distribution that means the probability of each possible value of X can be determined. Hence, n and p are called parameters of the binomial distribution. We denote binomial distribution having parameters n and p as $b(n, p)$.

Note : If we repeat an experiment having such Bernoulli trials N times and $p(x)$ is the probability of getting x successes in the experiment then expected frequency of number of successes in N repetitions $= N \cdot p(x)$

2.3.1 Properties of Binomial Distribution

(1) Binomial distribution is a discrete distribution.

(2) Its parameters are n and p.

(3) The mean of the distribution is np which denotes average (expected) number of successes in n Bernoulli trials.

(4) The variance of the distribution is npq and its standard deviation is $\sqrt{npq}$.

(5) For binomial distribution, mean is always greater than the variance and $\frac{\text{Variance}}{\text{Mean}} = q$ = probability of failure.

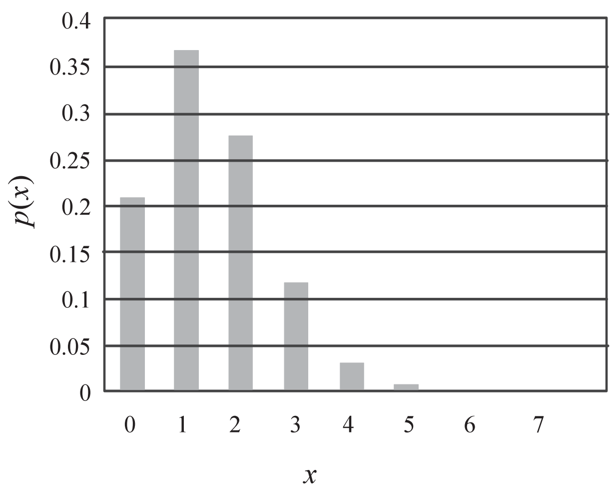
(6) If $p < \frac{1}{2}$ then the skewness of the distribution is positive for any value of n.

(7) If $p = \frac{1}{2}$ then the distribution becomes symmetric that means the skewness of the distribution is zero for any value of n.

(8) If $p > \frac{1}{2}$ then the skewness of the distribution is negative for any value of n.

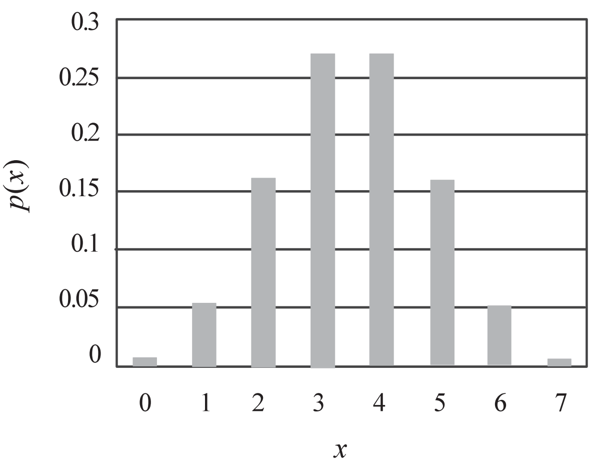
The properties (6), (7) and (8) can be clearly seen from the following graphs :

$p(x)$ for n = 7 and p = 0.2

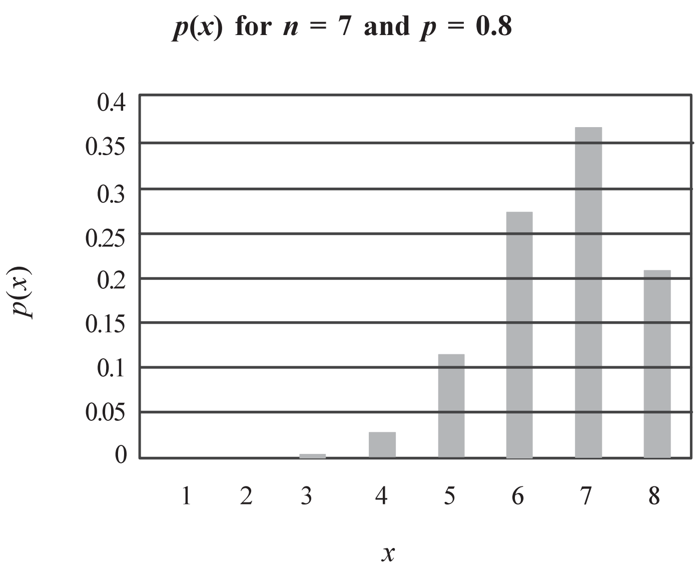


Positively Skewed Binomial Distribution

$p(x)$ for n = 7 and p = 0.5



Symmetric Binomial Distribution



Negatively Skewed Binomial Distribution

2.3.2 Illustrations of Binomial Distribution

Illustration 11 : There are 3 % defective items in the items produced by a factory. 4 items are selected at random from the items produced. What is the probability that there will not be any detective item ?

If the event that the selected items is defective is considered as success then the probability of success $p = 0.03$ and $n = 4$. None of the selected items is defective means $X = 0$.

Now,

$$p(x) = {}^nC_x p^x q^{n-x}, \ x = 0, 1, 2, ..., n$$

Putting the values of $n, p, q = 1-p$ and x in the formula,

$$P(X = 0) = {}^4c_0 (0.03)^0 \ (0.97)^{4-0}$$

$$= (0.97)^4$$

$$= 0.8853$$

Thus, the probability of getting no defective item in the selected 4 items is 0.8853.

Illustration 12 : The probability that a person living in a city is a non-vegetarian is 0.20. Find the probability of at the most two persons out of 6 persons randomly selected from the city is non-vegetarian.

If we consider the event that a person is non-vegetarian as success then we are given the probability of success $p = 0.20$ and $n = 6$.

If we take $X =$ number of non-vegetarians among the selected persons then the probability of $X \le 2$

is obtained by putting the values of n, p and x in the formula of binomial probability distribution

$$p(x) = {}^nC_x p^x q^{n-x},\ x = 0, 1, 2, ..., n$$

$$p(X \le 2) = p(X = 0 \text{ or } X = 1 \text{ or } X = 2)$$

$$= p(0) + p(1) + p(2)$$

$$= {}^6C_0 (0.20)^0 (0.80)^6 + {}^6C_1 (0.20)^1 (0.80)^{6-1} + {}^6C_2 (0.20)^2 (0.80)^{6-2}$$

$$= 0.2621 + 6(0.20)(0.3277) + 15(0.04)(0.4096)$$

$$= 0.2621 + 0.3932 + 0.2458$$

$$= 0.9011$$

Illustration 13 : The mean and variance of a binomial distribution are 3.9 and 2.73 respectively. Find the number of Bernoulli trials conducted in this distribution and write $p(x)$.

Here, variance $= npq = 2.73$ and mean $= np = 3.9$.

$$\therefore q = \frac{\text{Variance}}{\text{Mean}} = \frac{2.73}{3.9} = 0.7 \text{ and } p = 1 - q = 0.3$$

$$\text{Now } n = \frac{np}{p} = \frac{\text{Mean}}{p} = \frac{3.9}{0.3} = 13$$

Thus, the number of Bernoulli trials conducted in this distribution is 13. Since $n = 13$, $p = 0.3$ and $q = 0.7$ in the distribution, its $p(x)$ can be written as follow :

$$p(x) = {}^{13}C_x (0.3)^x (0.7)^{13-x},\ x = 0, 1, 2, ..., 13.$$

Illustration 14 : During a war, on an average one ship out of 9 got sunk in a certain voyage. Find the probability that exactly 5 out of a convoy of 6 ships would arrive safely.

Suppose X = the number of ships that arrive safely out of a convoy of 6 ships during a war.

$n =$ total number of ships in a convoy $= 6$

$p =$ probability that a ship arrives safely in a certain voyage $= \frac{8}{9}$

$\therefore$ The probability that exactly 5 out of a convoy of 6 ships would arrive safely can be obtained by putting corresponding values in the formula

$$p(x) = {}^nC_x p^x q^{n-x},\ x = 0, 1, 2, ..., n,$$

$$p(5) = {}^6C_5 \left(\frac{8}{9}\right)^5 \left(\frac{1}{9}\right)^1$$

$$= 6\left(\frac{32,768}{59,049}\right)\left(\frac{1}{9}\right)$$

$$= \frac{196608}{531441}$$

$$= 0.3700$$

Illustration 15 : Assume that on an average one line out of 4 telephone lines remains busy between 2 pm and 3 pm on week days. Find the probability that out of 6 randomly selected telephone lines (i) not more than 3 (ii) at least three of them will be busy.

Suppose $p =$ the probability of the event that the selected telephone line remains busy between 2 pm to 3 pm $= \frac{1}{4}$

and $X =$ the number of busy telephone lines out of 6 telephone lines between 2 pm to 3 pm.

It is given here that $n = 6$.

(i) The event that not more than 3 lines out of 6 randomlly selected telephone lines will be busy is the event that 3 or less telephone lines will be busy.

That is $X \le 3$.

$\therefore$ To find probability of this event we use the formula of binomial probability distribution

$$p(x) = {}^nC_x p^x q^{n-x}, \; x = 0, 1, 2, ..., n$$

$\therefore \quad P(X \le 3)$

$= P\,(X = 0 \text{ or } 1 \text{ or } 2 \text{ or } 3)$

$= 1 - P\,(X = 4 \text{ or } 5 \text{ or } 6)$

$= 1 - \left[p\,(4) + p(5) + p(6)\right]$

$$= 1 - \left[{}^6C_4 \left(\tfrac{1}{4}\right)^4 \left(\tfrac{3}{4}\right)^2 + {}^6C_5 \left(\tfrac{1}{4}\right)^5 \left(\tfrac{3}{4}\right)^1 + {}^6C_6 \left(\tfrac{1}{4}\right)^6 \left(\tfrac{3}{4}\right)^0\right]$$

$$= 1 - \left[15\left(\tfrac{1}{256}\right)\left(\tfrac{9}{16}\right) + 6\left(\tfrac{1}{1024}\right)\left(\tfrac{3}{4}\right) + \left(\tfrac{1}{4096}\right)\right]$$

$$= 1 - \left[\tfrac{135}{4096} + \tfrac{18}{4096} + \tfrac{1}{4096}\right]$$

$$= 1 - \tfrac{154}{4096} = \tfrac{3942}{4096} = 0.9624$$

(ii) The probability that at least 3 telephone lines will be busy

$= P(X \ge 3)$

$= P\,(X = 3 \text{ or } 4 \text{ or } 5 \text{ or } 6)$

$= p(3) + p\,(4) + p(5) + p(6)$

Now, from the above calculations we will get the values of $p\,(4)$, $p(5)$ and $p(6)$. So, we first find the value of $p(3)$.

$$p(3) = {}^6C_3 \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^3 = 20\left(\frac{1}{64}\right)\left(\frac{27}{64}\right)$$

$$= \frac{540}{4096}$$

Now, from the values of $p(4)$, $p(5)$ and $p(6)$ in question (i) and the value of $p(3)$ we obtained,

$$P(X \ge 3) = \frac{540}{4096} + \frac{135}{4096} + \frac{18}{4096} + \frac{1}{4096}$$

$$= \frac{694}{4096} = 0.1694$$

Illustration 16 : The parameters of binomial distribution of a random variable X are $n = 4$ and $p = \frac{1}{3}$. State the probability distribution of X in a tabular form and hence find the value of $P(X \le 2)$.

Here, parameters are $n = 4$ and $p = \frac{1}{3}$ $\therefore q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

Substituting the values of the parameters in the formula of binomial distribution,

$p(x) = {}^nC_x\, p^x q^{n-x}$, $x = 0, 1, 2, \ldots, n$ we have $p(x) = {}^4C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{4-x}$, $x = 0, 1, 2, 3, 4$.

Now, we calculate the values of $p(x)$ by putting the different values of x as 0, 1, 2, 3 and 4.

$$p(0) = {}^4C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{4-0} = \frac{16}{81}$$

$$p(1) = {}^4C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^{4-1} = 4\left(\frac{1}{3}\right)\left(\frac{8}{27}\right) = \frac{32}{81}$$

$$p(2) = {}^4C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{4-2} = 6\left(\frac{1}{9}\right)\left(\frac{4}{9}\right) = \frac{24}{81}$$

$$p(3) = {}^4C_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^{4-3} = 4\left(\frac{1}{27}\right)\left(\frac{2}{3}\right) = \frac{8}{81}$$

$$p(4) = {}^4C_4 \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^{4-4} = 1\left(\frac{1}{81}\right)1 = \frac{1}{81}$$

These can be put in the tabular form as follows :

$X = x$	0	1	2	3	4	Total
$p(x)$	$\frac{16}{81}$	$\frac{32}{81}$	$\frac{24}{81}$	$\frac{8}{81}$	$\frac{1}{81}$	1

Now, $P(X \le 2)$

$= p(X=0)+p(X=1)+p(X=2)$

$= \frac{16}{81}+\frac{32}{81}+\frac{24}{81}$

$= \frac{72}{81}$

$= \frac{8}{9}$

Illustration 17 : In a binomial distribution, for $P(X=x)=p(x)$, $n=8$ and $2p(4)=5p(3)$. Find the probability of getting success in all the trials for this distribution.

Here, we have $2p(4)=5p(3)$ and $n=8$

$\therefore$ Putting $n=8$ in the formula of binomial distribution, we get,

$p(x)={}^8C_x p^x q^{8-x}$, $x = 0, 1, 2,....., 8$

Putting the values of $p(4)$ and $p(3)$ from this formula in the given condition

$2p(4)=3p(3)$

$2\times {}^8C_4\, p^4 q^{8-4} = 5\times {}^8C_3 p^3 q^{8-3}$

$\therefore\ 2\times(70)\,p^4 q^4 = 5\times(56)\,p^3 q^5$

$\therefore\ 140\,p^4 q^4 = 280\,p^3 q^5$

$\therefore\ p=2q$

$\therefore\ p=2(1-p)$

$\therefore\ p=2-2p$

$\therefore 3p=2$

$\therefore\ p=\frac{2}{3}$ and $q=1-p=1-\frac{2}{3}=\frac{1}{3}$.

Now, getting success in all the trials means the event of getting 8 successes since we have total 8 trials..

The probability of this events is $p(8)$.

$\therefore\ p(8)={}^8C_8\left(\frac{2}{3}\right)^8\left(\frac{1}{3}\right)^{8-8}$

$=1\times\left(\frac{2}{3}\right)^8\times 1$

$=\frac{256}{6561}$

Thus, the probability of getting success in all the trials is $\frac{256}{6561}$.

Illustration 18 : For a binomial distribution, mean = 18 and variance = 4.5. Determine whether the skewness of this distribution is positive or negative.

Here, mean $= np = 18$ and variance $= npq = 4.5$

$\therefore q = \frac{\text{Variance}}{\text{Mean}} = \frac{4.5}{18} = 0.25 = \frac{1}{4}$

$\therefore p = 1 - \frac{1}{4} = \frac{3}{4}$

Since the value of p is greater than $\frac{1}{2}$, the skewness of binomial distribution will be negative.

Illustration 19 : A balacned die is tossed 7 times. If the event of getting a number 5 or more is called success and X denotes the number of success in 7 trials then (i) Write the probability distribution of X. (ii) Find the probability of getting 4 successes. (iii) Find the probability of getting at the most 6 successes.

The sample space associated with tossing of a balanced die once is $U = \{1, 2, 3, 4, 5, 6\}$ and probability of getting each number is $\frac{1}{6}$.

If the event of getting a number 5 or more is called success then probability of success

$p =$ probability of getting 5 or 6 on the die

$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

$\therefore q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

Here, total number of trials is 7. $\therefore n = 7$

(i) Using the probability distribution of X $p(x) = {}^nC_x p^x q^{n-x}$, $x = 0, 1, 2, ..., n$,

$$p(x) = {}^7C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{7-x}, \quad x = 0, 1, 2, 3, 4, 5, 6, 7$$

(ii) Probability of getting 4 successes

$$p(4) = {}^7C_4 \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^{7-4}$$

$$= 35\left(\frac{1}{81}\right)\left(\frac{8}{27}\right)$$

$$= \frac{280}{2187}$$

(iii) Probability of getting at the most 6 successes

$= p(X \le 6)$

$= 1 - p(X > 6)$

$= 1 - p(X = 7) \quad \because x = 0, 1, 2, ..., 7$

$= 1 - {}^{7}C_{7} \left(\frac{1}{3}\right)^{7} \left(\frac{2}{3}\right)^{7-7}$

$= 1 - \frac{1}{2187} = \frac{2186}{2187}$

Illustration 20 : A social worker claims that 10 % of the young children in a city have vision problem. A sample survey agency takes a random sample of 10 young children from the city to test the claim. If at the most one young child is affected by the vision problem, the claim of the social worker is rejected. Find (i) the probability that the claim of the social worker is rejected (ii) the expected number of young children having vision problem in the randomly selected 10 young children.

Suppose p = probability that a young child has eye problem

$= 0.10$ (by accepting the claim of social worker)

And X = the number of young childrens having eye problem in the randomly selected 10 young children

Here, putting $n = 10$ and $p = 0.10$ in the formula

$p(x) = {}^{n}C_{x}\ p^{x}\ q^{n-x}, \quad x = 0, 1, 2, ..., n$

of binomial distribution,

$p(x) = {}^{10}C_{x} (0.10)^{x} (0.90)^{10-x}, x = 0, 1, 2, ..., 10$

(i) Probability that at the most one young child has eye problem

$= p(0) + p(1)$

$= {}^{10}C_{0} (0.10)^{0} (0.90)^{10-0} + {}^{10}C_{1} (0.10)^{1} (0.90)^{10-1}$

$= 0.3487 + 10 (0.10) (0.3874)$

$= 0.3487 + 0.3874$

$= 0.7361$

Now, sample survey agency rejects the claim of the social worker if at the most one young child has eye problem.

$\therefore$ Probability of rejecting the claim of social worker by the sample survey agency $= 0.7361$.

(ii) The expected number of young children having eye problem in the randomly selected 10 young child

$= E(X) = np$

$= 10 \times$ probability that the selected young child has eye problem

$= 10 \times 0.10$

$= 1$

Illustration 21 : An experiment is conducted to toss five balanced coins simultaneously. If we consider occurrence of head (H) on the coin as success then obtain probability distribution of the number of successes. If such an experiment is repeated 3200 times then obtain expected frequency distribution of the number of successes. For this distribution, obtain expected value of the number successes and also obtain its standard deviation.

Since the coins are balanced, probability of getting head will be $\frac{1}{2}$.

$p =$ probability of success

$=$ probability of getting head $= \frac{1}{2}$.

$\therefore q = 1 - p = \frac{1}{2}$

Here, $n =$ number of coins $= 5$, $x =$ the number of successes in tossing of five coins.

Putting the values of n, p and q in the formula

$p(x) = {}^nC_x\, p^x q^{n-x}$, $x = 0, 1, 2, \ldots, n$

of the binomial distribution

$$p(x) = {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}, \quad x = 0, 1, 2, \ldots, 5$$

$$= {}^5C_x \left(\frac{1}{2}\right)^{x+5-x}$$

$$= {}^5C_x \left(\frac{1}{2}\right)^5$$

$$= \frac{{}^5C_x}{32}, \quad x = 0, 1, 2, \ldots, 5$$

Now, using the above formula, we calculate the probability for each x and the frequency for the number of successes in 3200 repetition of the experiment = $3200 \times p(x)$, $x = 0, 1, 2, ...,5$.

We present the calculations in the following table :

x	$p(x)$	Expected Frequency $= N \times p(x)$
0	$\frac{{}^5C_0}{32} = \frac{1}{32}$	$3200 \times \frac{1}{32} = 100$
1	$\frac{{}^5C_1}{32} = \frac{5}{32}$	$3200 \times \frac{5}{32} = 500$
2	$\frac{{}^5C_2}{32} = \frac{10}{32}$	$3200 \times \frac{10}{32} = 1000$
3	$\frac{{}^5C_3}{32} = \frac{10}{32}$	$3200 \times \frac{10}{32} = 1000$
4	$\frac{{}^5C_4}{32} = \frac{5}{32}$	$3200 \times \frac{5}{32} = 500$
5	$\frac{{}^5C_5}{32} = \frac{1}{32}$	$3200 \times \frac{1}{32} = 100$

(ii) Expected value of the number of successes

$= np$

$= 5\left(\frac{1}{2}\right) = 2.5$

(iii) Standard deviation of the number of successes

$= \sqrt{npq}$

$= \sqrt{5 \times \left(\frac{1}{2}\right) \times \left(\frac{1}{2}\right)} = \sqrt{\frac{5}{4}} = \sqrt{1.25}$

$= 1.118$

Illustration 22 : An advertisement company claims that 4 out of 5 house wives do not identify the difference between two different brands of butter. To check the claim, 5000 house wives are divided in groups, each group of 5 house wives. If the claim is true, in how many groups among these groups (i) at the most one house wife (ii) only two house wives can identify the difference between two different brands of butter ?

As per the claim made by an advertising company, 4 out of 5 house wives do not identify the difference between two different brands of butter.

$\therefore$ Its probability is $\frac{4}{5}$

That means the probability of identifying the difference between two different brands of butter by a house wife $= \frac{1}{5}$.

Let p = probability that the selected house wife can identify the difference between two different brands of butter $= \frac{1}{5}$.

To test the claim, selected 5000 housewives are divided in groups randomly, with each group having 5 house wives. So, there will be 1000 such groups.

If we take $X =$ the number of house wives who identify the difference between two different brands of butter in a group, $x = 0, 1, ..., 5$.

Thus, we have $n = 5$, $p = \frac{1}{5}$, $q = \frac{4}{5}$.

Putting the above values in the formula of binomial distribution

$$p(x) = {}^nC_x \ p^x q^{n-x}, \ x = 0, 1, 2, ..., n$$

We get the following $p(x)$

$$p(x) = {}^5C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{5-x}, \quad x = 0, 1, 2, ..., 5$$

Using this formula, we calculate the probabilities for different values of x and multiplying such probabilities by 1000 we get the number of groups out of 1000 groups in which 0, 1, 2, 3, 4 or 5 house wives can identify the difference between two different brands of the butter..

(i) The number of groups out of 1000 groups in which at the most one house wife can identify the difference between two different brands of butter

$$= 1000 \times [p(0) + p(1)]$$

$$= 1000 \times \left[{}^5C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{5-0} + {}^5C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^{5-1} \right]$$

$$= 1000 \times \left[\frac{1024}{3125} + 5 \times \left(\frac{1}{5}\right) \times \left(\frac{256}{625}\right) \right]$$

$$= 1000 \times \left[\frac{1024}{3125} + \frac{256}{625} \right]$$

$= 1000 \times [0.32768 + 0.4096]$

$= 1000 \times [0.73728]$

$= 737.28$

$\simeq 737$ groups

(ii) The number of groups out of 1000 groups in which only two house wives can identify the difference between the two brands of butter.

$= 1000 \times p(2)$

$= 1000 \times {}^5C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^{5-2}$

$= 1000 \times 10 \times \left(\frac{1}{25}\right) \times \left(\frac{64}{125}\right)$

$= 1000 \times \frac{640}{3125}$

$= 1000 \times 0.2048$

$= 204.8$

$\simeq 205$ groups

EXERCISE 2.2

1. For a symmetrical binomial distribution with $n = 8$, find $p(X \le 1)$.

2. Mean of a binomial distribution is 5 and its variance is equal to the probability of success. Find the parameters of this distribution and hence find the probability of the event of getting none of the failures for this distribution.

3. A person has kept 4 cars to run on rent. The probability that any car is rented during the day is 0.6. Find the probability that more than one but less than 4 cars are rented during a day.

4. There are 200 farms in a Taluka. Among the bore wells made in these 200 farms of the Taluka, salted water is found in 20 farms. Find the probability of the event of not getting salted water in 3 out of 5 randomly selected farms from the Taluka.

5. An example is given to 6 students to solve. The probability of getting correct solution of the problem by any student is 0.6. Students are trying to solve the problem independently. Find the probability of getting the correct solution by only 2 out of the 6 students.

*

Summary

- **Random Variable :** A function associating a real number with each outcome of the sample space of a random experiment is called random variable.
- **Discrete Random Variable :** A random variable X which can assume a finite or countable infinite number of values in the set R of real numbers is called a discrete random variable.
- **Continuous Random Variable :** A random variable X which can assume any value in R, the set of real numbers or in any interval of R is called continuous random variable.
- **Discrete Probability Distribution :** Suppose $X : U \to R$ is a random variable which assumes all values of a finite set $\{x_1, x_2, ..., x_n\}$ of R. Also suppose X assumes a value x_i with probability $p(x_i)$. If $p(x_i) > 0$ for $i = 1, 2, ..., n$ and $\Sigma\, p(x_i) = 1$ then the set of real values $\{x_1, x_2, ..., x_n\}$ and $\{p(x_1), p(x_2), ..., p(x_n)\}$ is called the discrete probability distribution of a random variable X which is expressed in a tabular form as follows :

$X = x$	x_1	x_2	....	x_i	...	x_n	Total
$p(x)$	$p(x_1)$	$p(x_2)$	....	$p(x_i)$	...	$p(x_n)$	1

Here $0 < p(x_i) < 1,\ i = 1, 2, ..., n$

- **Bernoulli Trials :** Suppose dichotomous random experiment has two outcomes success (S) and failure (F). If this experiment is repeated under identical conditions and the probability $p(0 < p < 1)$ of getting success at each trial is constant then such trials are called Bernoulli Trials.
- **Binomial Random Variable :** Suppose X denotes the number of successes in the sequence of success (S) and failure (F) obtained in n Bernoulli trials then X is called a binomial random variable.
- **Binomial Probability Distribution :** The probability distribution of a binomial random variable X is called binomial probability distribution.

List of Formulae

(1) Mean of discrete probability distribution $= \mu$

$= E(X)$

$= \Sigma\, x\, p(x)$

(2) Variance of discrete probability distribution $= \sigma^2$

$= V(X)$

$= E(X^2) - (E(X))^2$

where $E(X^2) = \Sigma\, x^2\ p(x)$

(3) Binomial Probability Distribution

$$P(X = x) = p(x) = {}^nC_x\, p^x q^{n-x},\ x = 0,\ 1,\ 2,\ ....,n.$$
$$0 < p < 1,\ q = 1 - p$$

(4) Mean of binomial probability distribution $= np$

(5) Variance of binomial probability distribution $= npq$

(6) Standard deviation of binomial probability distribution $= \sqrt{npq}$

(7) If an experiment having Bernoulli trials repeats N times and $p(x)$ is the probability of getting x successes in the experiment then expected frequency of the number of successes in N repetitions $= N \cdot p(x)$

EXERCISE 2

Section A

Find the correct option for the following multiple choice questions :

1. Which variable of the following will be an illustration of discrete variable ?

(a) Height of a student (b) Weight of a student

(c) Blood Pressure of a student (d) Birth year of a student

2. Which variable of the following will be an illustration of continuous variable ?

(a) Number of accidents occuring at any place

(b) Number of rainy days during a year

(c) Maximum temperature during a day

(d) Number of children in a family

3. A random variable X assume the values -1, 0 and 1 with respective probability $\frac{1}{5}$, K and $\frac{1}{3}$, where $0 < K < 1$ and X does not assume any value other than these values. What will be the value of $E(X)$?

(a) $\frac{2}{5}$ (b) $\frac{3}{5}$ (c) $\frac{2}{15}$ (d) $\frac{3}{15}$

4. A random variable X assumes the values -2, 0 and 2 only with respective probabilities $\frac{1}{5}, \frac{3}{5}$ and K. If $0 < K < 1$, what will be the value of K ?

(a) $\frac{1}{5}$ (b) $\frac{4}{5}$ (c) $\frac{2}{5}$ (d) $\frac{3}{5}$

5. Mean and variance of a discrete probability distribution are 3 and 7 respectively. What will be $E\left(X^2\right)$ for this distribution ?

(a) 10 (b) 4 (c) 40 (d) 16

6. For the probability distribution of a discrete random variable, $E(X)=5$ and $E\left(X^2\right)=35$. What will be the variance of this distribution ?

(a) 40 (b) 30 (c) 20 (d) 10

7. For a positively skewed binomial distribution with $n = 10$, which of the following values might be the value of mean ?

(a) 5 (b) 3 (c) 9 (d) 7

8. For which value of x, the value of $p(x)$ of binomial distribution with parameters $n = 4$ and $p = \frac{1}{2}$ becomes maximum ?

(a) 0 (b) 2 (c) 3 (d) 4

9. The binomial distribution has mean 5 and variance $\frac{10}{7}$. What will be the type of this distribution ?

(a) Positively skewed (b) Negatively skewed

(c) Symmetric (d) Nothing can be said about the distribution

10. Which of the following is the formula of probability of an event of not getting a success in the binomial distribution with parameters n and p ?

(a) ${}^nC_0 p^n q^0$ (b) ${}^nC_0 p^0 q^n$ (c) ${}^nC_0 pq^n$ (d) ${}^nC_0 p^n q$

Section B

Answer the following questions in one sentence :

1. Define discrete random variable.
2. Define continuous random variable.
3. Define discrete probability distribution.
4. State the formula to find mean of discrete variable.
5. State the formula to find variance of discrete variable.
6. Mean of a symmetrical binomial distribution is 7. Find the value of its parameter n.
7. The parameters of a binomial distribution are 10 and $\frac{2}{5}$. Calculate its variance.
8. State the relation between the probability of success and failure in Bernoulli trials.
9. State the relation between mean and variance of binomial distribution.
10. The probability of failure in a binomial distribution is 0.6 and the number of trials in it is 5. Find the probability of success.

Section C

Answer the following questions :

1. The probability distribution of a random variable X is as follows :

X	2	3	4	5
$p(x)$	0.2	0.3	$4C$	C

Determine the value of constant C.

2. Calculate mean of the discrete probability distribution $p(x) = \begin{cases} \frac{x-1}{6}; & x = 2, 3 \\ \frac{1}{2}; & x = 4 \end{cases}$

3. The probability distribution of a random variable is as follows :

$p(x) = \frac{x+3}{10}, \quad x = -2, 1, 2$

Hence calculate $E(X^2)$.

4. If $n=4$ for a symmetrical binomial distribution then find $p(4)$.

5. Define Bernoulli trials.

6. For a binomial distribution, if probability of success is double the probability of failure and $n=4$ then find variance of the distribution.

7. Find the standard deviation of the binomial distribution having $n=8$ and probability of failure $\frac{2}{3}$.

8. Find parameters of the binomial distribution where mean = 4 and variance = 2.

9. For a binomial distribution with $n=10$ and $q-p=0.6$, find mean of this distribution.

10. For a binomial distribution, standard deviation is 0.8 and probability of failure is $\frac{2}{3}$, find the mean of this distribution.

Section D

Answer the following questions :

1. The probability distribution of a random variable X is as follows :

$$p(x)=\begin{cases} K(x-1); & x=2, 3 \\ K; & x=4 \\ K(6-x)\ ; & x=5 \end{cases}$$

Find the value of constant K and the probability of the event that variable X assumes even numbers.

2. The probability distribution of a random variable X is as follows :

$p(x)=C(x^2+x),\ x=-2, 1, 2$

Find the value of C and show that $p(2)=3p(-2)$.

3. The distribution of a random variable X is $p(x)=K\cdot{}^5P_x,\ x=0, 1, 2, 3, 4, 5$

Find constant K and mean of this distribution.

4. What is discrete probability distribution ? State its properties.

5. State properties of binomial distribution.

6. In a game of hitting a target, the probability that Ramesh will fail in hitting the target is $\frac{2}{5}$. If he is given 3 trials to hit the target, find the probability of the event he hits the target successfully in 2 trials. State mean of this distribution.

7. A person is asked to select a number from positive integers 1 to 7. If the number selected by him is odd then he is entitled to get the prize. If he is asked to take 5 trials then find the probability of the event that he will be entitled to get a prize in only one trial.

8. The mean and variance of the binomial distribution are 2 and $\frac{6}{5}$ respectively. Find $p(1)$ and $p(2)$ for this binomial distribution.

9. 10 % apples are rotten in a box of apples. Find the probability that half of the 6 apples selected from the box with replacement will be rotten and find the variance of the number of rotten apples.

Section E

Solve the following :

1. The probability distribution of the monthly demand of laptop in a store is as follows :

Demand of laptop	1	2	3	4	5	6
Probability	0.10	0.15	0.20	0.25	0.18	0.12

Determine the expected monthly demand of laptop and find variance of the demand.

2. Two dice are thrown simultaneously once. Obtain the discrete probability distribution of the number of dice for which the number '6' comes up.

3. If the probability that any 50 year old person will die within a year is 0.01, find the probability that out of a group of 5 such persons

(i) none of them will die within a year

(ii) at least one of them will die within a year.

4. The probability that a student studying in 12th standard of science stream will get admission to engineering branch is 0.3. 5 students are selected from the students who studied in this stream. Find the probability of the event that the number of students admitted to engineering branch is more than the number of students who did not get admission to the engineering branch.

5. The probability that a bomb dropped from a plane over a bridge will hit the bridge is $\frac{1}{5}$. Two bombs are enough to destroy the bridge. If 6 bombs are dropped on the bridge, find the probability that the bridge will be destroyed.

6. Normally, 40 % students fail in one examination. Find the probability that at least 4 students in a group of 6 students pass in this examination.

7. There are 3 red and 4 white balls in a box. Four balls are selected at random with replacement from the box. Find the probability of the event of getting (i) 2 red balls and 2 white balls (ii) all four white balls among the selected balls using binomial distribution.

Section F

Solve the following :

1. There are one dozen mangoes in a box of which 3 mangoes are rotten. 3 mangoes are randomly selected from the box without replacement. If X denotes the number of rotten mangoes in the selected mangoes, obtain the probability distribution of X and hence find expected value and variance of the rotten mangoes in the selected mangoes.

2. It is known that 50 % of the students studying in the 10th standard have a habit of eating chocolate. To examine the information, 1024 investigators are appointed. Every investigator randomly selects 10 students from the population of such students and examines them for the habit of eating chocolate. Find the expected number of investigators who inform that less than 30 percent of the students have a habit of eating chocolate.

James Bernoulli
(1654 –1705)

James (Jacob) Bernoulli was born in Basel, Switzerland. He was one of the many prominent mathematicians in the Bernoulli family. Following his father's wish, he studied theology (divinity) and entered the ministry. But contrary to the desires of his parents, he also studied mathematics and astronomy. He travelled throughout Europe from 1676 to 1682; learning about the latest discoveries in mathematics and the sciences under leading figures of the time. He was an early proponent of Leibnizian calculus and had sided with Leibniz during the Leibniz-Newton calculus controversy. He is known for his numerous contributions to calculus, and along with his brother Johann, was one of the founders of the calculus of variations. However, his most important contribution was in the field of probability, where he derived the first version of the law of large numbers. He was appointed as professor of mathematics at the University of Basel in 1687, remained in this position for the rest of his life.

●

"Normal Distribution is father of all probability distributions. For larger sample size almost all theoritical distributions follow normal distribution".

– Unknown

3

Normal Distribution

Contents :

3.1 Normal Distribution : Introduction, Probability Density Function

In the previous chapter, we have studied the probability distribution for a discrete random variable. Now, we shall study the probability distribution for a continuous random variable. We know that, if a random variable X can assume any value of real set R or within any interval of real set R then it is called continuous random variable. If a random variable can assume any value between the definite interval a to b then it is denoted by $a < x < b$. A function for obtaining probability that a continuous random variable assumes value between specified interval is called probability density function of that variable and it satisfies the following two conditions:

(1) The probability that the value of random variable lies within the specified interval is non negative.

(2) The total probability that the random variable assumes any value within the specified interval is one.

Thus, probability density function is used to determine probability that the value of random variable X lies within the specified interval a to b and it is denoted as $P(a < x < b)$. It is necessary to note here that probability for the definite value of continuous random variable X obtained by the probability density function is always zero (0). i.e. $P(x = a) = 0$. Thus, the probabilities $P(a < x < b)$ and $P(a \le x \le b)$ obtained by using probability density function are always equal i.e. $P(a < x < b) = P(a \le x \le b)$.

Normal distribution is very important probability distribution among probability distributions for continuous random variable and is very useful distribution for higher statistical study. It can be defined as under:

If X is a random variable with mean μ and standard deviation σ and if its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; -\infty < x < \infty$$

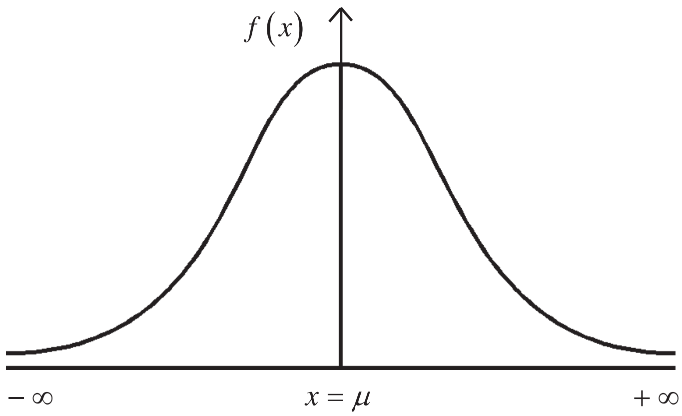
$$-\infty < \mu < \infty$$

$$0 < \sigma < \infty$$

where $\pi = 3.1416$ and $e = 2.7183$ are the constants

then X is called normal random variable and $f(x)$ is called probability density function of normal random variable. The distribution of this normal random variable X is called normal distribution and is denoted by $N(\mu, \sigma^2)$

A curve drawn by considering different values of normal random variable X and its respective values of probability density function $f(x)$ is called normal curve and is shown as under:



As shown in the above diagram, normal curve is completely bell shaped which shows that it is symmetric distribution.

3.2 Standard Normal Variable and Standard Normal Distribution

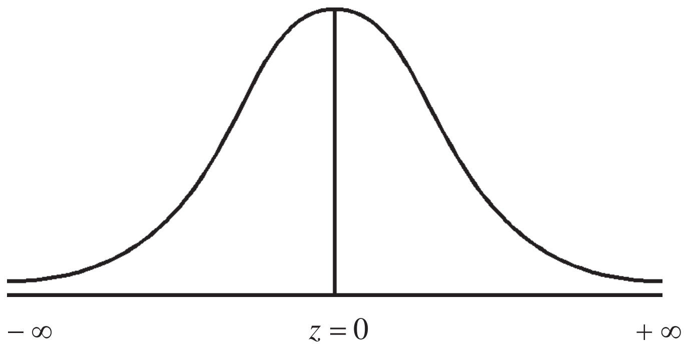
If X is a random normal variable with mean μ and standard deviation σ then random variable $Z = \frac{X-\mu}{\sigma}$ is called standard normal random variable and its probability density function is given below.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \; ; -\infty < z < \infty$$

It can be seen here that probability density function of standard normal variable is a normal density function with mean zero (0) and standard deviation 1.

Note : During the further study of this chapter, we shall call normal variable X instead of normal random variable X and standard normal variable Z instead of standard normal random variable Z.

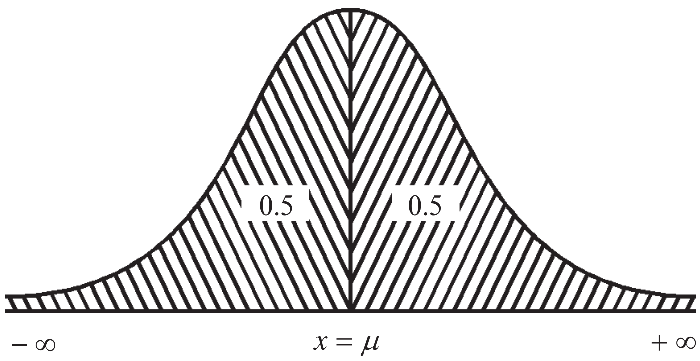
By plotting different values of standard normal variable Z and its respective values of $f(z)$ on graph paper, a completely bell shaped curve is obtained as under :



This curve is called standard normal curve and it is symmetrical to both the sides of $Z = 0$.

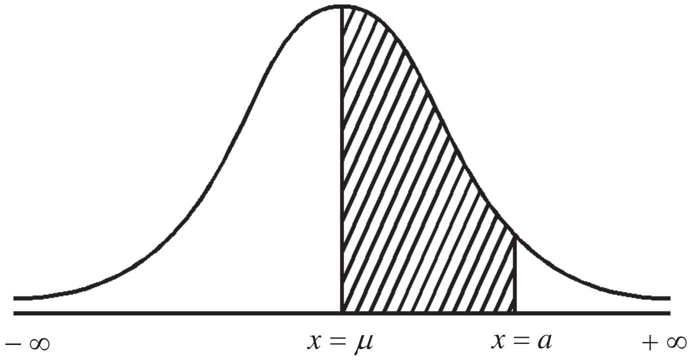
3.3 Method for Finding the Probability (area) from the Tables of Standard Normal Curve

We know that a normal curve is a curve of normal density function and it can be seen as under :



The area (probability) of the shaded region between the curve and X- axis is equal to 1. The normal curve of normal variable X is symmetrical about mean μ on both the sides and hence the perpendicular line at the point $X = \mu$ on X-axis divides the area (probability) of normal curve in two equal parts. The area (probability) to the right side of $X = \mu$ is 0.5 and is denoted by $P(X \geq \mu) = 0.5$, whereas the area (probability) to the left side of $X = \mu$ is 0.5 and is denoted by $P(X \leq \mu) = 0.5$.

In normal curve, probability that value of normal variable X lies between mean μ and its any specific value a $(a > \mu)$ can be shown by the area of shaded region between the x-axis and the perpendicular lines at $X = \mu$ and $X = a$. It can be shown as under :



In notation, this can be shown as $P(\mu \leq X \leq a)$.

For obtaining area under normal curve, first of all the normal variable X is changed into standard normal variable Z. By considering different positive values of standard normal variable Z, a table is prepared for obtaining area under normal curve for 0 to Z and by using this table the area can be obtained.

Note : A table for different values of standard normal variable is given on the last page of the book.

Suppose probability that a normal variable X assumes the value between mean μ and constant a $(a > \mu)$ is to be obtained then it is denoted as $P(\mu \leq X \leq a)$. Now, if standard deviation of normal variable X is σ,

When $X = \mu$ then $Z = \frac{X-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = \frac{0}{\sigma} = 0$ and

When $X = a$ then $Z = \frac{X-\mu}{\sigma} = \frac{a-\mu}{\sigma} = Z_1$

Thus, $P(\mu \leq X \leq a) = P(0 \leq Z \leq Z_1)$

= area between $Z = 0$ to $Z = Z_1$ obtained from tables of standard normal variable.

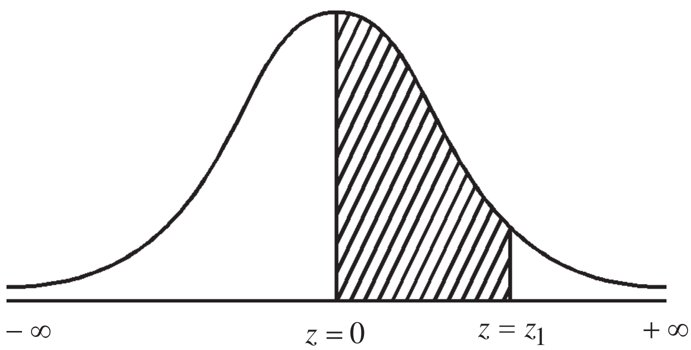


Illustration 1 : A normal distribution has mean 10 and standard deviation 2. Find the probabilities of (1) Normal variable X will take value between 10 and 12. (2) Normal variable X has the value between 8 and 10.

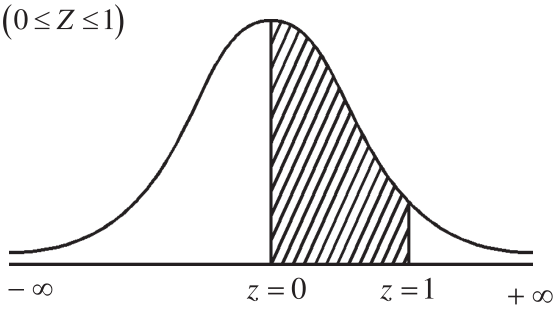
Here, mean $\mu = 10$ and standard deviation $\sigma = 2$.

(1) The probability that a normal variable X will take value between 10 and 12 is to be determined, i.e. to determine $P(10 \le X \le 12)$

$$\therefore P(10 \le X \le 12) = P\left(\frac{10-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{12-\mu}{\sigma}\right)$$

$$= P\left(\frac{10-10}{2} \le Z \le \frac{12-10}{2}\right)$$

$$= P(0 \le Z \le 1)$$



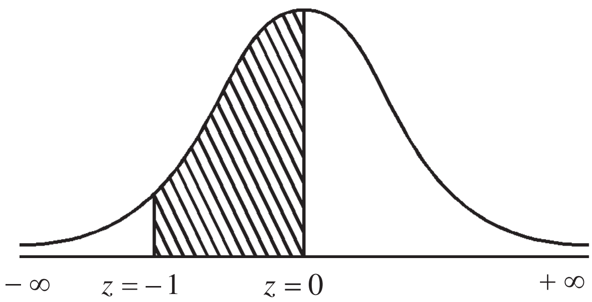
$$= 0.3413 \text{ (from the tables of standard normal variable)}$$

(2) The probability that a normal variable X will take value between 8 and 10 is to be determined, i.e. to determine $P(8 \le X \le 10)$

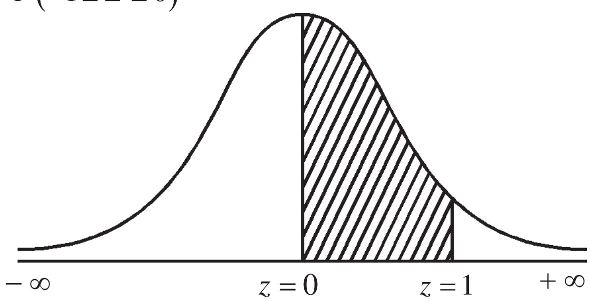
$$\therefore P(8 \le X \le 10) = P\left(\frac{8-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{10-\mu}{\sigma}\right)$$

$$= P\left(\frac{8-10}{2} \le Z \le \frac{10-10}{2}\right)$$

$$= P(-1 \le Z \le 0)$$



=



$$= P(0 \le Z \le 1) \quad (\because \text{ Symmetry})$$

$$= 0.3413 \text{ (from the tables of standard normal variable)}$$

Illustration 2 : A normal distribution has mean 20 and variance 16. Find the probabilities of (1) Normal variable X will take value less than 26 (2) Normal variable X has the value more than 14.

Here, mean $\mu = 20$ and variance $\sigma^2 = 16$.

$\therefore$ standard deviation $\sigma = 4$.

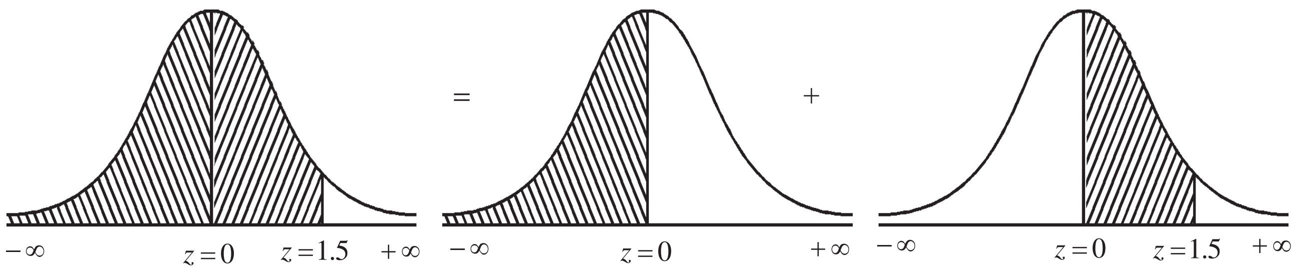
(1) The probability that a normal variable X will take value less than 26

$$= P(X \le 26)$$

$$= P\left(\frac{X-\mu}{\sigma} \le \frac{26-\mu}{4}\right)$$

$$= P\left(Z \le \frac{26-20}{4}\right)$$

$$= P(Z \le 1.5)$$



$= P(-\infty < Z \le 0) + P(0 \le Z \le 1.5)$

$=$ 0.5 + 0.4332 (from the tables of standard normal variable)

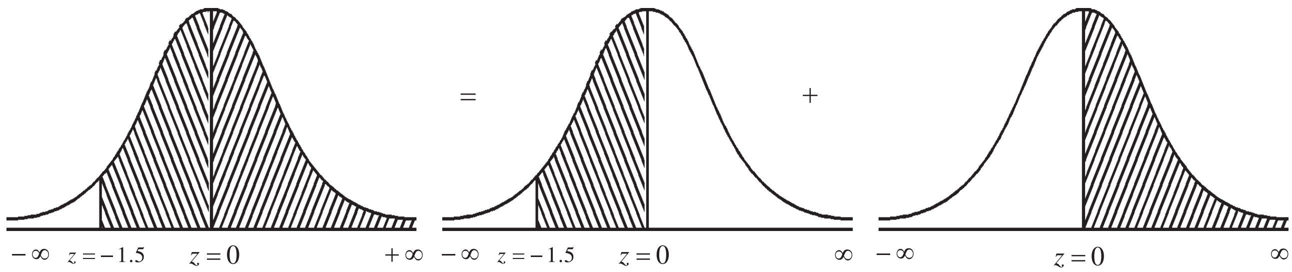
$=$ 0.9332

(2) The probability that a normal variable X will take value more than 14

$= P(X \ge 14) = P\left(\frac{X-\mu}{\sigma} \ge \frac{14-\mu}{\sigma}\right)$

$= P\left(Z \ge \frac{14-20}{4}\right)$

$= P(Z \ge -1.5)$



$= P(-1.5 \le Z \le 0) + P(0 \le Z < \infty)$

$= P(0 \le Z \le 1.5) + 0.5$ ($\because$ Symmetry)

$=$ 0.4332 + 0.5 (from the tables of standard normal variable)

$=$ 0.9332

Illustration 3 : The number of students in classes of higher secondary schools of a city follows normal distribution. Average number of students in the classes is 50 and standard deviation is 15. If a class is selected at random then find the following probabilities (i) a class consists of more than 68 students (ii) a class consists of less than 32 students.

It is given here that the number of students in the class follows normal distribution.

Normal variable $X =$ number of students in a class

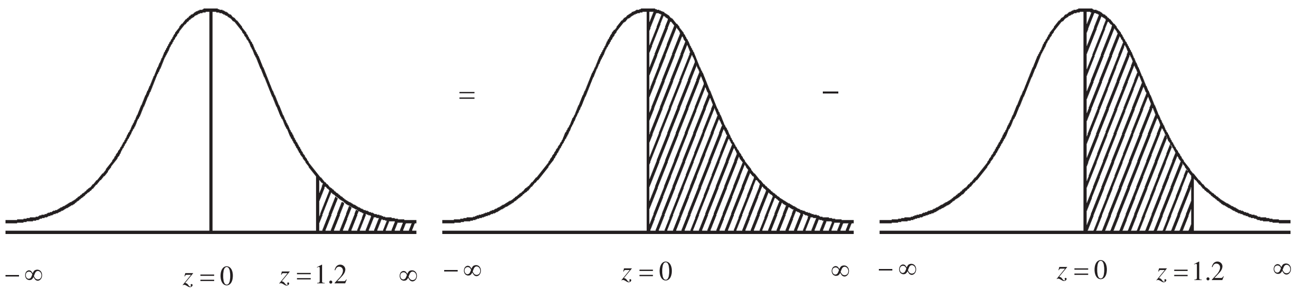
Also, mean $\mu = 50$ students and standard deviation $\sigma = 15$ students

(1) The probability that a randomly selected class consists of more than 68 students

$$= P(X \geq 68) \quad = P\left(\frac{X-\mu}{\sigma} \geq \frac{68-\mu}{\sigma}\right)$$

$$= P\left(Z \geq \frac{68-50}{20}\right)$$

$$= P(Z \geq 1.2)$$



$$= P(0 \leq Z < \infty) - P(0 \leq Z \leq 1.2)$$

= 0.5 – 0.3849 (from the tables of standard normal variable)

= 0.1151

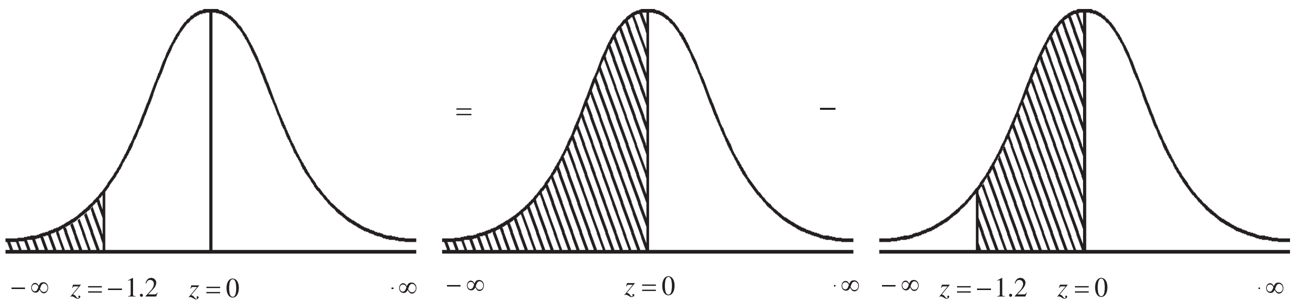
Thus, the probability that a randomly selected class consists of more than 68 students is 0.1151.

(2) The probability that a randomly selected class consists of less than 32 students

$$= P(X \leq 32) \quad = P\left(\frac{X-\mu}{\sigma} \leq \frac{32-\mu}{\sigma}\right)$$

$$= P\left(Z \leq \frac{32-50}{20}\right)$$

$$= P(Z \leq -1.2)$$



$$= P(-\infty < Z \leq 0) - P(-1.2 \leq Z \leq 0)$$

$$= 0.5 - P(0 \leq Z \leq 1.2) \ (\because \text{ Symmetry})$$

= 0.5 – 0.3849 (from the tables of standard normal variable)

= 0.1151

Thus, the probability that a randomly selected class consists of less than 32 students is 0.1151

Illustration 4 : The average weight of grown up children living in a large society is 50 kg and its standard deviation is 5 kg. If their weight follows normal distribution and a grown up child is selected at random then find

(1) the probability that his weight is between 55 kg and 65 kg.

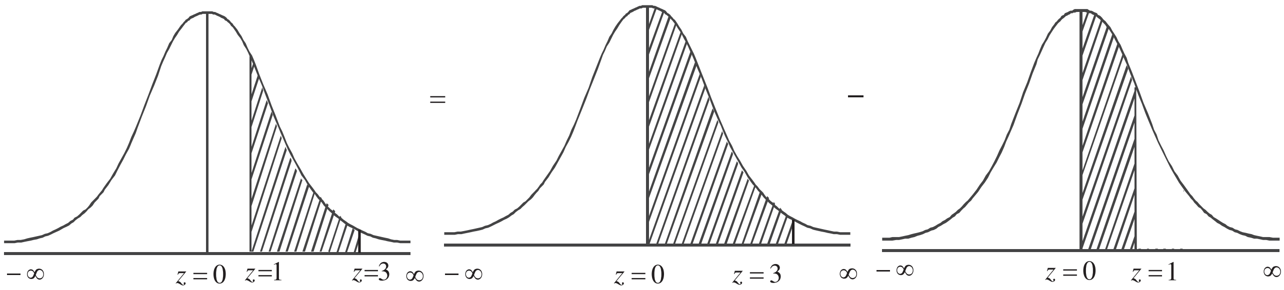
(2) the probability that his weight is between 35 kg and 45 kg.

Normal variable X = weight of a grown up child, average weight $\mu = 50$ kg and standard deviation $\sigma = 5$ kg.

(1) Probability that a randomly selected grown up child has weight between 55 kg to 65 kg

$$= P(55 \le X \le 65) = P\left(\frac{55-50}{5} \le \frac{X-\mu}{\sigma} \le \frac{65-50}{5}\right)$$

$$= P(1 \le Z \le 3)$$



$$= P(0 \le Z \le 3) - P(0 \le Z \le 1)$$

$$= 0.4987 - 0.3413$$

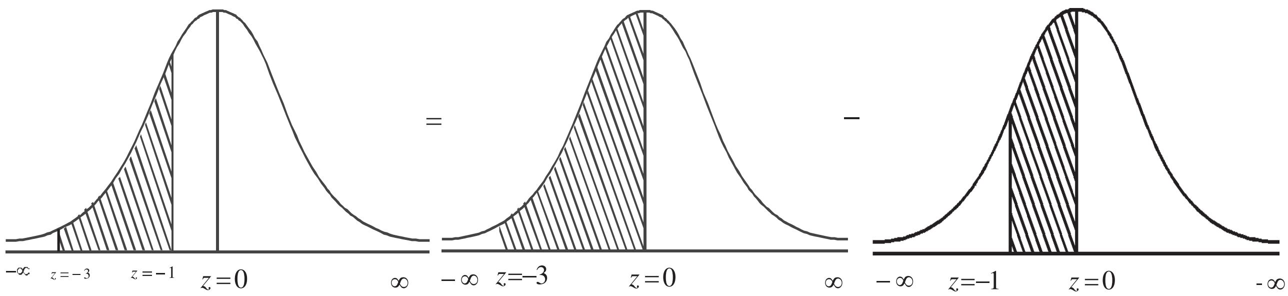
$$= 0.1574$$

Thus, the probability that a randomly selected grown up child has weight between 55 kg to 65 kg is 0.1574.

(2) Probability that a randomly selected grown up child has weight between 35 kg to 45 kg

$$= P(35 \le X \le 45) = P\left(\frac{35-50}{5} \le \frac{X-\mu}{\sigma} \le \frac{45-50}{5}\right)$$

$$= P(-3 \le Z \le -1)$$



$$= P(-3 \le Z \le 0) - P(-1 \le Z \le 0)$$

$$= P(0 \le Z \le 3) - P(0 \le Z \le 1) \quad (\because \text{ symmertry})$$

$$= 0.4987 - 0.3413$$

$$= 0.1574$$

Thus, the probability that a randomly selected grown up child has weight between 35 kg to 45 kg is 0.1574.

Note : From the above illustrations it is clear that the area under the normal curve for $Z = 0$ to $Z = a$ is equal to the area between $Z = -a$ to $Z = 0$. This is because the normal distribution is a symmetric distribution.

Illustration 5 : The monthly income of workers working in a production house follows normal distribution. Their average monthly income is ₹ 15,000 and standard deviation is ₹ 4000.

(1) If a worker is selected at random then find the probability that his monthly income is between ₹ 10,000 and ₹ 25,000.

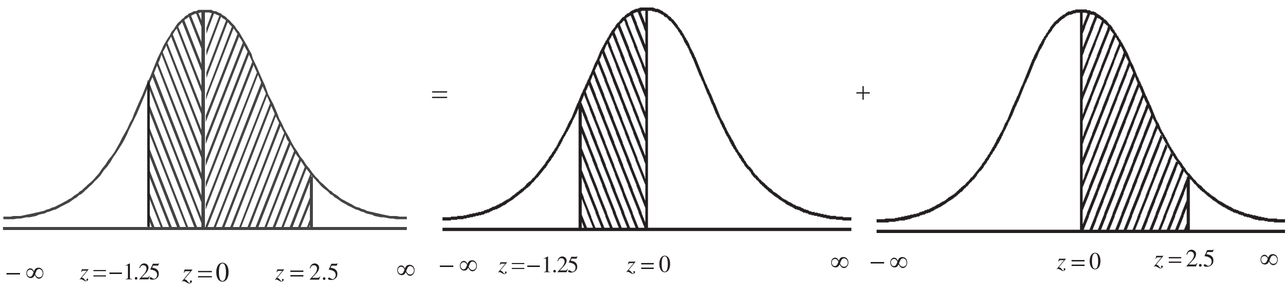
(2) Find the percentage of workers having monthly income between ₹ 12,000 and 22,000 in the production house.

Here, normal variable X = monthly income of worker, average income μ = ₹ 15,000 and standard deviation σ = ₹ 4000.

(1) The probability that a randomly selected worker has income between ₹ 10,000 and ₹ 25,000

$$= P(10000 \le X \le 25000) = P\left(\frac{10000-15000}{4000} \le \frac{X-\mu}{\sigma} \le \frac{25000-15000}{4000}\right)$$

$$= P(-1.25 \le Z \le 2.5)$$



$$= P(-1.25 \le Z \le 0) + P(0 \le Z \le 2.5)$$

$$= P(0 \le Z \le 1.25) + P(0 \le Z \le 2.5) \; (\because \text{ symmertry})$$

$$= 0.3944 + 0.4938$$

$$= 0.8882$$

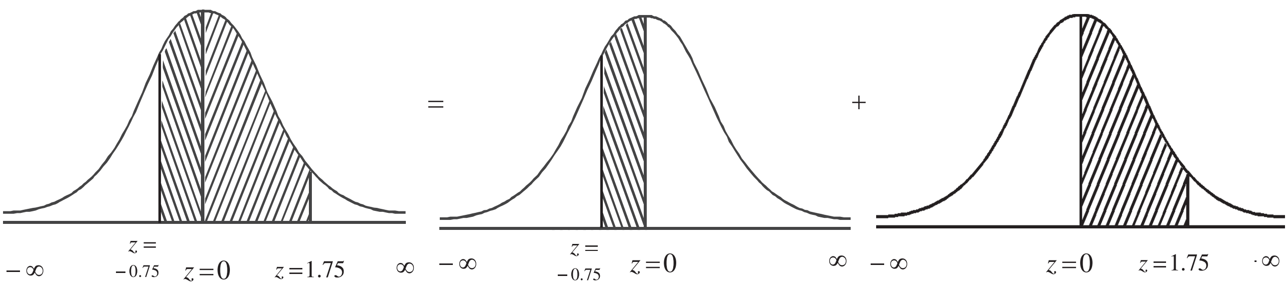
Thus, the probability that a randomly selected worker has monthly income between ₹ 10,000 and ₹ 25,000 is 0.8882.

(2) The probability that a randomly selected worker has income between ₹ 12,000 and ₹ 22,000

$$= P(12000 \le X \le 22000)$$

$$= P\left(\frac{12000-15000}{4000} \le \frac{X-\mu}{\sigma} \le \frac{22000-15000}{4000}\right)$$

$$= P(-0.75 \le Z \le 1.75)$$



$$= P(-0.75 \le Z \le 0) + P(0 \le Z \le 1.75)$$

$$= P(0 \le Z \le 0.75) + P(0 \le Z \le 1.75) \quad (\because \text{ symmertry})$$

$= 0.2734 + 0.4599$

$= 0.7333$

$\therefore$ The percentaged that a randomly selected worker has monthly income between ₹ 12,000 and ₹ 22,000

$= 0.7333 \times 100$

$= 73.33\ \%$

Thus, 73.33 % of the wokers in the production house have monthly income between ₹ 12,000 and ₹ 22,000.

Note : In order to express probability in percentage, probability is multiplied by 100.

3.4 Properties of Normal Distribution

Some important properties of normal distribution are as under :

(1) It is a distribution of continuous random variable.

(2) The constants μ and σ are the parameters of distribution which indicate mean and variance respectively.

(3) The distribution is symmetric about μ and its skewness is zero (0).

(4) For this distribution, the value of mean, median and mode are same. In notation, $\mu = M = M_0$

(5) For this distribution, quartiles are equidistant from median i.e. $Q_3 - M = M - Q_1$ and $M = \frac{Q_3+Q_1}{2}$

(6) The probability curve is completely bell shaped.

(7) Normal curve is asymptotic to X-axis. The tails never touch X-axis.

(8) The approximate value of quartiles of normal distribution can be obtained from the following formula

$Q_1 = \mu - 0.675\ \sigma$

$Q_3 = \mu + 0.675\ \sigma$

(9) For this distribution, quartile deviation = $\frac{2}{3}\sigma$ (approximately)

(10) For this distribution, mean deviation = $\frac{4}{5}\sigma$ (approximately)

(11) Important areas under normal curve are as below :

(i) Total area under normal curve is 1 and area on both the sides of perpendicular line at $X = \mu$ is 0.5

(ii) The area under the curve between the perpendicular lines at $\mu - \sigma$ and $\mu + \sigma$ is 0.6826 i.e. area under the normal curve between the perpendicular lines at $\mu \pm \sigma$ is 0.6826

(iii) The area under the curve between the perpendicular lines at $\mu - 2\sigma$ and $\mu + 2\sigma$ is 0.9545

(iv) The area under the curve between the perpendicular lines at $\mu - 3\sigma$ and $\mu + 3\sigma$ is 0.9973

(v) The area under the curve between the perpendicular lines at $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 0.95.

(vi) The area under the curve between the perpendicular lines at $\mu - 2.575\sigma$ and $\mu + 2.575\sigma$ is 0.99.

3.5 Properties of Standard Normal Distribution

Some important properties of standard normal distribution are as under :

(1) It is a distribution of continuous random variable.

(2) For this distribution, mean is zero (0) and its standard deviation is 1.

(3) The distribution is symmetric to $Z = 0$ and its skewness is zero (0).

(4) The probability curve is completely bell shaped and is asymptotic to X-axis.

(5) The approximate value of the first quartile of standard normal distribution is –0.675 and that of the third quartile is 0.675.

(6) For this distribution, quartile deviation $= \frac{2}{3}$ (approximately).

(7) For this distribution, mean deviation $= \frac{4}{5}$ (approximately).

(8) Important areas under normal curve are as below :

(i) Total area under normal curve is 1 and area on both the sides of perpendicular line at $Z = 0$ is 0.5

(ii) The area under the curve between the perpendicular lines at $Z = -1$ and $Z = +1$ is 0.6826 i.e. area under the normal curve between the perpendicular lines at $Z = \pm 1$ is 0.6826.

(iii) The area under the curve between the perpendicular lines at $Z = -2$ and $Z = +2$ is 0.9545.

(iv) The area under the curve between the perpendicular lines at $Z = -3$ and $Z = +3$ is 0.9973.

(v) The area under the curve between the perpendicular lines at $Z = -1.96$ and $Z = +1.96$ is 0.95

(vi) The area under the curve between the perpendicular lines at $Z = -2.575$ and $Z = +2.575$ is 0.99

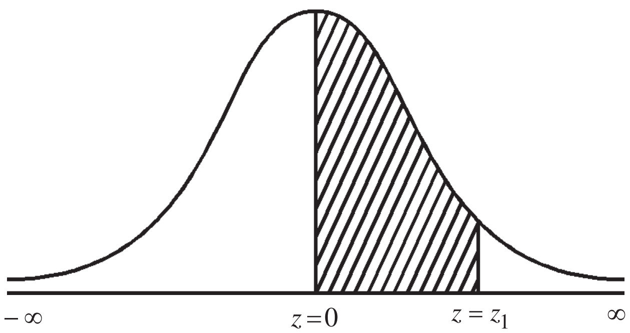
It should be noted here that the probability distribution of standard normal variable Z is a distribution of normal variable with mean zero and variance 1. Z is called standard score or Z-score and it is independent of unit of measurement.

We have seen earlier that when a value of normal variable X and values of parameters are known then corresponding value of Z-score is obtained and by using the table of standard normal variable, the respective probability can be obtained. Now, if the probability is known then to determine value of Z-score we shall study the following illustrations :

Illustration 6 : If the probability that value of standard normal variable Z lies between 0 and Z-score (z_1) is 0.3925 then obtain the possible values of Z-score (z_1).

The probability that value of standard normal variable Z lies between $Z = 0$ and $Z = z_1$ is 0.3925. This probability is equal to the area under the curve between $Z = 0$ and $Z = z_1$. The value of z_1 may be positive or negative.

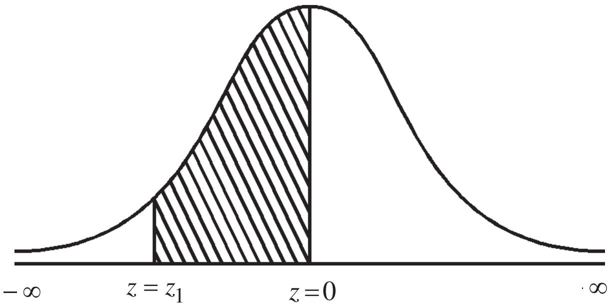
Suppose the value of z_1 is positive then $P(0 \le Z \le z_1) = 0.3925$.



For obtaining the value of z_1, see the first column of table of standard normal variable (Z-table). For $Z = 1.20$, the area is 0.384 which is less than 0.3925. Now, read the values in this row. For the value 0.3925, the corresponding value of Z is 1.24. Therefore, one possible value of Z-score is 1.24.

Now, suppose the value of z_1 is negative

$\therefore P(z_1 \le Z \le 0) = 0.3925$



Now, since the normal distribution is symmetric, $P(z_1 \le Z \le 0) = P(0 \le Z \le z_1) = 0.3925$. Thus, as above, $z_1 = 1.24$ but it can be seen in the above diagram that the perpendicular line at z_1 is to the left of $Z = 0$ hence Z-score $z_1 = -1.24$.

Thus, if the probability that the value of standard normal variable lies between $Z = 0$ and $Z = z_1$ is 0.3925 then the possible values of Z-score are ± 1.24.

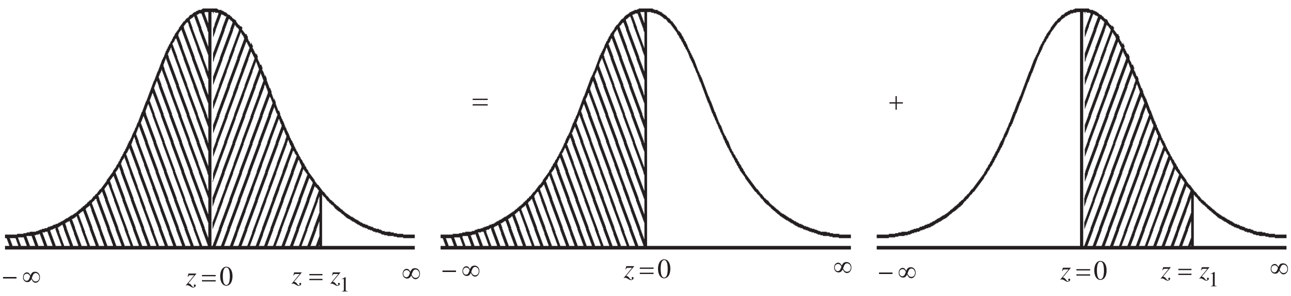
Thus, the sign of z_1 is positive if it is on right hand side of $Z = 0$ and negative if it is on left hand side of $Z = 0$.

Illustration 7 : If the probabilities for standard normal variable Z are as under then obtain the value of Z-socre (z_1) :

(1) Area to the left of $Z = z_1$ is 0.95

(2) Area to the right of $Z = z_1$ is 0.05.

(1) Area to the left of $Z = z_1$ is 0.95 i.e. $P(Z \le z_1) = 0.95$. In the curve of standard normal variable, the perpendicular line at $Z = z_1$ is to be drawn by moving from left to right so that the area under the curve is 0.95. The figure is as under :



$$\text{Thus} \quad P(Z \le z_1) = P(-\infty < Z \le 0) + P(0 \le Z \le z_1) = 0.95$$

$$\therefore 0.5 + P(0 \le Z \le z_1) = 0.95$$

$$\therefore P(0 \le Z \le z_1) = 0.95 - 0.5$$

$$\therefore P(0 \le Z \le z_1) = 0.45$$

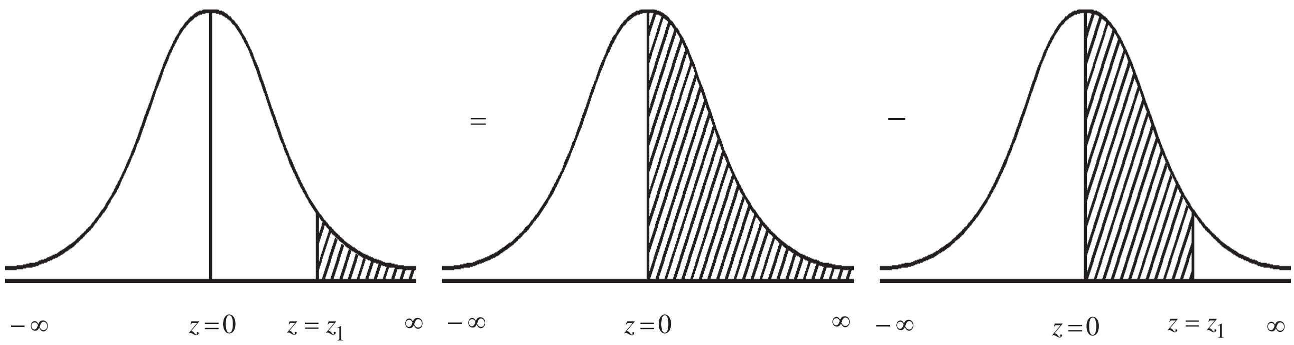
Now, corresponding to probability 0.45, it is not possible to obtain value of z_1 directly from the table of standard normal variable. Hence, the approximate value of z_1 is determined as follows

From table	Area	Z-score
Nearest value before 0.45	0.4495	1.64
Nearest value after 0.45	0.4505	1.65
Average value	0.4500	1.645

From the above table, it can be seen that $z_1 = 1.645$.

Thus, for $z_1 = 1.645$, $P(Z \le z_1) = 0.95$.

(2) Area to the right of $Z = z_1$ is 0.05 i.e. $P(Z \ge z_1) = 0.05$. In the curve of standard normal variable, the perpendicular line at $Z = z_1$ is to be drawn by moving from right to left so that the area under the curve is 0.05. The figure is as under



$$\therefore P(Z \ge z_1) = P(0 \le Z < \infty) - P(0 \le Z \le z_1) = 0.05$$

$$\therefore 0.5 - P(0 \le Z \le z_1) = 0.05$$

$$\therefore P(0 \le Z \le z_1) = 0.45$$

As calculated earlier, $z_1 = 1.645$.

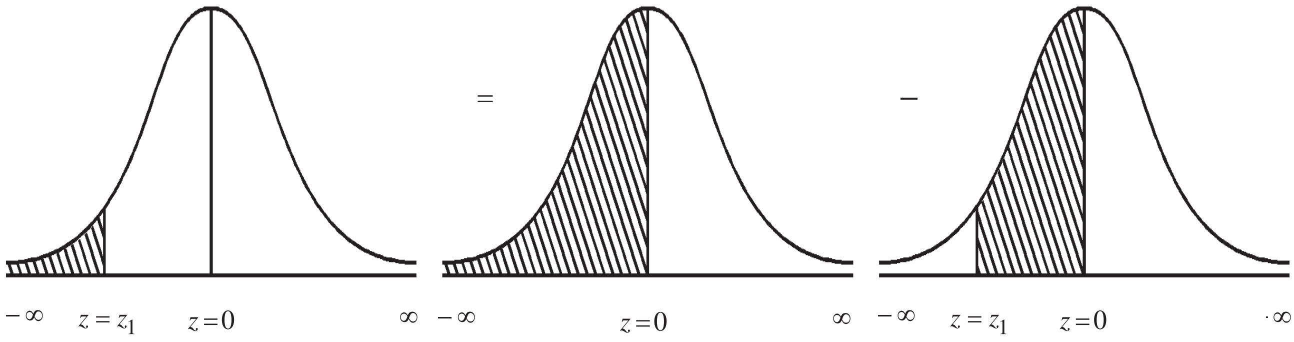
Thus, for $z_1 = 1.645$, $P(Z \ge z_1) = 0.05$.

Illustration 8 : If the probabilities for standard normal variable Z are as under then obtain the value of Z-score (z_1) :

(1) Area to the left of $Z = z_1$ is 0.10.

(2) Area to the right of $Z = z_1$ is 0.90.

(1) Area to the left of $Z = z_1$ is 0.10 i.e. $P(Z \le z_1) = 0.10$. In the curve of standard normal variable, the perpendicular line at $Z = z_1$ is to be drawn by moving from left to right so that the area under the curve is 0.10. The figure is as follows



$$\therefore \; P(Z \le z_1) = P(-\infty < Z \le 0) - P(z_1 \le Z \le 0) = 0.10$$

$$\therefore \; 0.50 - P(z_1 \le Z \le 0) = 0.10$$

$$\therefore \; P(z_1 \le Z \le 0) = 0.5 - 0.10$$

$$\therefore \; P(0 \le Z \le z_1) = 0.40 \quad (\because \text{ symmetry})$$

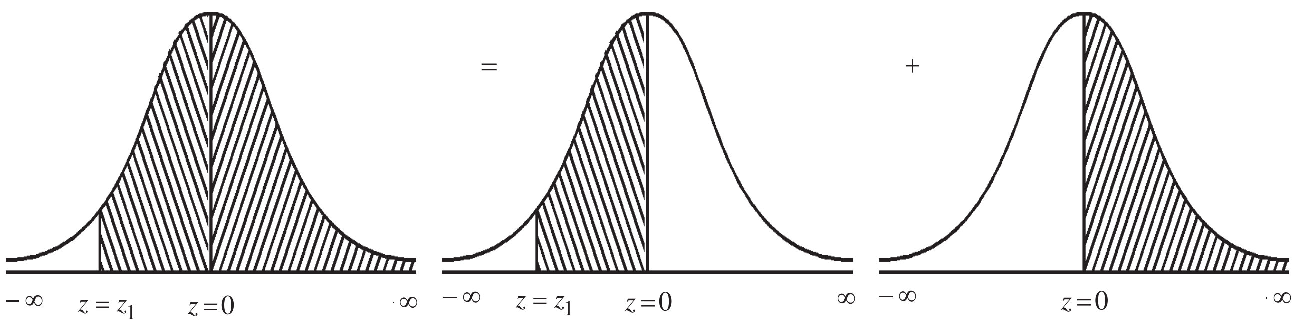
Now, corresponding to probability 0.4, it is not possible to obtain value of z_1 directly from the table of standard normal variable. Hence, the approximate value of z_1, is determined as under

From table	Area	Z-score
Nearest value before 0.40	0.3997	1.28
Nearest value after 0.40	0.4015	1.29
Average value	0.4006	1.285

From the above table, it can be seen that the nearest value to 0.40 is 0.3997 and the respective value of Z-score is 1.28. Also, z_1 is to the left of $Z = 0$, hence $z_1 = -1.28$.

Thus for $z_1 = -1.28$, $P(Z \le z_1) = 0.10$.

(2) Area to the right of $Z = z_1$ is 0.90 i.e $P(Z \ge z_1) = 0.90$. In the curve of standard normal variable, the perpendicular line at $Z = z_1$ is to be drawn by moving from right to left so that the area under the curve is 0.90. The figure is as under



$$P(Z \ge z_1) = P(z_1 \le Z \le 0) + P(0 \le Z < \infty) = 0.90$$

$$\therefore \; P(z_1 \le Z \le 0) + 0.50 = 0.90$$

$\therefore\ P(z_1 \le Z \le 0) = 0.40$

$\therefore\ P(0 \le Z \le z_1) = 0.40$ ($\because$ Symmetry)

As seen earlier $z_1 = -1.28$

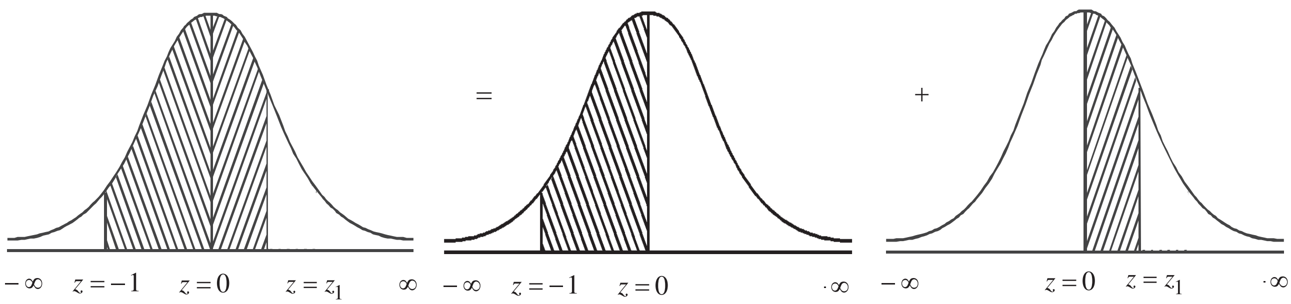
Thus, for $z_1 = -1.28$, $P(Z \ge z_1) = 0.90$

Illustration 9 : If Z is standard normal variable and z_1 is Z-score then obtain the values of z_1 satisfying the following conditions

(1) $P(-1 \le Z \le z_1) = 0.5255$

(2) $P(z_1 \le Z \le 2) = 0.7585$

(1) It is given that $P(-1 \le Z \le z_1) = 0.5255$. A perpendicular line at $Z = -1$ is drawn and then a perpendicular line at $Z = z_1$ is drawn to its right side so that the area between them is 0.5255. The figure is as under



$$P(-1 \le Z \le z_1) = P(-1 \le Z \le 0) + P(0 \le Z \le z_1) = 0.5255$$

$$\therefore\ P(0 \le Z \le 1) + P(0 \le Z \le z_1) = 0.5255 \quad (\because \text{ Symmetry})$$

$$\therefore\ 0.3413 + P(0 \le Z \le z_1) = 0.5255$$

$$\therefore\ P(0 \le Z \le z_1) = 0.5255 - 0.3413$$

$$\therefore\ P(0 \le Z \le z_1) = 0.1842$$

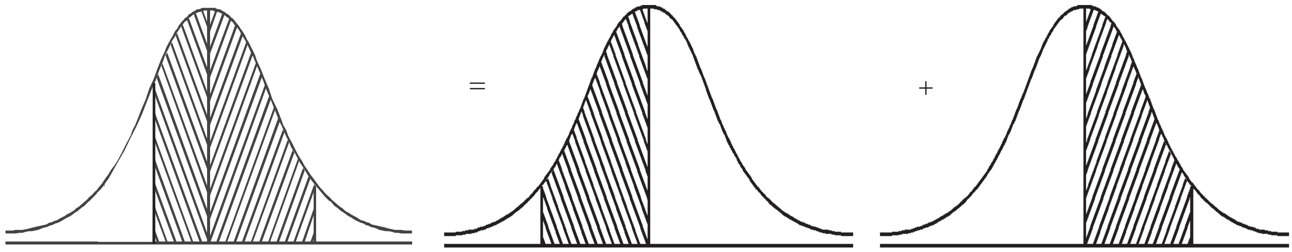
By using the table of standard normal variable the estimated value of Z-score, z_1 can be determined as under

From table	**Area**	**Z-score**
Nearest value before 0.1842	0.1808	0.47
Nearest value after 0.1842	0.1844	0.48
Average value	0.1826	**0.475**

From the above table, it can be seen that the nearest value to 0.1842 is 0.1844 and the respective value of Z-score is 0.48. Hence we take $z_1 = 0.48$.

Thus, for $z_1 = 0.48$, $P(-1 \le Z \le z_1) = 0.5255$

(2) It is given that $P(z_1 \le Z \le 2) = 0.7585$. A perpendicular line at $Z = 2$ is drawn and then a perpendicular line at $Z = z_1$ is drawn to its left so that the area between them is 0.7585. The figure is as under



$$\therefore P(z_1 \le Z \le 2) = P(z_1 \le Z \le 0) + P(0 \le Z \le 2) = 0.7585$$

$$\therefore P(z_1 \le Z \le 0) + 0.4772 = 0.7585$$

$$\therefore P(z_1 \le Z \le 0) = 0.7585 - 0.4772$$

$$\therefore P(0 \le Z \le z_1) = 0.2813 \quad (\because \text{ Symmetry})$$

By using the table of standard normal variable the estimated value of Z-score, z_1 can be determined as under

From table	**Area**	***Z*-score**
Nearest value before 0.2813	0.2794	0.77
Nearest value after 0.2813	0.2823	0.78
Average value	0.2809	0.775

From the above table, it can be seen that the nearest value to 0.2813 is 0.2809 and the respective value of Z-score is 0.775. Since Z-score is to the left of $Z = 0$, therefore $z_1 = -0.775$.

Thus, for $z_1 = -0.775$, $P(z_1 \le Z \le 2) = 0.7585$

Activity

For 30 persons residing around your residence, collect information of their weight (in kg) and obtain its mean (in kg) and standard deviation (in kg). Assuming that the weight of selected person follows normal distribution with the obtained mean and standard deviation, estimate (1) minimum weight of 5% persons having maximum weight (2) maximum weight of 15% of persons having minimum weight.

3.6 Illustrations

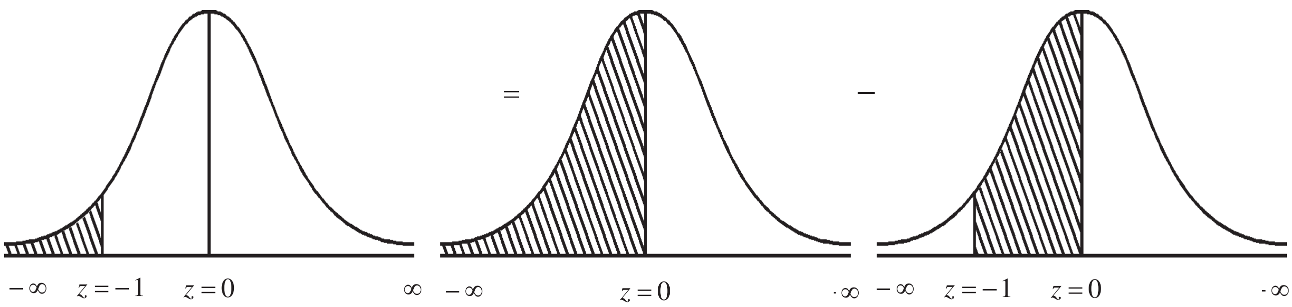
Illustration 10 : In a city, daily sale of petrol at a petrol pump follows normal distribution and its mean and standard deviation are 33,000 litre and 3000 litre respectively. (1) Obtain the percentage of days of a month during which the daily sales of petrol is less than 30,000 litre. (2) During the month of May, how many days are expected so that the sale of petrol is between 32,000 litre to 35,000 litre?

Here, X = daily sale of petrol at petrol pump (in litre). Also $\mu = 33{,}000$ litre and $\sigma = 3000$ litre.

(1) Probability that the sale of petrol is less than 30,000 litre

$$= P(X \le 30000) \quad = P\left(\frac{X-\mu}{\sigma} \le \frac{30000-33000}{3000}\right)$$

$$= P(Z \le -1)$$



$$= P(-\infty < Z \le 0) - P(-1 \le Z \le 0)$$

$$= 0.5 - P(0 \le Z \le 1) \; (\because \text{ Symmetry})$$

$$= 0.5 - 0.3413$$

$$= 0.1587$$

$\therefore$ Percentage of days during a certain month where the daily sale of petrol is less than 30,000 litre

$$= 0.1587 \times 100$$

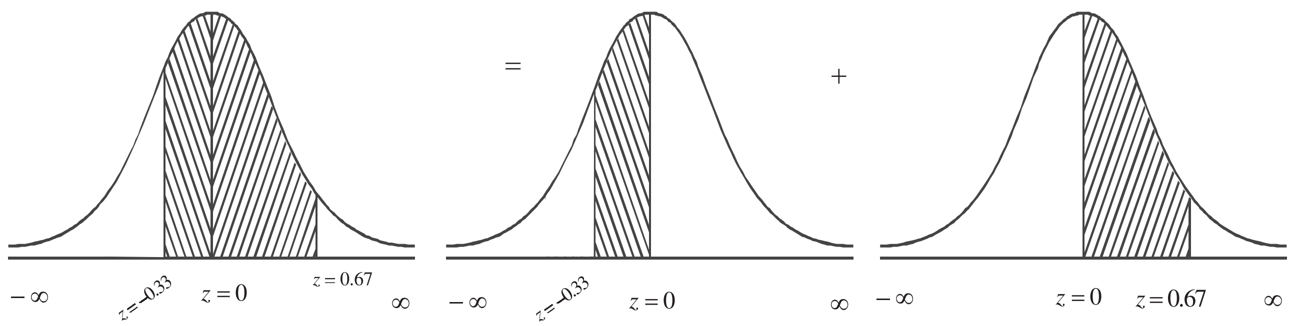
$$= 15.87 \%$$

Thus, during 15.87 % of the days of a month, the daily sale of petrol is less than 30,000 litres.

(2) The probability that during the month of May, the daily demand of petrol is between 32,000 and 35,000 litre

$$= P(32000 \le X \le 35000) \quad = P\left(\frac{32000-33000}{3000} \le \frac{X-\mu}{\sigma} \le \frac{35000-33000}{3000}\right)$$

$$= P(-0.33 \le Z \le 0.67)$$



$= P(-0.33 \le Z \le 0) + P(0 \le Z \le 0.67)$

$= P(0 \le Z \le 0.33) + 0.2486$ $\quad (\because$ Symmetry)

$= 0.1293 + 0.2486$

$= 0.3779$

The number of days in the month of May is N = 31. Therefore, the expected number of days in the month of May during which the daily sale of petrol is between 32,000 litter to 35,000 litre $= 31 \times 0.3779$

$= 11.71$

$\simeq$ 12 Days (approximately)

Thus, in the month of May, approximately during 12 days the demand of petrol is between 32,000 litre and 35,000 litre.

Illustration 11 : 200 students are selected from all the students of a school and the marks obtained by them in an examination of 100 marks follows normal distribution. The mean marks of the distribution is 60 and its standard deviation is 8.

(1) If 70 or more marks are required for the special scholarship then obtain the number of students getting special scholarship.

(2) Obtain the minimum marks of 10% of the students getting maximum marks.

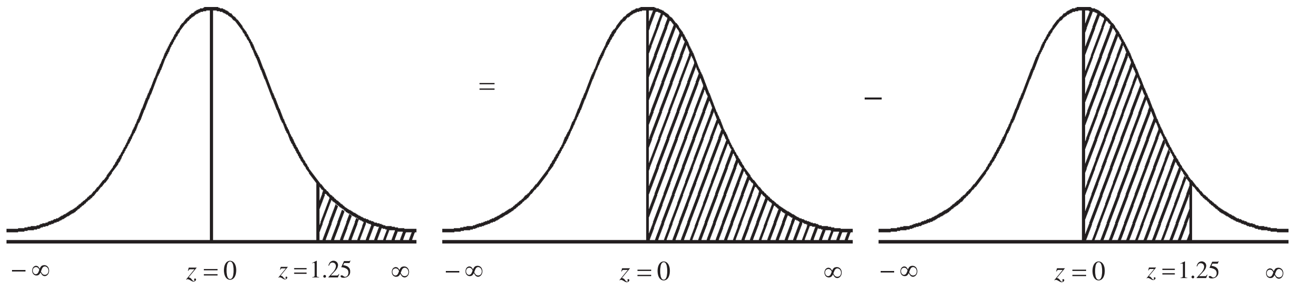
Here, X = marks obtained by a student

Also, N = 200, $\mu = 60$ and $\sigma = 8$.

(1) Probability that the marks of the student is 70 or more

$= P(X \ge 70) \quad = P\left(\frac{X-\mu}{\sigma} \ge \frac{70-60}{8}\right)$

$= P(Z \ge 1.25)$



$= P(0 \le Z < \infty) - P(0 \le Z \le 1.25)$

$= 0.5 - 0.3944$

$= 0.1056$

$\therefore$ the expected number of students getting 70 or more marks

$= 200 \times 0.1056$

$= 21.12$

$\simeq$ 21 (approximately)

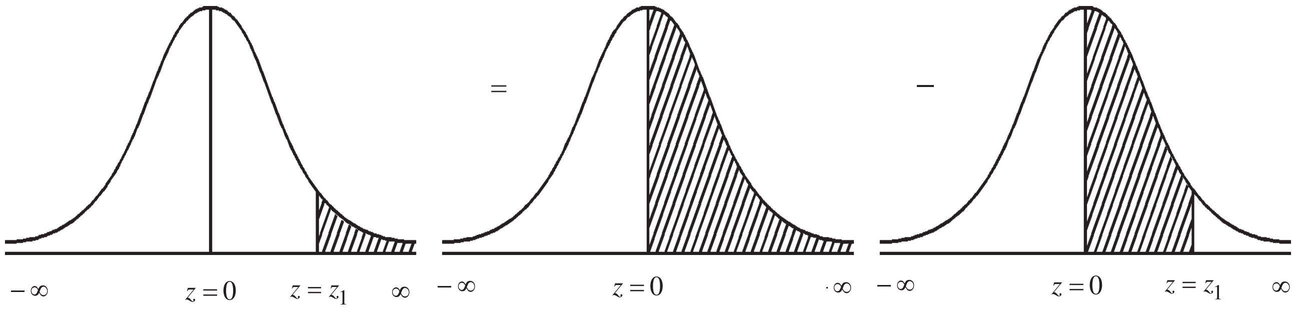
Thus, the approximate number of students getting special scholarship is 21.

(2) Suppose the minimum marks of 10% of the students getting maximum marks is x_1. The probability that a student gets x_1 or more marks is 0.10.

$\therefore \quad P(X \ge x_1) = 0.10$

$$\therefore \quad P\left(\frac{X-\mu}{\sigma} \geq \frac{x_1-60}{8}\right) = 0.10$$

$$\therefore \quad P(Z \geq z_1) = 0.10, \quad \text{where} \quad z_1 = \frac{x_1-60}{8}$$



$$\therefore 0.10 = P(0 \leq Z < \infty) - P(0 \leq Z \leq z_1)$$

$$\therefore 0.10 = 0.5 - P(0 \leq Z \leq z_1)$$

$$\therefore P(0 \leq Z \leq z_1) = 0.40$$

By using the table of standard normal variable the estimated value of z_1 can be determined as under

From table	Area	Z-score
Nearest value before 0.40	0.3997	1.28
Nearest value after 0.40	0.4015	1.29
Average value	0.4006	1.285

From the above table it can be seen that the nearest value to 0.40 is 0.3997 and the respective value of Z-score is 1.28.

Therefore $z_1 = 1.28$

$$\therefore \frac{x_1-60}{8} = 1.28$$

$$\therefore x_1 - 60 = 10.24$$

$$\therefore x_1 = 70.24$$

Thus, the minimum marks of most intelligent 10% of the students is $70.24 \simeq 70$.

Illustration 12 : The monthly income of a group of employees follows normal distribution. The mean of the distribution is ₹ 15,000 and its standard deviation is ₹ 4000. From this information, (1) obtain range of monthly income for middle 60% of the employees. (2) if monthly income of 250 employees is between ₹ 15000 and certain fixed income ₹ x_1 then find the value of x_1.

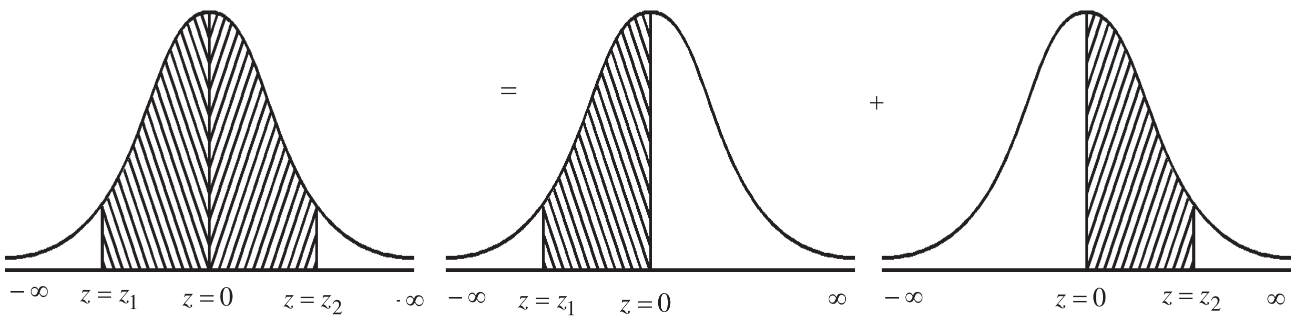
Here, X = monthly income of an employee. Also, N = 1000, $\mu = ₹\, 15,000$ and $\sigma = ₹\, 4000$.

(1) Suppose the range of monthly income of exactly middle 60% of employee is ₹ x_1 and ₹ x_2 where x_1 and x_2 are at equal distance from mean μ. Now the probability that the monthly income of employee is between x_1 and x_2 is 0.60.

i.e. $P(x_1 \le X \le x_2) = 0.60.$

$$\therefore\ P\left(\frac{x_1-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right) = 0.60$$

$$\therefore\ P(z_1 \le Z \le Z_2) = 0.60 \text{ where } z_1 = \frac{x_1 - 15000}{4000} \text{ and } z_2 = \frac{x_2 - 15000}{4000}$$



$$0.60 = P(z_1 \le Z \le 0) + P(0 \le Z \le z_2)$$

Now, since x_1 and x_2 are at equal distance from mean μ, the perpendicular line at $Z = 0$ divides the total area (probability) under the curve between $Z = z_1$ and $Z = z_2$ into two equal parts. So, $z_1 = -z_2$ and also $P(z_1 \le Z \le 0) = 0.30$ and $P(0 \le Z \le z_2) = 0.30$.

By using the table of standard normal variable the estimated value of z_1 and z_2 can be obtained as under

From table	Area	Z-score
Nearest value before 0.30	0.2995	0.84
Nearest value after 0.30	0.3023	0.85
Average value	0.3009	0.845

The nearest value to 0.30 is 0.2995 and the respective value of Z-score is 0.84.

$\therefore\ z_1 = -0.84$ and $z_2 = 0.84$

$$\therefore\ \frac{x_1 - 15000}{4000} = -0.84 \text{ and } \frac{x_2 - 15000}{4000} = 0.84$$

$\therefore\ x_1 - 15000 = -3360$ and $\therefore\ x_2 - 15000 = 3360$

$\therefore\ x_1 = 11640$ and $\therefore\ x_2 = 18360$

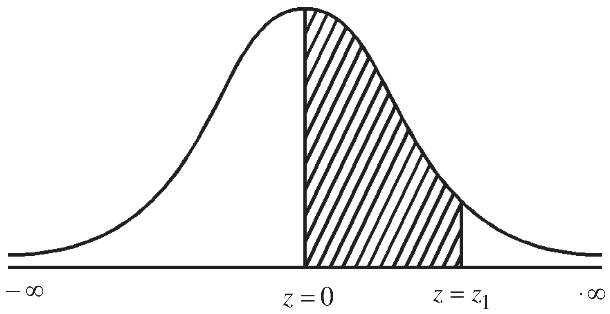
Thus, the range of monthly income for middle 60 % of the employees will be ₹ 11,640 to ₹ 18,360.

(2) Monthly income of 250 employees is between ₹ 15,000 and ₹ x_1

Therefore $P(15000 \le X \le x_1) = \frac{250}{1000}$

$\therefore\ P\left(\frac{15000-15000}{4000} \le \frac{X-\mu}{\sigma} \le \frac{x_1-15000}{4000}\right) = 0.25$

$\therefore\ P(0 \le Z \le z_1) = 0.25$ where $z_1 = \frac{x_1-15000}{4000}$



By using the table of standard normal variable the estimated value of z_1 can be obtained as under

From table	**Area**	***Z*-score**
Nearest value before 0.25	0.2486	0.67
Nearest value after 0.25	0.2518	0.68
Average value	0.2502	0.675

It is clear from the above table that $z_1 = 0.675$

$\therefore\ \frac{x_1-15000}{4000} = 0.675$

$\therefore\ x_1 - 15000 = 2700$

$\therefore\ x_1 = 17700$

Thus, monthly income of 250 employees will be between ₹ 15,000 and ₹ 17,700.

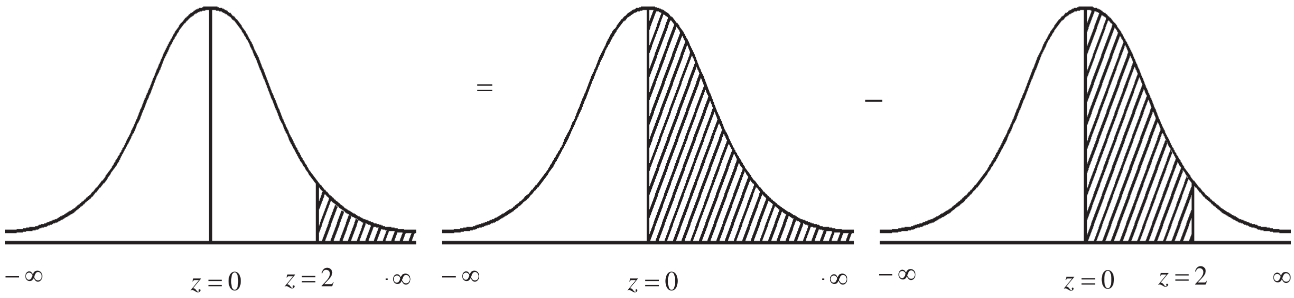
Illustration 13 : The bill amount of purchase by the customers in departmental store follows normal distribution and its mean is ₹ 800 and standard deviation is ₹ 200. On a day, 57 customers had the bill amount more than ₹ 1200. Estimate the number of customers who visited the store on that day.

Here X = bill amount of purchase by the customer. $\mu = 800$ and $\sigma = 200$. Suppose N customers visited that departmental store during that day.

The probability that the bill amount of purchase by the customer is more than ₹ 1200

$$= P(X \geq 1200) \quad = P\left(\frac{X-\mu}{\sigma} \geq \frac{1200-800}{200}\right)$$

$$= P(Z \geq 2)$$



$$= P(0 \leq Z < \infty) - P(0 \leq Z \leq 2)$$

$= 0.5 - 0.4772$ (From Z-table)

$= 0.0228$

Now, the expected number of customer whose bill amount of purchase is more than ₹ 1200 $= N \times P(x \geq 1200)$

$57 = N \times 0.0228$

$\therefore \quad N = \dfrac{57}{0.0228}$

$\therefore \quad N = 2500$

$\therefore$ 2500 customers visited the store on that day.

Illustration 14 : For a group of 1000 persons, the average height is 165 cms and variance is 100 (cms)^2. The distribution of height of these persons follows normal distribution. From this information, determine the third decile and the 60th percentile and interpret it.

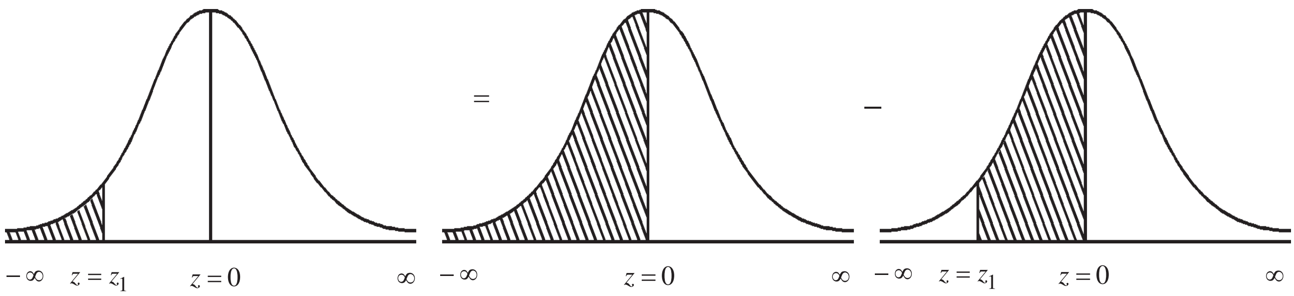
Here, X = height of a person in the group. Also, $\mu = 165$ and $\sigma^2 = 100$ therefore $\sigma = 10$

The third decile (D_3) is to be determined. According to the definition of D_3, 30% of the observations in the data have the value less than or equal to D_3.

$$\therefore P(X \leq D_3) = \frac{30}{100}$$

$$\therefore P\left(\frac{X-\mu}{\sigma} \leq \frac{D_3-165}{10}\right) = 0.30$$

$$\therefore P(Z \leq z_1) = 0.30 \text{ where } z_1 = \frac{D_3-165}{10}$$



$$0.30 = P(-\infty < Z \le 0) - P(z_1 \le Z \le 0)$$

$$0.30 = 0.50 - P(z_1 \le Z \le 0)$$

$$\therefore P(z_1 \le Z \le 0) = 0.50 - 0.30$$

$$\therefore P(0 \le Z \le z_1) = 0.20 \quad (\because \text{ Symmetry})$$

By using the table of standard normal variable the estimated value of z_1 can be obtained as under

From table	**Area**	***Z*-score**
Nearest value before 0.2	0.1985	0.52
Nearest value after 0.2	0.2019	0.53
Average value	0.2002	0.525

The nearest value to 0.2 is 0.2002 and the respective value of *Z*-score is 0.525 and it is to the left of $Z = 0$

Therefore $z_1 = -0.525$

$$\therefore \quad \frac{D_3 - 165}{10} = -0.525$$

$$\therefore \quad D_3 - 165 = -5.25$$

$$\therefore \quad D_3 = 159.75$$

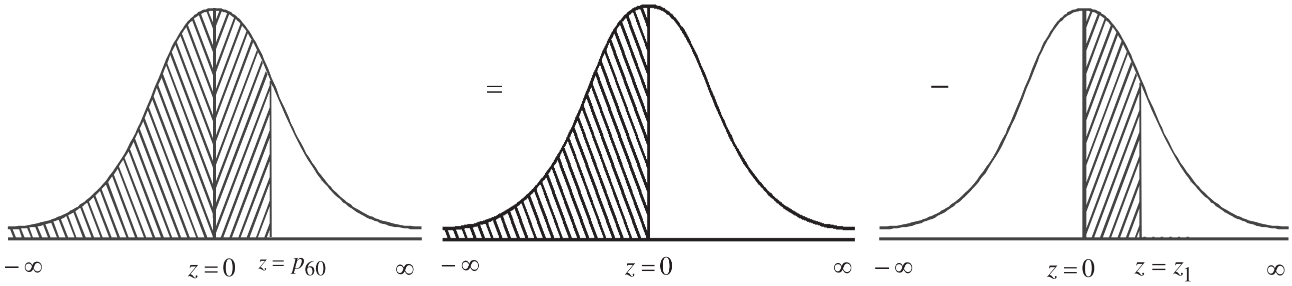
Thus, 30% of the persons in the group have height less than or equal to 159.75 cms.

Now, according to the definition of the 60th percentile (P_{60}), 60% of the observations in the given data have the value less than or equal to P_{60}.

$$\therefore P(X \le P_{60}) = \frac{60}{100}$$

$$\therefore P\left(\frac{X-\mu}{\sigma} \le \frac{P_{60}-165}{10}\right) = 0.60$$

$$\therefore P(Z \le z_1) = 0.60 \quad \text{where} \quad z_1 = \frac{P_{60}-165}{10}$$



$$0.60 = P(-\infty < Z \le 0) + P(0 \le Z \le z_1)$$

$$0.60 = 0.50 + P(0 \le Z \le z_1)$$

$$\therefore P(0 \le Z \le z_1) = 0.10$$

By using the table of standard normal variable the estimated value of z_1 can be obtained as under

From table	Area	*Z*-score
Nearest value before 0.10	0.0987	0.25
Nearest value after 0.10	0.1026	0.26
Average value	0.10065	0.255

The nearest value to 0.10 is 0.10065 and the respective value of *Z*-score is 0.255.

$$\therefore z_1 = 0.255$$

$$\therefore \frac{P_{60} - 165}{10} = 0.255$$

$$\therefore P_{60} - 165 = 2.55$$

$$\therefore P_{60} = 167.55$$

Thus, 60% of the persons in the group have height less than or equal to 167.55 cms.

Illustration 15 : A manufacturing company produces electric bulb and life of the electric bulb (in hours) follows normal distribution. Its average life is 2040 hours. If 3.36% of bulbs have life more than 2150 hours then find variance of the life of bulbs.

Here, X = life of electric bulb. Also, $\mu = 2040$. Suppose its variance is σ^2.

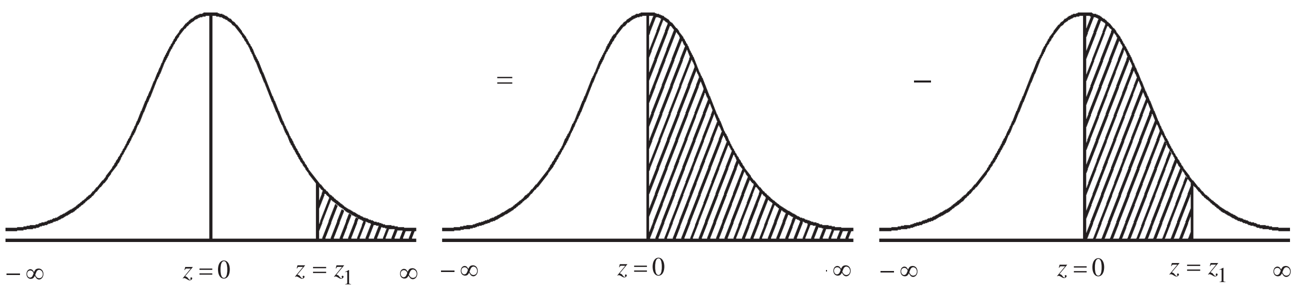
Now, 3.36% of the bulbs have life more than 2150 hours.

$$\therefore P(X \ge 2150) = \frac{3.36}{100}$$

$$\therefore P\left(\frac{X-\mu}{\sigma} \ge \frac{2150-2040}{\sigma}\right) = 0.0336$$

$$\therefore P\left(Z \ge \frac{110}{\sigma}\right) = 0.0336$$

$$\therefore P(Z \ge z_1) = 0.0336 \quad \text{where} \quad z_1 = \frac{110}{\sigma}$$



$$0.0336 = P(0 \le Z < \infty) - P(0 \le Z \le z_1)$$

$$0.0336 = 0.5 - P(0 \le Z \le z_1)$$

$$\therefore P(0 \le Z \le z_1) = 0.5 - 0.0336$$

$$\therefore P(0 \le Z \le z_1) = 0.4664$$

From the table of standard normal variable, for Z-score 1.83, $P(0 \le Z \le 1.83) = 0.4664$

$$\therefore z_1 = 1.83$$

$$\therefore \frac{110}{\sigma} = 1.83$$

$$\therefore \sigma = \frac{110}{1.83}$$

$$\therefore \sigma = 60.11$$

$$\therefore \sigma^2 = 3613.21$$

Thus, the variance of the life of electric bulbs produced is 3613.21 $(\text{hours})^2$.

Illustration 16 : The profit in daily business of a businessman having grocery shop follows normal distribution. Variance of profit is 22500 (₹)^2, and the probability that the daily profit is less than Rs. 1000 is 0.0918. Find the average daily profit.

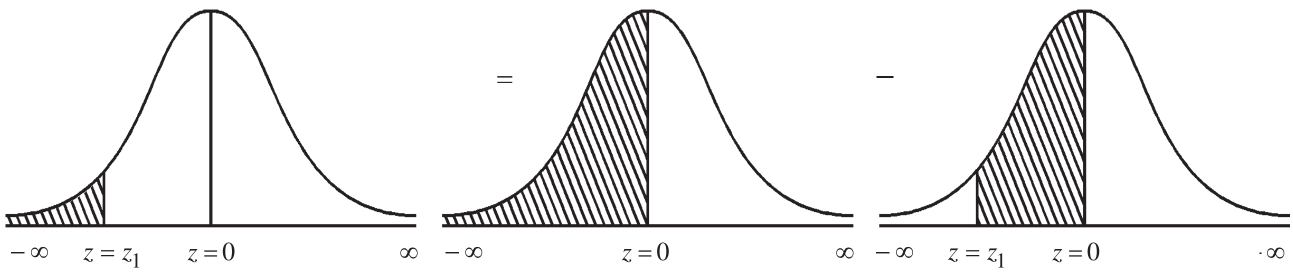
Here, X = daily profit of the businessman in his business. As $\sigma^2 = 22{,}500$, therefore $\sigma = 150$ and suppose the average profit is μ.

Now, the probability that the daily profit is less than ₹ 1000 = 0.0918

$$\therefore P(X \le 1000) = 0.0918$$

$$\therefore P\left(\frac{X-\mu}{\sigma} \le \frac{1000-\mu}{150}\right) = 0.0918$$

$$\therefore P(Z \le z_1) = 0.0918 \text{ where } z_1 = \frac{1000-\mu}{150}$$



$$0.0918 = P(-\infty < Z \le 0) - P(z_1 \le Z \le 0)$$

$$0.0918 = 0.5 \ - P(z_1 < Z \le 0)$$

$$\therefore \ P(z_1 \le Z \le 0) = 0.5 - 0.0918$$

$$\therefore \ P(0 \le Z \le z_1) = 0.4082 \quad (\because \ \text{Symmetry})$$

From the table of standard normal variable, Z-score is 1.33.

$$\therefore \ z_1 = -1.33$$

$$\therefore \ \frac{1000-\mu}{150} = -1.33$$

$$\therefore \ \mu = 800.5$$

Thus, the average daily profit of the businessman in his business is ₹ 800.50.

Illustration 17 : The maximum temperature of a city during summer follows normal distribution. On a particular day, the probability that the maximum temperature of the city is more than 31° Celsius is 0.3085, whereas the probability that during some other day, the maximum temperature is less than 27° is 0.0668. Find mean and standard deviation of the maximum temperature of the city.

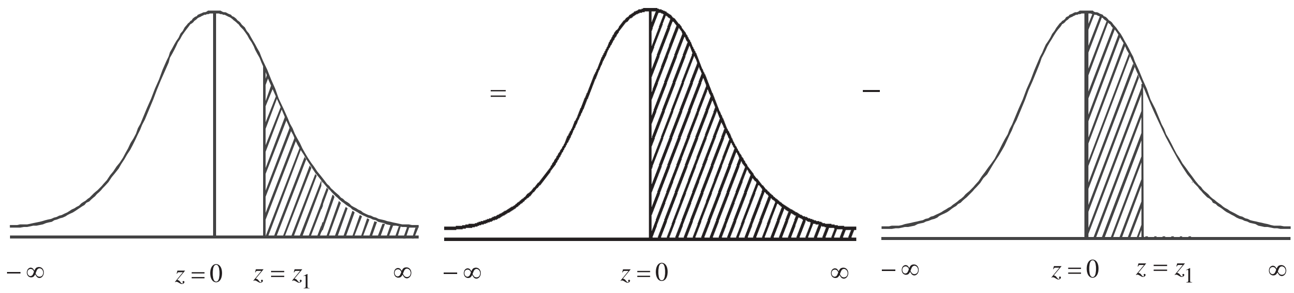
Here, X = maximum temperature (in Celsius) of the city. Suppose μ and σ are the mean and standard deviation.

Now, the probability that the maximum temperature is more than 31° Celsius = 0.3085

$$\therefore \ P(X \ge 31) = 0.3085$$

$$\therefore \ P\left(\frac{X-\mu}{\sigma} \ge \frac{31-\mu}{\sigma}\right) = 0.3085$$

$$\therefore \ P(Z \ge z_1) = 0.3085 \quad \text{where} \quad z_1 = \frac{31-\mu}{\sigma}$$



$$0.3085 = P(0 \le Z < \infty) - P(0 \le Z \le z_1)$$

$$0.3085 = 0.5 \ - P(0 \le Z \le z_1)$$

$$\therefore \ P(0 \le Z \le z_1) = 0.5 - 0.3085$$

$$\therefore \ P(0 \le Z \le z_1) = 0.1915$$

From the table of standard normal variable, Z-score is 0.5.

$$\therefore \ z_1 = 0.5$$

$$\therefore \ \frac{31-\mu}{\sigma} = 0.5$$

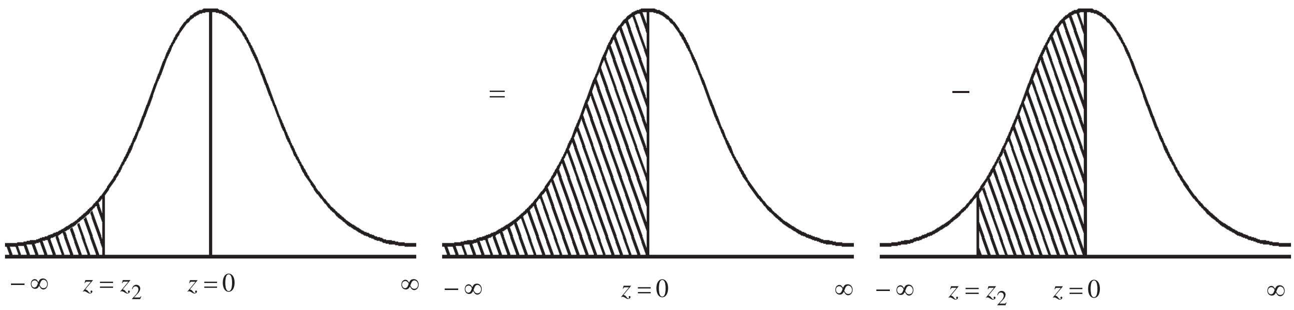
$$\therefore \ 31 - \mu = 0.5\,\sigma \ \ldots\ldots\ldots\ldots\ldots\ldots \ (1)$$

The probability that the maximum temperature is less than 27° Celsius = 0.0668

$$\therefore \ P(X \le 27) = 0.0668$$

$$\therefore \ P\left(\frac{X-\mu}{\sigma} \le \frac{27-\mu}{\sigma}\right) = 0.0668$$

$$\therefore \ P(Z \le z_2) = 0.0668 \text{ where } z_2 = \frac{27-\mu}{\sigma}$$



$$0.0668 = P(-\infty < Z \le 0) - P(z_2 \le Z \le 0)$$

$$\therefore \ 0.0668 = 0.5 - P(z_2 \le Z \le 0)$$

$$\therefore \ P(z_2 \le Z \le 0) = 0.5 - 0.0668$$

$$\therefore \ P(0 \le Z \le z_2) = 0.4332 \ (\because \text{ Symmetry})$$

From the table of standard normal variable, Z-score is 1.5

$$\therefore \ z_2 = -1.5$$

$$\therefore \ \frac{27-\mu}{\sigma} = -1.5$$

$$\therefore \ 27 - \mu = -1.5\sigma \ \text{.................} \ (2)$$

Solving equations (1) and (2),

$$31 - \mu = 0.5\,\sigma$$

$$27 - \mu = -1.5\,\sigma$$

$$- \quad + \quad +$$

$$4 = 2\sigma$$

$$\therefore \ \sigma = 2$$

By substituting $\sigma = 2$ in equation (1),

$$31 - \mu = 0.5\,(2)$$

$$\therefore \ 31 - \mu = 1$$

$$\therefore \ \mu = 30$$

Thus, the mean of maximum temperature of a city is 30° Celsius and its standard deviation is 2° Celsius.

Illustration 18 : The probability density function of a normal variable is as under

$$f(x) = \frac{1}{4\sqrt{2\pi}} \; e^{-\frac{1}{32}(x-50)^2} \; ; \quad -\infty < x < \infty$$

Obtain parameters of this distribution and find the values of following :

(1) $P(52 \le X \le 58)$ (2) $P(|X-45| \le 4)$

By comparing the given probability density function with the probability density function of normal variable X

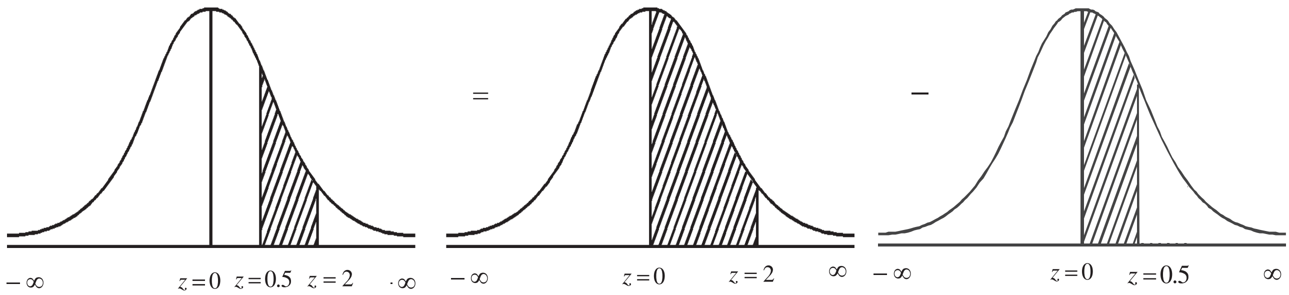
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \; ; \quad -\infty < x < \infty$$

Here, $\sigma\sqrt{2\pi} = 4\sqrt{2\pi}$ and $\mu = 50$

$\therefore \quad \sigma = 4$

(1) $P(52 \le X \le 58) = P\left(\frac{52-50}{4} \le \frac{X-\mu}{\sigma} \le \frac{58-50}{4}\right)$

$= P(0.5 \le Z \le 2)$



$= P(0 \le Z \le 2) - P(0 \le Z \le 0.5)$

$= 0.4772 - 0.1915$

$= 0.2857$

Thus, $P(52 \le X \le 58) = 0.2857$

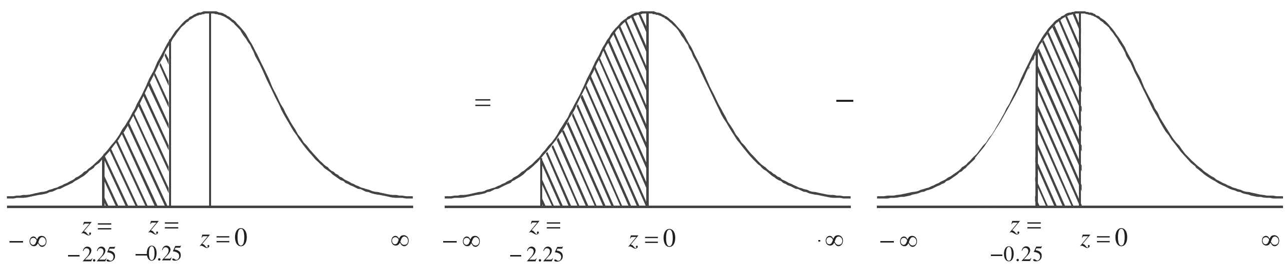
(2) $P(|X-45| \le 4) = P(-4 \le (X-45) \le 4)$ (definition of modulus)

$= P(-4+45 \le (X-45)+45 \le 4+45)$

$= P(41 \le X \le 49)$

$= P\left(\frac{41-50}{4} \le \frac{X-\mu}{\sigma} \le \frac{49-50}{4}\right)$

$= P(-2.25 \le Z \le -0.25)$



$= P(-2.25 \le Z \le 0) - P(-0.25 \le Z \le 0)$

$= P(0 \le Z \le 2.25) - P(0 \le Z \le 0.25)$ ($\because$ Symmetry)

$= 0.4878 - 0.0987$

$= 0.3891$

Thus, $P(|X - 45| \le 4) = 0.3891$.

Illustration 19 : The probability density function of a normal variable X is defined as under

$$f(x) = \text{constant} \cdot e^{-\frac{1}{2}\left(\frac{x-25}{10}\right)^2} \quad ; \quad -\infty < x < \infty$$

From this normal distribution estimate the values of the following :

(1) Third quartile (2) Quartile deviation (3) Mean deviation

By comparing the given probability density function with the probability density function of normal variable X,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad -\infty < x < \infty$$

Here, $\mu = 25$ and $\sigma = 10$.

(1) Third quartile $Q_3 = \mu + 0.675\,\sigma$

$= 25 + 0.675\ (10)$

$= 25 + 6.75$

$= 31.75$

(2) Quartile deviation $= \frac{2}{3}\ \sigma$

$= \frac{2}{3}\ (10)$

$= \frac{20}{3}$

(3) Mean deviation $= \frac{4}{5}\ \sigma$

$= \frac{4}{5}\ (10)$

$= 8$

Thus, for the given normal distribution the estimates of the required values are 31.75, $\frac{20}{3}$ and 8 respectively.

Illustration 20 : The extreme quartiles for a normal distribution are 20 and 50 respectively. Obtain the limits which include 95% of the observations of the distribution.

Here, $Q_1 = 20$ and $Q_3 = 50$. For the normal distribution

Mean = Median = Mode = $\frac{Q_3+Q_1}{2}$

$$\therefore \mu = \frac{50+20}{2}$$

$$\therefore \mu = 35$$

Quartile deviation = $\frac{2}{3}\ \sigma$

$\therefore \frac{Q_3-Q_1}{2} = \frac{2}{3}\ \sigma$

$\therefore \frac{50-20}{2} = \frac{2}{3}\ \sigma$

$\therefore \frac{50-20}{2} \times \frac{3}{2} = \sigma$

$\therefore \sigma = 22.5$

For the normal distribution the limits including 95% of the observations are $\mu \pm 1.96\sigma$. Hence, the interval (limits) is

$$(\mu - 1.96\sigma,\ \mu + 1.96\sigma)$$

$$\therefore (35 - 1.96\,(22.5),\ 35 + 1.96\,(22.5))$$

$$\therefore (35 - 44.1,\ 35 + 44.1)$$

$$\therefore (-9.1,\ 79.1)$$

Thus, from the given information, the limits including 95% of the observations are –9.1 to 79.1.

Illustration 21 : For a normal distribution, the first quartile and the mean deviation are 20 and 24 respectively. Obtain an estimate of the value of mode.

Here, Q_1 = 20 and mean deviation = 24.

$$\therefore \frac{4}{5}\ \sigma = 24$$

$$\therefore \sigma = 24 \times \frac{5}{4}$$

$$\therefore \sigma = 30$$

Now, quartile deviation = $\frac{2}{3}\ \sigma$

$\therefore \frac{Q_3-Q_1}{2} = \frac{2}{3}\ \sigma$

$\therefore \frac{Q_3-20}{2} = \frac{2}{3}\ (24)$

$\therefore Q_3 - 20 = 16 \times 2$

$\therefore Q_3 = 32 + 20$

$\therefore Q_3 = 52$

Now, for normal distribution, Mean = Median = Mode = $\frac{Q_3+Q_1}{2}$

$$= \frac{52+20}{2}$$

$$= 36$$

Thus, from the given information the estimated value of mode is 36.

Illustration 22 : The number of vehicles arriving at toll station during busy hours of national highway follows normal distribution. The mean of this distribution is μ and its standard deviation is σ. The number of vehicles arriving at two different busy time periods are 88 and 64 and if the respective values of *Z*-score for these values are 0.8 and –0.4 then find mean and standard deviation of number of vehicles arriving at the toll station during busy period.

Here, X = number of vehicles arriving at toll station during busy period.

The mean of this distribution is μ and its standard deviation is σ

$$Z\text{-score} = \frac{X-\mu}{\sigma}$$

When $X = 88$ then $Z = 0.8$ therefore $0.8 = \frac{88-\mu}{\sigma}$

$$\therefore\ 0.8\,\sigma = 88 - \mu \qquad (1)$$

When $X = 64$ then $Z = -0.4$ therefore $-0.4 = \frac{64-\mu}{\sigma}$

$$\therefore -0.4\,\sigma = 64 - \mu \qquad (2)$$

Solving equations (1) and (2),

$$0.8\,\sigma = 88 - \mu$$
$$-\quad 0.4\,\sigma = 64 - \mu$$
$$+ \qquad - \quad +$$

$$1.2\sigma = 24$$

$$\therefore\ \sigma = 20$$

By putting $\sigma = 20$ in equation (1), $0.8(20) = 88 - \mu$

$$\therefore\ 16 = 88 - \mu$$

$$\therefore\ \mu = 72$$

Thus. the mean of given data is $\mu = 72$ vehicles and its standard deviation is $\sigma = 20$ vehicles.

Activity

Collect the information of average monthly expenses of 30 families residing around your residence. Assuming that the average monthly expense of these families follows normal distribution with the mean and standard deviation determined by you,

(1) Obtain the limits of average monthly income of middle 60% of the families.

(2) Find the percentage of observations lying between the range $\mu \pm \sigma$ from your data.

Summary

- A function for obtaining probability that a continuous random variable assumes value between specified interval is called probability density function of that variable.
- The probability for a definite value of continuous random variable X obtained by the probability density function is always zero (0).
- A curve drawn by considering different values of normal variable X and its respective values of probability density function $f(x)$ is called normal curve.
- Normal curve is completely bell shaped and its skewness is zero.
- If X is a random normal variable with mean μ and standard deviation σ then $Z = \frac{X-\mu}{\sigma}$ is called standard normal variable.
- Standard normal probability distribution is a probability distribution of normal variable with mean zero and standard deviation 1.
- The observed value of standard normal variable Z is called standard score or Z-score and it is independent of unit of measurement.
- Normal distribution is also defined as $N(\mu, \sigma^2)$ where μ and σ are parameters of the distribution which indicate its mean and variance respectively.
- In order to express probability in percentage, the probability is multiplied by 100.
- In order to obtain expected number of observations, the probability is multiplied by the total number of observations (N).

List of Formulae

If X is a normal variable with mean μ and standard deviation σ then

(1) Standard normal variable $Z = \frac{X-\mu}{\sigma}$

(2) Mean = Median = Mode = $\frac{Q_3+Q_1}{2}$

(3) Approximate value of the first quartile $Q_1 = \mu - 0.675\,\sigma$

(4) Approximate value of the third quartile $Q_3 = \mu + 0.675\,\sigma$

(5) Quartile deviation = $\frac{2}{3}\sigma$ (approximately)

(6) Mean deviation = $\frac{4}{5}\sigma$ (approximately)

EXERCISE 3

Section A

Find the correct option for the following multiple choice questions :

1. Which of the following is probability density function for normal variable X with mean μ and standard deviation σ ?

(a) $f(x)=\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)}$; $-\infty<x<\infty$ (b) $f(x)=\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$; $-\infty<x<\infty$

(c) $f(x)=\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$; $-\infty<x<\infty$ (d) $f(x)=\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$; $0\le x<\infty$

2. For a normal variable X with mean μ and standard deviation σ, which of the following is standard normal variable Z for it ?

(a) $Z = \frac{x-\sigma}{\mu}$ (b) $Z = \frac{\sigma-x}{\mu}$ (c) $Z = \frac{\mu-x}{\sigma}$ (d) $Z = \frac{x-\mu}{\sigma}$

3. Which of the following is probability density function for standard normal variable?

(a) $f(z)=\ e^{-\frac{1}{2}z^2}$; $-\infty<z<\infty$ (b) $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$ $-\infty<z<\infty$

(c) $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$; $0<z<\infty$ (d) $f(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2}$ $-\infty<z<\infty$

4. Which of the following are mean and variance of standard normal variable?

(a) Mean = 0, Variance = 1 (b) Mean = 1, Variance = 0

(c) Mean = 0, Variance = 0 (d) Mean = 1, Variance = 1

5. What is the total area under normal curve among the following ?

(a) –1 (b) 0 (c) 1 (d) 0.5

6. What is the area under the normal curve to the right hand side of perpendicular line at $X = \mu$?

(a) 0 (b) 0.5 (c) 1 (d) –0.5

7. In normal distribution, usually which limits include 99 % of the observations?

(a) $\mu\pm1.96\,\sigma$ (b) $\mu\pm2\,\sigma$ (c) $\mu\pm3\,\sigma$ (d) $\mu\pm2.575\sigma$

8. In normal distribution, usually what percentage of the observations are included in the limits $\mu\pm\sigma$?

(a) 34.13 % (b) 95.45 % (c) 68.26 % (d) 50 %

9. Which of the following is approximate value of mean deviation for normal variable?

(a) $\frac{4}{5}\sigma$ (b) $\frac{4}{5}\mu$ (c) $\frac{2}{3}\sigma$ (d) $\frac{2}{3}\mu$

10. Which of the following is approximate value of quartile deviation for standard normal variable?

(a) $\frac{2}{3}\sigma$ (b) $\frac{2}{3}$ (c) $\frac{4}{5}\sigma$ (d) $\frac{4}{5}$

11. Mean and the first quartile for a normal distribution are 11 and 3 respectively. Which of the following is the value of the third quartile?

(a) 8 (b) 14 (c) 19 (d) 10

12. For a normal distribution, approximate value of mean deviation is 20. Which of the following is the value of quartile deviation?

(a) $\frac{25}{3}$ (b) $\frac{32}{3}$ (c) 24 (d) $\frac{50}{3}$

13. In usual notation of normal distribution, $x = 25$, $\mu = 20$ and $\sigma = 5$ then which of the following is the value of standard normal variable?

(a) 1 (b) –1 (c) 4 (d) $\frac{10}{3}$

14. Mean of a normal variable X is 50. If the value of Z–score is –2.5 for x = 25 then which of the following is a variance of the distribution ?

(a) 10 (b) 100 (c) 50 (d) 25

15. If the distribution of normal variable is shown as $N(20, 4)$ then which of the following intervals includes 99.73% of observations?

(a) (18, 22) (b) (16, 24) (c) (14, 26) (d) (12, 28)

Section B

Answer the following questions in one sentence :

1. Give the values of the constants used in probability density function of normal variable.

2. What is the probability that a continuous random variable takes definite value?

3. What is the shape of normal curve?

4. What is the skewness of normal distribution?

5. "Standard score is independent of unit of measurement". Is this statement true or false?

6. For which value of standard normal variable, the standard normal curve is symmetric on both the sides?

7. Which value of normal variable divides the area of normal curve in two equal parts?

8. What percentage of area is covered under the normal curve within the range $\mu - 2\sigma$ to $\mu + 2\sigma$?

9. Mean of a normal distribution is 13.25 and its standard deviation is 10. Estimate the value of its third quartile.

10. For a normal distribution having mean 10 and standard deviation 6, estimate the value of quartile deviation.

11. The approximate value of mean deviation for a normal distribution is 8. Find its standard deviation.

12. For a normal distribution, the estimated value of quartile deviation is 12. Find the value of its standard deviation.

13. For a probability distribution of standard normal variable, state the estimated limits for the middle 50 % observations.

14. The extreme quartiles of normal distribution are 20 and 30. Find its mean.

15. The monthly expense of a group of persons follows normal distribution with mean ₹ 10,000 and standard deviation ₹ 1000. A student has obtained a *Z*- score = ₹ 1 for randomly selected person having monthly expense more than 11,000. Is this calculation of *Z*- score true ? Give reason.

16. The age of a group of persons follows normal distribution with mean 45 years and standard deviation 10 years. Calculate *Z*-score for a randomly selected person having age 60 years.

17. Marks obtained by students of a school in Economics subject follows normal distribution with mean μ and standard deviation σ. The value of standard score that a randomly selected student obtained 60 marks is 1. If the variance of variable is 100 (marks)^2 then find average marks.

Section C

Answer the following questions :

1. Define probability density function of continuous random variable.

2. Write the conditions for probability density function for continuous variable.

3. How is the normal curve drawn?

4. Define probability density function for normal variable.

5. What is the shape of standard normal curve? To which value of variable it is symmetric?

6. Define standard normal variable and write its probability density function.

7. A normal variable X has the probability density function as,

$$f(x) = \text{constant} \times e^{-\frac{1}{50}(x-10)^2} \quad ; \quad -\infty < x < \infty$$

Find the first quartile from this information.

8. The extreme quartiles of a normal variable are 10 and 30. Find its mean deviation.

9. For a normal variable, mean deviation is 48 and its third quartile is 120. Estimate its first quartile.

10. A normal variable X has the probability density function as,

$$f(x) = \frac{1}{10\sqrt{2\pi}} \; e^{-\frac{1}{2}\left(\frac{x-100}{10}\right)^2} \quad ; \quad -\infty < x < \infty$$

For this distribution, obtain the limits which include middle 68.26% of the observations.

11. The probability that the value of standard normal variable lies between 0 and *Z*-score (z_1) is 0.4950. Find the possible values of *Z*-score.

Section D

Answer the following questions :

1. Define normal distribution and state the characteristics of normal curve.

2. State the properties of normal distribution

3. State the properties of standard normal distribution.

4. A normal distribution has mean 50 and variance 9. Find the probability that

(1) The value of normal variable X lies between 50 and 53.

(2) The value of normal variable X lies between 47 and 53.

5. If X is a normal variable with mean 100 and standard deviation 15 then find the percentage of observations

(1) Having value more than 85.

(2) Having value less than 130.

6. The weight of randomly selected 500 adult persons from a region of a city follows normal distribution. The average weight of these persons is 55 kg and its standard deviation is 7 kg.

(1) Estimate the number of persons having weight between 41 kg to 62 kg.

(2) Estimate the number of persons having weight less than 41 kg.

7. If probabilities for the value of standard normal variable Z are as under then estimate the value of Z-score (z_1) :

(1) Area to the left of $Z = z_1$ is 0.9928.

(2) Area to the rightt of $Z = z_1$ is 0.0250.

8. If Z is a standard normal variable then estimate the value of Z-score (z_1) such that the following conditions are satisfied :

(1) Area to the left of $Z = z_1$ is 0.15.

(2) Area to the rightt of $Z = z_1$ is 0.75.

9. If Z is a standard normal variable and z_1 represents the Z-score then estimate the value of z_1 so that the following conditions are satisfied :

(1) $P(-2 \le Z \le z_1) = 0.2857$ (2) $P(z_1 \le Z \le 1.75) = 0.10$

10. The monthly production of units in a factory is normally distributed with mean μ and standard deviation σ. The Z-scores corresponding to the production of 2400 units and 1800 units are 1 and –0.5 respectively. Find its mean and standard deviation.

11. A normal variable X has the following probability density function

$$f(x) = \frac{1}{6\sqrt{2\pi}} \cdot e^{-\frac{1}{72}(x-100)^2} \quad ; \quad -\infty < x < \infty$$

For this distribution, obtain the estimated limits for the exact middle 50% of the observations.

12. For a normal distribution, the first quartile is 35 and its third quartile is 65. Estimate the limits that includes exactly middle 95.45% of the observations.

13. For a normal distribution, the third quartile and quartile deviations are 36 and 24 respectively. Find the mean of the distribution.

14. A normal variable X has mean 200 and variance 100.

(1) Estimate the values of extreme quartiles.

(2) Find the approximate value of quartile deviation.

(3) Find the approximate value of mean deviation.

Section E

Solve the following :

1. An amount of purchase of a customer in a mall of a city follows normal distribution with mean ₹ 800 and standard deviation ₹ 200. If a customer is selected at random then find the probabilities for the following events :

(1) Amount of purchase made by him is in between ₹ 850 to ₹ 1200.

(2) Amount of purchase made by him is in between ₹ 600 to ₹ 750.

2. The average weight of 500 persons of age between 20 years and 26 years of certain area is 55 kg and its variance is 100 $(kg)^2$. The weight of these persons follows normal distribution. According to the weight of persons they can be categorized as under :

(1) Person having weight more than 70 kg is in the fat persons group

(2) Person having weight between 50 kg to 60 kg is in the healthy persons group.

(3) Person having weight less than 35 kg is in the physically weak person's group

From this information, estimate the number of fat persons, number of healthy persons and number of physically weak persons in that area.

3. The average monthly expense of students residing in university hostel is ₹ 2000 and its standard deviation is ₹ 500. If the monthly expense of a student follows normal distribution then

(1) Find percentage of students having expense between ₹ 750 and ₹ 1250.

(2) Find percentage of students having expense more than ₹ 1800.

(3) Find percentage of students having expense less than ₹ 2400.

4. The monthly average salary of workers working in a production house is ₹ 10,000 and its standard deviation is ₹ 2000. By assuming that the monthly salary of a worker follows normal distribution, estimate the maximum salary of 20% of the workers having lowest salary. Also estimate the minimum salary of 10% of the workers having highest salary.

5. A normal distribution has mean 52 and variance 64. Obtain estimated limits which include exactly middle 25% of the observations.

6. In a big showroom of electronic items, on an average 52 electronic units are sold every week and its variance is 9 $(\text{unit})^2$. Sale of electronic items follows normal distribution. The probability that the sale of electronic items during a week out of 52 weeks is from x_1 units to 61 units is 0.1574. Estimate the value of x_1. Also estimate the number of weeks during which the sale of electronic items is more than 55 units.

7. It is known that on an average a person spends 61 minutes in a painting exhibition. If this time is normally distributed and 20% of the persons spent less than 30 minutes in the exhibition then find variance of the distribution. Also determine the probability that a person spends more than 90 minutes in the exhibition.

8. If the diameter of pipes produced by a company manufacturing pipes is 20 mm to 22 mm then it is accepted by specified group of customers. The standard deviation of produced pipes is 4 mm and it is known that 70% of the pipes produced in the unit have diameter more than 19.05 mm. Find the average diameter of the produced pipes. Also find the percentage of pipes rejected by specified group of customers.

Note : The diameter of the produced pipes follows normal distribution.

9. A normal variable X has mean 400 and variance 900. Find the fourth decile and 90th percentile for this distribution and also interpret the values.

10. A normal variable X has the following density function :

$$f(x) = \frac{1}{50\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-150}{50}\right)^2} \quad ; \quad -\infty < x < \infty$$

For this distribution,

(i) If $P(x_1 \le x \le 250) = 0.4772$ then estimate x_1.

(ii) If $P(75 < x \le x_2) = 0.3539$ then estimate x_2.

Section F

Solve the following :

1. An intelligence test is conducted for 500 children and it is found that the average marks are 68 and standard deviations is 22. If the marks obtained by the children is normally distributed then (1) find the number of children getting marks more than 68. (2) Find the percentage of children getting marks between 70 and 90. (3) Find the minimum score of most intelligent 50 children.

2. Age of 500 employees working in a private company follows normal distribution with mean 40 years and standard deviation 6 years. The company wants to reduce its staff by 25% in the following manner :

(i) To retrench 5% of the employees having minimum age

(ii) After retrenching 5% of employees having minimum age, next 10% of the employees are to be transferred to another company.

(iii) To retire 10% of employees having maximum age.

From this information, find the age of employees who are to be retrenched, transferred and retired from the company.

3. An entrance test of 200 marks is conducted for higher study. 20,000 students remain presents in the examination and the marks obtained by them follows normal distribution with mean 120 and standard deviation 20. The rules for the result are as under :

(a) Students who acquire less than 40 percent marks are failed.

(b) An additional test is conducted for the students acquiring marks between 40 percent and 48 percent.

(c) Students who acquire mark between 48 percent and 75 percent are called for personal interview.

(d) Students who acquire marks more than 75 percent get direct admission for the higher studies.

Find approximate number of students who : (1) failed in test (2) appeared for additional 100 marks test (3) appeared for personal interview and (4) got direct admission for the higher studies.

4. The monthly income of a group of persons follows normal distribution with mean ₹ 20,000 and standard deviation ₹ 5000. If the minimum monthly income of 50 richest person is ₹ 31,625 then how many persons are in the group? Also, what is the maximum income of 50 persons having lowest monthly income?

5. Analysis of result of 12th standard students of a school is as under :

Pass with distinction	:	15 % of total students
Pass without distinction	:	75 % of total students
Fail	:	10 % of total students

For passing the examination, minimum 40 % of the total marks and for distinction minimum 80 % marks are required. If the percentage of result of the students follows normal distribution then find mean and standard deviation and by using it determine the percentage marks for which 75 % of the students have less than that percentage marks.

6. The monthly bill amount of regular customers of a provision store follows normal distribution. If 7.78 % customers have monthly bill amount less than ₹ 3590 and 94.52 % customers have bill amount less than ₹ 5100 then determine the parameters of the normal distribution. Also determine the interval for monthly bill amount of exactly middle 95% customers.

7. A normal variable X has following probability density function :

$$f(x) = \frac{1}{\sqrt{5000\pi}} e^{-\frac{1}{5000}(x-75)^2} \quad ; \quad -\infty \le x \le \infty$$

From this, answer the following questions :

(i) If $P(60 \le x \le x_2) = 0.5670$ then find x_2.

(ii) If $P(x_1 \le x \le 125) = 0.3979$ then find x_1.

(iii) Find $P(|x - 50| \le 10)$.

8. A normal variable X has following probability density function :

$$f(x) = \text{ constant } \cdot e^{-\frac{1}{200}(x-50)^2} \quad ; \quad -\infty < x < \infty$$

From this distribution, answer the following questions :

(1) Find median.

(2) Find estimated values of the extreme quartiles.

(3) Find approximate value of quartile deviation.

(4) Find approximate value of mean deviation.

Johann Carl Friedrich Gauss
(1777 – 1855)

Carl Friedrich Gauss was a German mathematician who contributed significantly to many fields, including number theory, algebra, statistics, analysis, differential geometry, geodesy, geophysics, mechanics, electrostatics astronomy, matrix theory, and optics. He was referred to as the Princeps mathematicorum. (Latin, "the foremost of mathematicians") and "greatest mathematician since antiquity". Gauss had an exceptional influence in many fields of mathematics and science and has several theories and results in his name.

In the area of probability and statistics, Gauss introduced what is now known as Gaussian or normal distribution, the Gaussian function and the Gaussian error curve. He showed how probability could be represented by a bell shaped or "normal" curve, which peaks around the mean or expected value and quickly falls off towards plus/minus infinity, which is basic to descriptions of statistically distributed data.

4

Limit

Contents :

4.1 Introduction

We have studied function in 11th Standard. We studied in the chapter that when we substitute a particular value of a variable in the function, we got the corresponding value of the function. For example, if we substitute $x = 2$ in the function $f(x) = 2x + 3$, we get $f(2) = 7$. And if we substitute $x = 1$ in the function $f(x) = \frac{3-x}{3x+2}$, we get $f(1) = \frac{2}{5}$. But this is not possible for all functions and all values of x. Let us consider a function $f(x) = \frac{x^2-9}{x-3}$ and if we substitute $x = 3$ in $f(x)$, we get $f(3) = \frac{0}{0}$ which is an indeterminate value. To find approximate value of $f(3)$ for this function, we need to know the concept of limit of a function. So, limits can be used to approximate the value of a function when the value of the function is indeterminate.

We consider the following illustration to clarify the above concept.

Assume that we are watching a football game through internet. Unfortunately, the connection is choppy and we missed what happened at 14:00 (14 minutes after the commencement of match.)

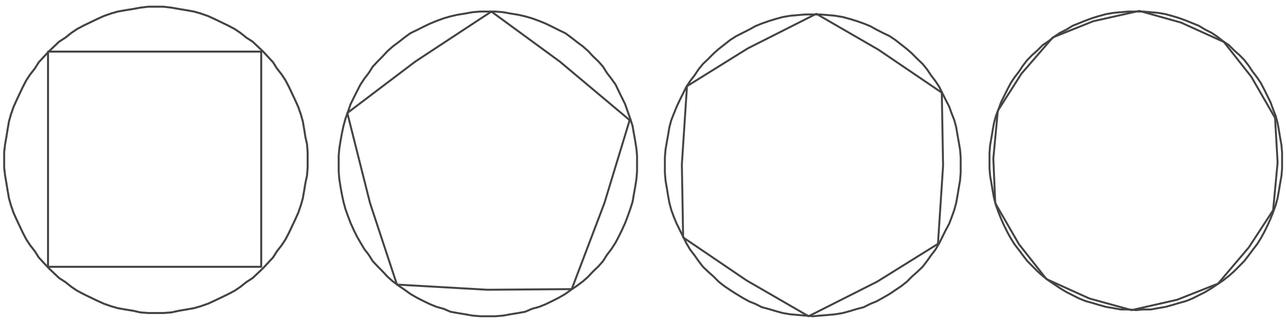
What would be the position of the ball at 14:00 ? We have seen the position of the ball at 13:58 (13 minutes and 58 seconds after the commencement of match), 13:59, 14:01, 14:02.

We will see the neighbouring instants of 14:00, (13:59 and 14:01) and estimate the position of the ball at 14:00. Our estimation is "At 14:00, the ball was somewhere between its position at 13:59 and 14:01." With a slow-motion camera, we might even say "At 14:00, the ball was somewhere between its positions at 13:59.99 and 14:00.01". It means that our estimation improves as we take closer and closer instants to 14:00. The approximate position of the ball thus obtained will be the limiting value of the position of the ball.

Thus, we can say that, "Limit is a method for finding confident approximate value."

We consider one more illustration.

Suppose we want to find the area of a circle. We can estimate the area of circle from the area of polygon drawn inside the circle.



4 side   5 side   6 side   10 side

We can see from the above figures that as the number of sides of polygon increases, area of the polygon approaches nearer the area of circle. The limiting value of the area of polygon is the best approximate value of the area of the circle.

Thus, limit can be used to approximate the unknown values by using its nearby values. Closer the neighbouring values, better is the approximation.

To understand the concept of limit, we shall understand the following basic terms.

4.2 Real Line and its Interval

Real line : The real line or real number line is a line where its points are the real numbers.

Interval : A set of real numbers between any two real numbers is an interval. We shall study different types of intervals.

Closed Interval : If $a \in R$, $b \in R$ and $a < b$ then the set of all real numbers between a and b including a and b is called a closed interval. It is denoted by $[a, b]$.

$$[a, b] = \{x \mid a \le x \le b, \; x \in R\}$$

Open Interval : If $a \in R$, $b \in R$ and $a < b$ then the set of all real numbers between a and b not including a and b is called an open interval. It is denoted by (a, b).

$$(a, b) = \{x \mid a < x < b, \; x \in R\}$$

Closed-Open Interval : If $a \in R$, $b \in R$ and $a < b$ then the set of all real numbers between a and b including a but not including b is called a closed open interval. It is denoted by $[a, b)$.

$[a, b) = \{x \mid a \le x < b,\ x \in R\}$

Open-Closed Interval : If $a \in R$, $b \in R$ and $a < b$ then set of all real numbers between a and b not including a but including b is called an open closed interval. It is denoted by $(a, b]$.

$(a, b] = \{x \mid a < x \le b,\ x \in R\}$

4.3 Modulus

If $x \in R$ then

$|x| = x$ if $x \ge 0$

$= -x$ if $x < 0$

Modulus of any real number is always non-negative.

e.g. $|3| = 3$, $|-4| = 4$, $|0| = 0$

Meaning of $|x - a| < \delta$ (*Delta*)

Using the definition of modulus

$|x-a| < \delta = (x-a) < \delta$ if $x \ge a$ **or** $x < a + \delta$ if $x \ge a$

$= (a-x) < \delta$ if $x < a$ **or** $x > a - \delta$ if $x < a$

$\therefore\ |x-a| < \delta \Leftrightarrow x \in (a-\delta,\ a+\delta)$

4.4 Neighbourhood

Any open interval containing a, $a \in R$ is called a **neighbourhood** of a.

δ neighbourhood of a :

If $a \in R$ and δ is non-negative real number then the open interval $(a-\delta,\ a+\delta)$ is called δ neighbourhood of a. It is denoted by $N(a, \delta)$.

Here, it can be understood that

$$N(a, \delta) = \{x \mid a - \delta < x < a + \delta,\ x \in R\}$$

$$= \{x \mid |x-a| < \delta,\ x \in R\}$$

δ neighbourhood of a can be expressed in the following different ways.

Neighbourhood form	Modulus form	Interval form
$N(a, \delta)$	$\|x-a\| < \delta$	$(a-\delta,\ a+\delta)$

Illustration 1 : Express $N(5, 0.2)$ in modulus and interval form.

Comparing $N(5, 0.2)$ with $N(a, \delta)$, we get $a = 5$ and $\delta = 0.2$.

Modulus form : $|x-a| < \delta$

Putting $a = 5$ and $\delta = 0.2$,

$N(5, 0.2) = |x - 5| < 0.2$

Interval form : $(a - \delta, a + \delta)$

Putting $a = 5$ and $\delta = 0.2$,

$N(5, 0.2) = (5 - 0.2, 5 + 0.2)$
$= (4.8, 5.2)$

Illustration 2 : Express 0.001 neighbourhood of 3 in modulus and interval form.

Comparing 0.001 neighbourhood of 3 with δ neighbourhood of a, we get $a = 3$ and $\delta = 0.001$.

Modulus form : $|x - a| < \delta$

Putting $a = 3$ and $\delta = 0.001$,

0.001 neighbourhood of 3 $= |x - 3| < 0.001$

Interval form : $(a - \delta, a + \delta)$

Putting $a = 3$ and $\delta = 0.001$,

0.001 neighbourhood of 3 $= (3 - 0.001, 3 + 0.001)$
$= (2.999, 3.001)$

Illustration 3 : Express $|x + 1| < 0.5$ in neighbourhood and interval form.

Comparing $|x + 1| < 0.5$ with $|x - a| < \delta$, we get $a = -1$ and $\delta = 0.5$.

Neighbourhood form : $N(a, \delta)$

Putting $a = -1$ and $\delta = 0.5$,

$|x + 1| < 0.5 = N(-1, 0.5)$

Interval form : $(a - \delta, a + \delta)$

Putting $a = -1$ and $\delta = 0.5$,

$|x + 1| < 0.5 = (-1 - 0.5, -1 + 0.5)$
$= (-1.5, -0.5)$

Illustration 4 : Express $(0.9, 1.1)$ in neighbourhood and modulus form.

Comparing $(0.9, 1.1)$ with $(a - \delta, a + \delta)$, we get $a - \delta = 0.9$ and $a + \delta = 1.1$.

Adding $a - \delta = 0.9$ and $a + \delta = 1.1$, we get $2a = 2$ $\therefore a = 1$.

Putting $a = 1$ in $a + \delta = 1.1$, we get $\delta = 0.1$.

Neighbourhood form : $N(a, \delta)$

Putting $a = 1$ and $\delta = 0.1$,

$$(0.9,\ 1.1) = N(1,\ 0.1)$$

Modulus form : $|x-a| < \delta$

Putting $a = 1$ and $\delta = 0.1$,

$$(0.9,\ 1.1) = |x-1| < 0.1$$

Punctured δ neighbourhood of *a* :

If $a \in R$ and δ is a non-negative real number then the open interval $(a-\delta,\ a+\delta) - \{a\}$ is called punctured δ neighbourhood of *a*. It is denoted by $N^*(a,\ \delta)$.

Here, it can be understood that

$$N^*(a,\ \delta) = N(a,\ \delta) - \{a\}$$

$$= \{x \mid a-\delta < x < a+\delta,\ x \neq a,\ x \in R\}$$

$$= \{x \mid |x-a| < \delta,\ x \neq a,\ x \in R\}$$

e.g. $N^*(5,\ 2) = N(5,\ 2) - \{5\}$

$$= \{x \mid 3 < x < 7,\ x \neq 5,\ x \in R\}$$

$$= \{x \mid |x-5| < 2,\ x \neq 5,\ x \in R\}$$

EXERCISE 4.1

1. Express the following in modulus and interval form :

(1) 0.4 neighbourhood of 4 (2) 0.02 neighbourhood of 2

(3) 0.05 neighbourhood of 0 (4) 0.001 neighbourhood of –1

2. Express the following in interval and neighbourhood form :

(1) $|x-2| < 0.01$ (2) $|x+5| < 0.1$

(3) $|x| < \frac{1}{3}$ (4) $|x+3| < 0.15$

3. Express the following in modulus and neighbourhood form :

(1) $3.8 < x < 4.8$ (2) $1.95 < x < 2.05$

(3) $-0.4 < x < 1.4$ (4) $1.998 < x < 2.002$

4. Express $N(16,\ 0.5)$ in the interval and modulus form.

5. If $N(3,\ b) = (2.95,\ k)$ then find the values of b and k.

6. If $|x-10| < k_1 = (k_2,\ 10.01)$ then find the values of k_1 and k_2.

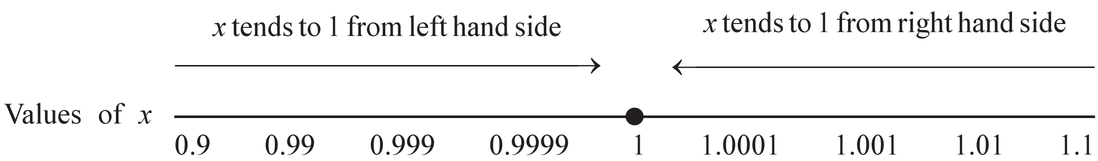
*

Meaning of $x \to a$:

If the value of variable x is brought very close to a number 'a' by increasing or decreasing its value then it can be said that x tends to a. It is denoted by $x \to a$.

It is necessary here to note that $x \to a$ means value of x approaches very close to a but it will not be equal to a.

e.g. Let us understand the meaning of $x \to 1$.

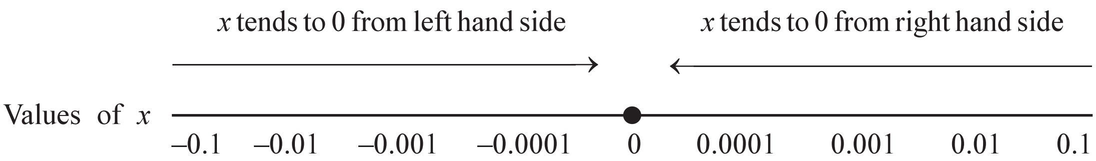


Meaning of $x \to 0$:

If by decreasing the positive value of a variable x or by increasing negative value of the variable x, the value of x is brought very close to '0' then it can be said that x tends to 0. It is denoted by $x \to 0$.

It is necessary here to note that $x \to 0$ means, the value of x approaches very close to 0 but it will not be equal to 0.

Let us understand the meaning of $x \to 0$.



4.5 Limit of a function

When the value of a variable x is brought closer and closer to a number 'a', the value of function $f(x)$ reaches closer and closer to a definite number 'l' then we can say that as x tends to a, $f(x)$ tends to l. i.e. as $x \to a$, $f(x) \to l$. Symbolically it can be written as $\lim_{x \to a} f(x) = l$. l is called the limiting value of the function.

Definition : The function $f(x)$ has a limit l as x tends to 'a' if for each given predetermined $\varepsilon > 0$, however small, we can find a positive number δ such that $\left| f(x) - l \right| < \varepsilon$ (*Epsilon*) for all x such that $| x - a | < \delta$.

Now, we shall understand how limit of a function is obtained.

Suppose, we want to find the value of the function $f(x) = \frac{x^2-1}{x-1}$ at $x = 1$.

If we put $x = 1$ in $f(x) = \frac{x^2-1}{x-1}$ we get $f(1) = \frac{0}{0}$ which is indeterminate. So, we cannot find the value of $f(1)$ but assuming value of x very close to 1, we can approximate the value of $f(1)$. Let us see the changes in $f(x)$ as x tends to 1.

x (towards 1 from LHS of 1)	$f(x)$	x (towards 1 from RHS of 1)	$f(x)$
0.9	1.9	1.1	2.1
0.99	1.99	1.01	2.01
0.999	1.999	1.001	2.001
0.9999	1.9999	1.0001	2.0001
⋮	⋮	⋮	⋮

We can assume any value of x close to 1. Generally, we start with a value at a distance 0.1 on both sides of $x = 1$. i.e. we start with $x = 0.9$ and 1.1 and bring values of x closer to 1 from both the sides.

It is clear from the table that when the value of x is brought nearer to 1 by increasing or decreasing its values, the value of $f(x)$ approaches to 2.

This can symbollically be expressed as $\lim\limits_{x \to 1} \frac{x^2-1}{x-1} = 2$.

Limit of a function is obtained by putting different values of x in $f(x)$ and tabulating them. So, this method of obtaining the limit of a function is called a **tabular method.**

Illustration 5 : Find $\lim\limits_{x \to 3} 2x+5$ by tabular method.

We have $f(x) = 2x+5$. We shall take the values of x very near to 3 and prepare a table in the following way :

x	$f(x)$	x	$f(x)$
2.9	10.8	3.1	11.2
2.99	10.98	3.01	11.02
2.999	10.998	3.001	11.002
2.9999	10.9998	3.0001	11.0002
⋮	⋮	⋮	⋮

It is clear from the table that when the value of x is brought nearer to 3 by increasing or decreasing its values, the value of $f(x)$ approaches to 11. That is, when $x \to 3$, $f(x) \to 11$.

$\therefore \lim\limits_{x \to 3} 2x+5 = 11$

Illustration 6 : Find $\lim_{x \to -1} \frac{x^2-1}{x+1}$, $x \in R - \{-1\}$ by preparing table.

We have $f(x) = \frac{x^2-1}{x+1}$. We shall take the values of x very near to –1 and prepare a table in the following way :

x	$f(x)$	x	$f(x)$
–1.1	–2.1	–0.9	–1.9
–1.01	–2.01	–0.99	–1.99
–1.001	–2.001	–0.999	–1.999
–1.0001	–2.0001	–0.9999	–1.9999
. . .	. . .	. . .	. . .

It is clear from the table that when the value of x is brought nearer to –1 by increasing or decreasing its value, the value of $f(x)$ approaches to –2. That is, when $x \to -1$, $f(x) \to -2$.

$\therefore \lim_{x \to -1} \frac{x^2-1}{x+1} = -2$

Illustration 7 : Find $\lim_{x \to 0} \frac{2x^2+3x}{x}$, $x \in R - \{0\}$ using tabular method.

We have $f(x) = \frac{2x^2+3x}{x}$. We shall take the values of x very near to 0 and prepare a table in the following way :

x	$f(x)$	x	$f(x)$
–0.1	2.8	0.1	3.2
–0.01	2.98	0.01	3.02
–0.001	2.998	0.001	3.002
–0.0001	2.9998	0.0001	3.0002
. . .	. . .	. . .	. . .

It is clear from the table that when the value of x is brought nearer to 0 by increasing or decreasing its value, the value of $f(x)$ approaches to 3. That is, when $x \to 0$, $f(x) \to 3$.

$\therefore \lim_{x \to 0} \frac{2x^2+3x}{x} = 3$

Illustration 8 : Find $\lim_{x \to 1} \frac{1}{x-1}$, $x \in R - \{1\}$ by tabular method.

We have $f(x) = \frac{1}{x-1}$. We shall take the values of x very near to 1 and prepare a table in the following way :

x	$f(x)$	x	$f(x)$
0.9	–10	1.1	10
0.99	–100	1.01	100
0.999	–1000	1.001	1000
0.9999	–10000	1.0001	10000
. . .	. . .	. . .	. . .

It is clear from the table that when the value of x is brought nearer to 1 by increasing or decreasing its value, the value of $f(x)$ does not approach to a particular value. That is, when $x \to 1$, $f(x)$ does not tend to a particular value. Thus, limit of the function does not exist.

$\therefore \lim_{x \to 1} \frac{1}{x-1}$ does not exist.

Illustration 9 : Find $\lim_{x \to 2} \frac{3x^2 - 4x - 4}{x^2 - 4}$, $x \in R - \{2\}$ by tabular method.

We have $f(x) = \frac{3x^2 - 4x - 4}{x^2 - 4}$. We can obtain the value of limit as calculated in previous illustrations. But for simplification we shall obtain the value of limit of $f(x)$ after eliminating the common factor $(x-2)$ from numerator and denominator.

$$\lim_{x \to 2} \frac{3x^2 - 4x - 4}{x^2 - 4} = \lim_{x \to 2} \frac{(x-2)(3x+2)}{(x-2)(x+2)}$$

$$= \lim_{x \to 2} \frac{3x+2}{x+2} \qquad (\because \ x - 2 \neq 0)$$

We shall take the values of x very near to 2 and prepare a table in the following way :

x	$f(x)$	x	$f(x)$
1.9	1.9744	2.1	2.02439
1.99	1.9975	2.01	2.002494
1.999	1.9997	2.001	2.0002499
1.9999	1.9999	2.0001	2.000025
. . .	. . .	. . .	. . .

It is clear from the table that when the value of x is brought very near to 2 by increasing or decreasing its value, the value of $f(x)$ approaches to 2. That is, when $x \to 2$, $f(x) \to 2$.

$$\therefore \lim_{x \to 2} \frac{3x^2 - 4x - 4}{x^2 - 4} = 2$$

EXERCISE 4.2

1. Find the values of the following using tabular method :

(1) $\lim_{x \to 1} 2x+1$

(2) $\lim_{x \to 3} \frac{x^2 - 2x - 3}{x - 3}$

(3) $\lim_{x \to 2} \frac{2x^2 + 3x - 14}{x - 2}$

(4) $\lim_{x \to -3} \frac{2x^2 + 9x + 9}{x + 3}$

(5) $\lim_{x \to 2} x$

2. Using tabular method, show that $\lim_{x \to 3} \frac{2}{x - 3}$ does not exist.

3. If $y = \frac{x^2 + x - 6}{x - 2}$, show that as $x \to 2$ then $y \to 5$ using tabular method.

4. If $y = 5 - 2x$, show that as $x \to -1$ then $y \to 7$ using tabular method.

*

4.6 Working rules of limit

The following rules will be accepted without proof :

If $f(x)$ and $g(x)$ are two real valued functions of a real variable x and $\lim_{x \to a} f(x) = l$ and $\lim_{x \to a} g(x) = m$, then

(1) $\lim_{x \to a} [f(x) \pm g(x)] = l \pm m$

The limit of the sum (or difference) of two functions is equal to the sum (or difference) of their limits.

(2) $\lim_{x \to a} [f(x) \times g(x)] = l \times m$

The limit of the product of two functions is equal to the product of their limits.

(3) $\lim_{x \to a} \left[\frac{f(x)}{g(x)}\right] = \frac{l}{m}, \quad m \neq 0$

The limit of the division of two functions is equal to the division of their limits, provided the limit of the function in denominator is not zero.

(4) $\lim_{x \to a} k\, f(x) = kl$, k is the constant.

The limit of the product of a function with a constant is equal to the product of the limit of the function with the same constant.

4.7 Standard forms of limit

(1) **Limit of a polynominl**

Let $f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$ then using the working rules of limit

$$\lim_{x \to b} f(x) = a_0 + a_1 b + a_2 b^2 + ... + a_n b^n$$

(2) $\lim_{x \to a} \left[\frac{x^n - a^n}{x - a}\right] = n a^{n-1}, \quad n \in Q$

We will see some illustrations based on the standard forms and working rules of limit.

Illustration 10 : Find the value of $\lim_{x \to 0} \frac{x^2 + 5x + 6}{x^2 + 2x + 3}$.

$$\lim_{x \to 0} \frac{x^2 + 5x + 6}{x^2 + 2x + 3} = \frac{(0)^2 + 5(0) + 6}{(0)^2 + 2(0) + 3}$$

$$= \frac{6}{3}$$

$$= 2$$

Illustration 11 : Find the value of $\lim_{x \to 2} \frac{2x + 3}{x - 1}$.

$$\lim_{x \to 2} \frac{2x + 3}{x - 1} = \frac{2(2) + 3}{2 - 1}$$

$$= \frac{7}{1}$$

$$= 7$$

Illustration 12 : Find the value of $\lim_{x \to 3} \frac{x^2 - 2x - 3}{x^2 - 5x + 6}$.

If we put $x = 3$ in the function $f(x)$, we get the value of the function as $\frac{0}{0}$, which is indeterminate. Hence, we shall factorize numerator and denominator. Since $x \to 3$, $(x - 3)$ will be a common factor of numerator and denominator.

Note : If we put $x = a$ in the given function and we get $\frac{0}{0}$ then $(x - a)$ will be the common factor of numerator and denominator.

Numerator $= x^2 - 2x - 3$

$= x^2 - 3x + x - 3$

$= x(x - 3) + 1(x - 3)$

$= (x - 3)(x + 1)$

Denominator $= x^2 - 5x + 6$

$= x^2 - 3x - 2x + 6$

$= x(x - 3) - 2(x - 3)$

$= (x - 3)(x - 2)$

Now, $\lim_{x \to 3} \frac{x^2 - 2x - 3}{x^2 - 5x + 6} = \lim_{x \to 3} \frac{(x - 3)(x + 1)}{(x - 3)(x - 2)}$

$= \lim_{x \to 3} \frac{(x + 1)}{(x - 2)} \qquad (\because x - 3 \neq 0)$

$$= \frac{3+1}{3-2}$$

$$= \frac{4}{1}$$

$$= 4$$

Illustration 13 : Find the value of $\lim_{x \to 1} \frac{2x^2 + x - 3}{x^2 - 1}$.

If we put $x = 1$ in the function $f(x)$, we get the value of the function as $\frac{0}{0}$, which is indeterminate.

Numerator $= 2x^2 + x - 3$

$= 2x^2 + 3x - 2x - 3$

$= x(2x + 3) - 1(2x + 3)$

$= (2x + 3)\ (x - 1)$

Denominator $= x^2 - 1$

$= (x + 1)\ (x - 1)$

Now, $\lim_{x \to 1} \frac{2x^2 + x - 3}{x^2 - 1} = \lim_{x \to 1} \frac{(2x + 3)(x - 1)}{(x + 1)(x - 1)}$

$= \lim_{x \to 1} \frac{2x + 3}{x + 1} \qquad (\because\ x - 1 \neq 0)$

$= \frac{2(1) + 3}{1 + 1}$

$= \frac{5}{2}$

Illustration 14 : Find the value of $\lim_{x \to -3} \frac{2x^2 + 7x + 3}{3x^2 + 8x - 3}$.

If we put $x = -3$ in the function $f(x)$, we get the value of the function as $\frac{0}{0}$, which is indeterminate.

Numerator $= 2x^2 + 7x + 3$

$= 2x^2 + 6x + x + 3$

$= 2x(x + 3) + 1(x + 3)$

$= (x + 3)\ (2x + 1)$

Denominator $= 3x^2 + 8x - 3$

$= 3x^2 + 9x - x - 3$

$= 3x(x + 3) - 1(x + 3)$

$= (x + 3)\ (3x - 1)$

Now, $\lim_{x \to -3} \frac{2x^2 + 7x + 3}{3x^2 + 8x - 3} = \lim_{x \to -3} \frac{(x+3)(2x+1)}{(x+3)(3x-1)}$

$= \lim_{x \to -3} \frac{2x+1}{3x-1} \qquad (\because \ x + 3 \neq 0)$

$= \frac{2(-3)+1}{3(-3)-1}$

$= \frac{-6+1}{-9-1}$

$= \frac{-5}{-10}$

$= \frac{1}{2}$

Illustration 15 : Find the value of $\lim_{x \to -\frac{1}{2}} \frac{2x^2 - x - 1}{4x^2 + 8x + 3}$.

If we put $x = -\frac{1}{2}$ in the function $f(x)$, we get the value of the function as $\frac{0}{0}$, which is indeterminate.

Numerator $= 2x^2 - x - 1$

$= 2x^2 - 2x + x - 1$

$= 2x(x-1) + 1(x-1)$

$= (x-1)(2x+1)$

Denominator $= 4x^2 + 8x + 3$

$= 4x^2 + 6x + 2x + 3$

$= 2x(2x+3) + 1(2x+3)$

$= (2x+3)(2x+1)$

Now, $\lim_{x \to -\frac{1}{2}} \frac{2x^2 - x - 1}{4x^2 + 8x + 3} = \lim_{x \to -\frac{1}{2}} \frac{(x-1)(2x+1)}{(2x+3)(2x+1)}$

$= \lim_{x \to -\frac{1}{2}} \frac{x-1}{2x+3} \qquad (\because \ 2x + 1 \neq 0)$

$= \frac{-\frac{1}{2} - 1}{2\left(-\frac{1}{2}\right) + 3}$

$= \frac{-\frac{3}{2}}{-1+3}$

$= \frac{-\frac{3}{2}}{2}$

$= -\frac{3}{4}$

Illustration 16 : Find the value of $\lim\limits_{x \to 2} \left[\dfrac{1}{x-2} - \dfrac{2}{x^2-2x} \right]$.

$$\lim_{x \to 2} \left[\frac{1}{x-2} - \frac{2}{x^2-2x} \right] = \lim_{x \to 2} \left[\frac{1}{x-2} - \frac{2}{x(x-2)} \right]$$

$$= \lim_{x \to 2} \left[\frac{x-2}{x(x-2)} \right]$$

$$= \lim_{x \to 2} \frac{1}{x} \qquad (\because \; x - 2 \neq 0)$$

$$= \frac{1}{2}$$

Illustration 17 : Find the value of $\lim\limits_{x \to 0} \dfrac{1}{x} \left[\dfrac{2x+3}{3x-5} + \dfrac{3}{5} \right]$.

$$\lim_{x \to 0} \frac{1}{x} \left[\frac{2x+3}{3x-5} + \frac{3}{5} \right] = \lim_{x \to 0} \frac{1}{x} \left[\frac{5(2x+3) + 3(3x-5)}{5(3x-5)} \right]$$

$$= \lim_{x \to 0} \frac{1}{x} \left[\frac{10x + 15 + 9x - 15}{5(3x-5)} \right]$$

$$= \lim_{x \to 0} \frac{1}{x} \left[\frac{19x}{5(3x-5)} \right]$$

$$= \lim_{x \to 0} \frac{19}{5(3x-5)} \qquad (\because \; x \neq 0)$$

$$= \frac{19}{5[3(0) - 5]}$$

$$= \frac{19}{5(-5)}$$

$$= -\frac{19}{25}$$

Illustration 18 : If $f(x) = x^2 + x$ then find the value of $\lim\limits_{x \to 2} \dfrac{f(x) - f(2)}{x^2 - 4}$.

Here, $f(x) = x^2 + x$

$\therefore \; f(2) = (2)^2 + 2$

$= 4 + 2$

$= 6$

Now, $\lim_{x\to 2} \dfrac{f(x) - f(2)}{x^2 - 4} = \lim_{x\to 2} \dfrac{\left(x^2 + x\right) - 6}{x^2 - 4}$

Numerator $= x^2 + x - 6$

$= x^2 + 3x - 2x - 6$

$= x(x+3) - 2(x+3)$

$= (x+3)(x-2)$

Denominator $= x^2 - 4$

$= (x+2)(x-2)$

So, $\lim_{x\to 2} \dfrac{x^2 + x - 6}{x^2 - 4} = \lim_{x\to 2} \dfrac{(x+3)(x-2)}{(x+2)(x-2)}$

$= \lim_{x\to 2} \dfrac{x+3}{x+2} \quad (\because \ x - 2 \neq 0)$

$= \dfrac{2+3}{2+2}$

$= \dfrac{5}{4}$

Illustration 19 : If $f(x) = x^3$ then find the value of $\lim_{h\to 0} \dfrac{f(3+h) - f(3-h)}{2h}$.

Here, $f(x) = x^3$

$\therefore \ f(3+h) = (3+h)^3$

$= 27 + 27h + 9h^2 + h^3$

and

$f(3-h) = (3-h)^3$

$= 27 - 27h + 9h^2 - h^3$

Now, $\lim_{h\to 0} \dfrac{f(3+h) - f(3-h)}{2h} = \lim_{h\to 0} \dfrac{\left(27 + 27h + 9h^2 + h^3\right) - \left(27 - 27h + 9h^2 - h^3\right)}{2h}$

$= \lim_{h\to 0} \dfrac{27 + 27h + 9h^2 + h^3 - 27 + 27h - 9h^2 + h^3}{2h}$

$$= \lim_{h \to 0} \frac{54h + 2h^3}{2h}$$

$$= \lim_{h \to 0} \frac{h\left(54 + 2h^2\right)}{2h}$$

$$= \lim_{h \to 0} \frac{54 + 2h^2}{2} \qquad (\because \ h \neq 0)$$

$$= \frac{54 + 2(0)^2}{2}$$

$$= \frac{54}{2}$$

$$= 27$$

Illustration 20 : Find the value of $\lim_{x \to 0} \frac{\sqrt{3+x} - \sqrt{3}}{x}$.

$$\lim_{x \to 0} \frac{\sqrt{3+x} - \sqrt{3}}{x}$$

(multiplying numerator and denominator by $\sqrt{3+x} + \sqrt{3}$)

$$= \lim_{x \to 0} \frac{\sqrt{3+x} - \sqrt{3}}{x} \times \frac{\sqrt{3+x} + \sqrt{3}}{\sqrt{3+x} + \sqrt{3}}$$

$$= \lim_{x \to 0} \frac{\left(\sqrt{3+x}\right)^2 - \left(\sqrt{3}\right)^2}{x\left(\sqrt{3+x} + \sqrt{3}\right)}$$

$$= \lim_{x \to 0} \frac{3 + x - 3}{x\left(\sqrt{3+x} + \sqrt{3}\right)}$$

$$= \lim_{x \to 0} \frac{x}{x\left(\sqrt{3+x} + \sqrt{3}\right)}$$

$$= \lim_{x \to 0} \frac{1}{\left(\sqrt{3+x} + \sqrt{3}\right)} \qquad (\because \ x \neq 0)$$

$$= \frac{1}{\sqrt{3+0} + \sqrt{3}}$$

$$= \frac{1}{\sqrt{3} + \sqrt{3}}$$

$$= \frac{1}{2\sqrt{3}}$$

Illustration 21 : Find the value of $\lim\limits_{x \to 2} \dfrac{\sqrt{x+7} - 3}{x - 2}$.

$$\lim_{x \to 2} \frac{\sqrt{x+7} - 3}{x - 2}$$

(multiplying numerator and denominator by $\sqrt{x+7} + 3$)

$$= \lim_{x \to 2} \frac{\sqrt{x+7} - 3}{x - 2} \times \frac{\sqrt{x+7} + 3}{\sqrt{x+7} + 3}$$

$$= \lim_{x \to 2} \frac{\left(\sqrt{x+7}\right)^2 - (3)^2}{(x-2)\left(\sqrt{x+7} + 3\right)}$$

$$= \lim_{x \to 2} \frac{x + 7 - 9}{(x-2)\left(\sqrt{x+7} + 3\right)}$$

$$= \lim_{x \to 2} \frac{x - 2}{(x-2)\left(\sqrt{x+7} + 3\right)}$$

$$= \lim_{x \to 2} \frac{1}{\sqrt{x+7} + 3} \qquad (\because \ x - 2 \neq 0)$$

$$= \frac{1}{\sqrt{2+7} + 3}$$

$$= \frac{1}{\sqrt{9} + 3}$$

$$= \frac{1}{3 + 3}$$

$$= \frac{1}{6}$$

Illustration 22 : Find the value of $\lim\limits_{h \to 0} \dfrac{f(x+h) - f(x)}{h}$ where $f(x) = \sqrt{x}$, $x > 0$.

$$f(x) = \sqrt{x}$$

$$\therefore \ f(x+h) = \sqrt{x+h}$$

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h}$$

(multiplying numerator and denominator by $\sqrt{x+h} + \sqrt{x}$)

$$= \lim_{h \to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \times \frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}}$$

$$= \lim_{h \to 0} \frac{\left(\sqrt{x+h}\right)^2 - \left(\sqrt{x}\right)^2}{h\left(\sqrt{x+h} + \sqrt{x}\right)}$$

$$= \lim_{h \to 0} \frac{x + h - x}{h\left(\sqrt{x+h} + \sqrt{x}\right)}$$

$$= \lim_{h \to 0} \frac{h}{h\left(\sqrt{x+h} + \sqrt{x}\right)}$$

$$= \lim_{h \to 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} \qquad (\because \quad h \neq 0)$$

$$= \frac{1}{\sqrt{x+0} + \sqrt{x}}$$

$$= \frac{1}{\sqrt{x} + \sqrt{x}}$$

$$= \frac{1}{2\sqrt{x}}$$

Illustration 23 : Find the value of $\lim_{x \to 2} \frac{x^3 - 8}{\sqrt{x} - \sqrt{2}}$ **.**

$$\lim_{x \to 2} \frac{x^3 - 8}{\sqrt{x} - \sqrt{2}}$$

(multiplying numerator and denominator by $\sqrt{x} + \sqrt{2}$)

$$= \lim_{x \to 2} \frac{x^3 - 8}{\sqrt{x} - \sqrt{2}} \times \frac{\sqrt{x} + \sqrt{2}}{\sqrt{x} + \sqrt{2}}$$

$$= \lim_{x \to 2} \frac{(x-2)\left(x^2 + 2x + 4\right)\left(\sqrt{x} + \sqrt{2}\right)}{\left(\sqrt{x}\right)^2 - \left(\sqrt{2}\right)^2}$$

$$= \lim_{x \to 2} \frac{(x-2)\left(x^2 + 2x + 4\right)\left(\sqrt{x} + \sqrt{2}\right)}{(x-2)}$$

$$= \lim_{x \to 2} \left(x^2 + 2x + 4\right)\left(\sqrt{x} + \sqrt{2}\right) \qquad (\because \quad x - 2 \neq 0)$$

$$= \left[(2)^2 + 2(2) + 4\right]\left[\sqrt{2} + \sqrt{2}\right]$$

$$= (4 + 4 + 4)\left(2\sqrt{2}\right)$$

$$= 12\left(2\sqrt{2}\right)$$

$$= 24\sqrt{2}$$

Illustration 24 : Find the value of $\lim_{x\to 1} \frac{\sqrt{x+3}-2}{\sqrt{x+8}-3}$.

$$\lim_{x\to 1} \frac{\sqrt{x+3}-2}{\sqrt{x+8}-3}$$

(multiplying numerator and denominator by $\sqrt{x+3}+2$ and $\sqrt{x+8}+3$)

$$= \lim_{x\to 1} \frac{\sqrt{x+3}-2}{\sqrt{x+8}-3} \times \frac{\sqrt{x+3}+2}{\sqrt{x+3}+2} \times \frac{\sqrt{x+8}+3}{\sqrt{x+8}+3}$$

$$= \lim_{x\to 1} \frac{\left(\sqrt{x+3}\right)^2 - (2)^2}{\left(\sqrt{x+8}\right)^2 - (3)^2} \times \frac{\sqrt{x+8}+3}{\sqrt{x+3}+2}$$

$$= \lim_{x\to 1} \frac{(x+3-4) \times \left(\sqrt{x+8}+3\right)}{(x+8-9) \times \left(\sqrt{x+3}+2\right)}$$

$$= \lim_{x\to 1} \frac{(x-1)\left(\sqrt{x+8}+3\right)}{(x-1)\left(\sqrt{x+3}+2\right)}$$

$$= \lim_{x\to 1} \frac{\sqrt{x+8}+3}{\sqrt{x+3}+2} \qquad (\because \ x-1 \neq 0)$$

$$= \frac{\sqrt{1+8}+3}{\sqrt{1+3}+2}$$

$$= \frac{\sqrt{9}+3}{\sqrt{4}+2}$$

$$= \frac{3+3}{2+2}$$

$$= \frac{6}{4}$$

$$= \frac{3}{2}$$

Illustration 25 : Find the value of $\lim_{x\to 2} \frac{x^5-32}{x-2}$.

$$\lim_{x\to 2} \frac{x^5-32}{x-2} = \lim_{x\to 2} \frac{x^5-2^5}{x-2}$$

$$= 5(2)^{5-1} \qquad \left[\because \lim_{x\to a} \frac{x^n-a^n}{x-a} = na^{n-1}\right]$$

$$= 5(2)^4$$

$$= 5(16)$$

$$= 80$$

Illustration 26 : Find the value of $\lim\limits_{x \to 3} \dfrac{x^5 - 243}{x^3 - 27}$.

$$\lim_{x \to 3} \frac{x^5 - 243}{x^3 - 27} = \lim_{x \to 3} \frac{x^5 - 3^5}{x^3 - 3^3}$$

(multiplying numerator and denominator by $(x - 3)$)

$$= \lim_{x \to 3} \frac{x^5 - 3^5}{x - 3} \times \frac{x - 3}{x^3 - 3^3}$$

$$= \lim_{x \to 3} \left[\frac{x^5 - 3^5}{x - 3} \div \frac{x^3 - 3^3}{x - 3} \right]$$

$$= \frac{5(3)^{5-1}}{3(3)^{3-1}} \quad \left[\because \lim_{x \to a} \frac{x^n - a^n}{x - a} = n a^{n-1} \right]$$

$$= \frac{5(3)^4}{3(3)^2}$$

$$= \frac{5 \times 81}{3 \times 9}$$

$$= 15$$

Illustration 27 : Find the value of $\lim\limits_{x \to -2} \dfrac{x^7 + 128}{x + 2}$.

$$\lim_{x \to -2} \frac{x^7 + 128}{x + 2} = \lim_{x \to -2} \frac{x^7 - (-2)^7}{x - (-2)}$$

$$= 7(-2)^{7-1} \quad \left[\because \lim_{x \to a} \frac{x^n - a^n}{x - a} = n a^{n-1} \right]$$

$$= 7(-2)^6$$

$$= 7(64)$$

$$= 448$$

Illustration 28 : Find the value of $\lim\limits_{h \to 0} \dfrac{(x + h)^5 - x^5}{h}$.

$$\lim_{h \to 0} \frac{(x + h)^5 - x^5}{h}$$

(Taking $x + h = t$, when $h \to 0$, $t \to x$)

$$= \lim_{t \to x} \frac{t^5 - x^5}{t - x} \quad (\because x + h = t)$$

$$= 5(x)^{5-1} \quad \left[\because \lim_{x \to a} \frac{x^n - a^n}{x - a} = n a^{n-1} \right]$$

$$= 5x^4$$

Illustration 29 : Find the value of $\lim_{x \to 0} \frac{\sqrt[n]{x+1} - 1}{x}$.

$$\lim_{x \to 0} \frac{\sqrt[n]{x+1} - 1}{x} = \lim_{x \to 0} \frac{(x+1)^{\frac{1}{n}} - 1^{\frac{1}{n}}}{x}$$

(Taking $x + 1 = t$, when $x \to 0$, $t \to 1$)

$$= \lim_{t \to 1} \frac{t^{\frac{1}{n}} - 1^{\frac{1}{n}}}{t - 1} \quad (\because x + 1 = t \quad \therefore x = t - 1)$$

$$= \frac{1}{n} (1)^{\frac{1}{n} - 1} \qquad \left[\because \lim_{x \to a} \frac{x^n - a^n}{x - a} = n a^{n-1}\right]$$

$$= \frac{1}{n} \times 1$$

$$= \frac{1}{n}$$

Summary

- **Neighbourhood :** Let $a \in R$. Any open interval containing 'a' is called a neighbourhood of 'a'.
- δ **neighbourhood of** a **:** If $a \in R$ and δ is a non-negative real number then the open interval $(a - \delta, a + \delta)$ is called δ neighbourhood of a.
- **Meaning of** $x \to a$ **:** If the value of a variable x is brought very close to a number 'a' by increasing or decreasing then it can be said that x tends to a. It is symbolically denoted by $x \to a$.
- **Meaning of** $x \to 0$ **:** If by decreasing the positive values of a variable x or by increasing negative values of a variable x is brought very close to '0' then it can be said that x tends to 0. It is symbolically denoted by $x \to 0$.
- **Limit of a function**

  The function $f(x)$ has limit l as x tends to 'a' if for each given predetermined $\varepsilon > 0$, however small, we can find a positive number δ such that $|f(x) - l| < \varepsilon$ for all x such that $|x - a| < \delta$.

List of Formulae :

- $\lim_{x \to a} \left[f(x) \pm g(x) \right] = l \pm m$
- $\lim_{x \to a} \left[f(x) \times g(x) \right] = l \times m$
- $\lim_{x \to a} \left[\frac{f(x)}{g(x)} \right] = \frac{l}{m}, \quad m \neq 0$
- $\lim_{x \to a} k\, f(x) = kl, \quad k$ is constant.
- If $f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$ then

  $\lim_{x \to b} f(x) = a_0 + a_1 b + a_2 b^2 + ... + a_n b^n$
- $\lim_{x \to a} \frac{x^n - a^n}{x - a} = n a^{n-1}, \quad n \in Q$

EXERCISE 4

Section A

Choose the correct option for the following multiple choice questions :

1. What is the modulus form of 0.3 neighbourhood of 3 ?

(a) $|x - 0.3| < 3$ (b) $|x - 3| < 0.3$ (c) $|x + 3| < 0.3$ (d) $|x - 3| > 0.3$

2. What is the interval form of 0.02 neighbourhood of –2 ?

(a) $(1.98, 2.02)$ (b) $(-1.98, 2.02)$ (c) $(-2.02, -1.98)$ (d) $(-2.02, 1.98)$

3. What is the interval form of $|x - 5| < 0.25$?

(a) $(4.75, 5.25)$ (b) $(-4.75, +5.25)$ (c) $(-5.25, -4.75)$ (d) $(-5.25, 4.75)$

4. What is the interval form of $|2x + 1| < \frac{1}{5}$?

(a) $\left(-\frac{6}{5}, -\frac{4}{5}\right)$ (b) $\left(-\frac{6}{10}, -\frac{4}{10}\right)$ (c) $\left(\frac{4}{10}, \frac{6}{10}\right)$ (d) $\left(-\frac{6}{10}, \frac{4}{10}\right)$

5. What is the modulus form of $N(5, 0.02)$?

(a) $|x + 5| < 0.02$ (b) $|x - 0.02| < 5$ (c) $|x - 5| > 0.02$ (d) $|x - 5| < 0.02$

6. If modulus form of $N(a, 0.07)$ is $|x - 10| < k$ then what will be the value of k?

(a) a (b) 0.7 (c) 0.07 (d) 9.93

7. What is the value of $\lim_{x \to 3} 3x - 1$?

(a) 9 (b) 10 (c) $\frac{4}{3}$ (d) 8

8. What is the value of $\lim_{x \to 4} \sqrt{4x+9}$?

(a) 5 (b) 25 (c) $\frac{7}{4}$ (d) 7

9. What is the value of $\lim_{x \to -2} 10$?

(a) 10 (b) –2 (c) 8 (d) Indeterminate

10. What is the value of $\lim_{x \to 3} \frac{x^4 - 81}{x - 3}$?

(a) 192 (b) 324 (c) 36 (d) 108

11. What is the value of $\lim_{x \to -1} \frac{x^5 + 1}{x + 1}$?

(a) –5 (b) 5 (c) 4 (d) –4

12. If $y = 10 - 3x$ and $x \to -3$ then y tends to which value ?

(a) 1 (b) 9 (c) 19 (d) 7

Section B

Answer the following questions in one sentence :

1. Express 0.09 neighbourhood of 0 in interval form.

2. Express 0.001 neighbourhood of –5 in modulus form.

3. Express $|x - 10| < \frac{1}{10}$ in neighbourhood form.

4. Express $|2x| < \frac{1}{2}$ in interval form.

5. Express $N(50,\ 0.8)$ in modulus form.

6. If $N(a,\ 0.2) = |x - 7| < b$ then find the value of a.

7. If $|x + 4| < 0.04 = (k,\ -3.96)$ then find the value of k.

8. Find the value of $\lim_{x \to 5} (3x + 5)$.

9. Find the value of $\lim_{x \to -3} \sqrt[3]{2 - 2x}$.

10. Find the value of $\lim_{x \to 0} \left(\frac{3x^2 - 4x + 10}{2x + 5} \right)$.

11. Find the value of $\lim_{x \to 2} \frac{x^5 - 32}{x - 2}$.

12. Find the value of $\lim_{x \to -a} \frac{x^m + a^m}{x + a}$ where m is an odd number.

13. If $\lim_{x \to -1} 4x + k = 6$ then find the value of k.

14. If $\lim_{x \to 3} \frac{2}{3x + k} = \frac{1}{7}$ then find the value of k.

Section C

Answer the following questions :

1. Define an open interval.

2. Define the δ neighbourhood of a.

3. Define the punctured δ neighbourhood of a.

4. Express the interval form $(-0.5,\ 0.5)$ in modulus form.

5. Express the interval form $(-8.75,\ -7.25)$ in neighbourhood form.

6. If $N(k_1,\ 0.5) = (19.5,\ k_2)$ then find the value of k_1 and k_2.

7. Express $|3x + 1| < 2$ in neighbourhood and interval form.

8. If $|x - A_1| < 0.09 = (A_2,\ 4.09)$ then find the value of A_1 and A_2.

9. Explain the meaning of $x \to a$.

10. Explain the meaning of $x \to 0$.

11. Define limit of a function.

12. State multiplication working rule of limit.

13. State division working rule of limit.

14. State the standard form of limit of a polynomial.

Section D

Find the values of the following :

1. $\lim_{x \to 1} \frac{3x^2 - 4x + 1}{x - 1}$

2. $\lim_{x \to 3} \frac{x - 3}{2x^2 - 3x - 9}$

3. $\lim_{x \to -1} \frac{3x^2 - 2x - 5}{x + 1}$

4. $\lim_{x \to 1} \frac{x^2 + 2x - 3}{x^2 - 1}$

5. $\lim_{x \to \frac{1}{2}} \frac{2x^2 + 5x - 3}{4x^2 - 1}$

6. $\lim_{x \to -3} \frac{2x^2 + 9x + 9}{2x^2 + 7x + 3}$

7. $\lim_{x \to -\frac{1}{2}} \frac{2x^2 + 3x + 1}{2x^2 - x - 1}$

8. $\lim_{x \to -2} \frac{9x^2 + 5x - 26}{5x^2 + 17x + 14}$

9. $\lim_{x \to 0} \frac{1}{x}\left[\frac{5x + 14}{3x + 7} - 2\right]$

10. $\lim_{x \to 2} \left[\frac{2}{x - 2} - \frac{4}{x^2 - 2x}\right]$

11. $\lim_{x \to 0} 1 + \frac{2}{3 + \frac{4}{x}}$

12. $\lim_{x \to -p} \frac{x^4 - p^4}{x^3 + p^3}$

13. $\lim_{x \to 3} \frac{x^6 - 729}{x^4 - 81}$

14. $\lim_{x \to -2} \frac{x^{10} - 1024}{x^5 + 32}$

15. $\lim_{x \to -1} \frac{x^{2017} + 1}{x^{2018} - 1}$

16. $\lim_{x \to 1} \frac{x^{\frac{7}{2}} - 1}{x^{\frac{3}{2}} - 1}$

17. $\lim_{x \to 1} \frac{\sqrt[3]{x} - 1}{\sqrt{x} - 1}$

Section E

I. Answer the following :

1. If $y = 5x + 7$ then using tabular method, prove that when $x \to 2$, $y \to 17$.

2. If $y = \frac{3x^2 + 16x + 16}{x + 4}$ then using tabular method, prove that when $x \to -4$, $y \to -8$.

3. Using tabular method, prove that $\lim_{x \to -1} \frac{3}{x + 1}$ does not exist.

II. Find the values of the following using tabular method :

1. $\lim_{x \to 5} \frac{x^2 - 3x - 10}{x - 5}$

2. $\lim_{x \to 1} \frac{2x^2 + 3x - 5}{x - 1}$

3. $\lim_{x \to -1} \frac{4x^2 + 5x + 1}{x + 1}$

4. $\lim_{x \to 0} 3x - 1$

III. Find the values of the following :

1. $\lim_{h \to 0} \frac{(x + h)^7 - x^7}{h}$

2. $\lim_{x \to 0} \frac{\sqrt[10]{1 + x} - 1}{x}$

3. $\lim_{x \to 0} \frac{(1 + x)^n - 1}{x}$

4. $\lim_{x \to \frac{1}{2}} \frac{f(x) - f\left(\frac{1}{2}\right)}{2x - 1}$ where $f(x) = x^2 + x - 1$

5. $\lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$ where $f(x) = x^3$

6. $\lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$ where $f(x) = x^7$

7. $\lim_{x \to 2} \frac{f(x) - f(2)}{x - 2}$ where $f(x) = \sqrt{x + 7}$

8. $\lim_{h \to 0} \frac{f(2 + h) - f(2)}{h}$ where $f(x) = 2x^2 + 3$

9. $\lim_{x \to 0} \frac{f(2 + x) - f(2 - x)}{2x}$ where $f(x) = x^2$

10. $\lim_{x \to 2} \frac{f(x) - f(2)}{x - 2}$ where $f(x) = x^2 + x$

Srinivasa Ramanujan
(1887 - 1920)

Srinivasa Ramanujan was one of the greatest mathematical geniuses of India. He made substantial contributions to the analytical theory of numbers and worked on elliptic functions, continued fractions and infinite series. Ramanujan independently discovered results of Gauss, Kummer and others on hyper geometric series. Ramanujan initially developed his own mathematical research in isolation; it was quickly recognized by Indian mathematicians. When his skills became obvious and known to the wider mathematical community, centered in Europe at the time, he began a famous partnership with the English mathematician G. H. Hardy, who realized that Ramanujan had rediscovered previously known theorems in addition to producing new ones. On 18th February 1918, Ramanujan was elected as fellow of the Cambridge Philosophical Society. On the 125th anniversary of his birth, India declared the birthday of Ramanujan, December 22nd as 'National Mathematics Day and also declared that the year 2012 would be celebrated as the National Year of Mathematics.

●

5

Differentiation

Contents :

5.1 Introduction

We have studied about functions in Standard 11. Let $f(x)$ be a function of x. Differentiation is a technique which is used for analyzing the way in which function $f(x)$ changes and how much does it change with a change in the value of x. That is, we can know how rapidly a function is changing at any point using differentiation. In real life, we have functions like production cost, revenue, profit, etc. and it is often important to know how rapidly these functions change with respect to change in produced units or sold units x.

Consider $y = f(x) = 2x^2 + 3$, a function of x. If the value of independent variable (x) is changed there will be a corresponding change in the dependent variable (y). If the value of x is 2 then the value of dependent variable y will be 11. Now we shall find the increase in y for a small increase in x. For a small increase in value of x, i.e. if we take values of x as 2.1, 2.01, 2.001, 2.0001, ... etc then we get corresponding value of y as 11.82, 11.082, 11.008, 11.0008, ..., etc. We denote the increase in x by δ_x and increase in y by δ_y. The ratio $\frac{\delta_y}{\delta_x}$ is termed as incrementary ratio. Let us observe this incrementary ratio for the above values of x and corresponding values of y.

x	δ_x	$y = f(x)$	δ_y	$\frac{\delta_y}{\delta_x}$
2.1	0.1	11.82	0.82	8.2
2.01	0.01	11.0802	0.0802	8.02
2.001	0.001	11.0080	0.0080	8.002
2.0001	0.0001	11.0008	0.0008	8.0002
. . .	. . .	. . .	. . .	. . .

We make the following observations from the above table :

(i) δ_y varies when δ_x varies

(ii) $\delta_y \to 0$ when $\delta_x \to 0$

(iii) The ratio $\frac{\delta_y}{\delta_x}$ tends to 8.

Hence, this example illustrates that $\delta_y \to 0$ when $\delta_x \to 0$ but $\frac{\delta_y}{\delta_x}$ tends to a finite value, not necessarily zero. The limit of $\frac{\delta_y}{\delta_x}$ is represented by $\frac{dy}{dx}$ and is called the derivative of y with respect to x.

In the above example $\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta_y}{\delta_x} = 8$

In many business problems, we are interested in the rate of change of a function and, in particular, the range of values of independent variable for which the rate of change of a function may be positive or negative.

Differentiation is used in production, replacement, pricing and other management decision problems.

In short, differentiation is used to determine the rate of change in the dependent variable (function of independent variable) with respect to the independent variable.

5.2 Definition : Differentiation and Derivative

Let us consider a function $y = f(x)$.

When we take $x = a$, the value of the function will be $f(a)$. Now, when the value of x changes from a to $a + h$, the value of the function will change from $f(a)$ to $f(a+h)$. So, for a change of $(a+h) - a = h$ in the value of x, there is a change of $f(a+h) - f(a)$ in the value of $f(x)$. If there is a change of h in value of a then the relative change in the function will be $\frac{f(a+h) - f(a)}{h}$. If h is made very small then the limit of this relative change is called derivative of $f(x)$ with respect to x at $x = a$ and it is denoted by $f'(a)$.

Definition : Let $f : A \to R$ and $a \in A$, where A is an open interval of R. If $\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$ exists, then this limit of a function f is called **derivative** at $x = a$. It is denoted by $f'(a)$.

The process of obtaining derivative of a function is called **differentiation**.

Thus, $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$.

For any value of x of the domain of f, we have $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$. $f'(x)$ is called a derivative of $f(x)$ with respect to x.

If y is a function of x then its derivative is denoted by $\frac{dy}{dx}$.

We shall now find derivatives of some functions using this definition of derivative.

Illustration 1 : Obtain the derivative of $f(x) = x$ with the help of definition.

Here, $f(x) = x$

$\therefore \quad f(x+h) = x + h$

Now, $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

$$= \lim_{h \to 0} \frac{(x+h) - x}{h}$$

$$= \lim_{h \to 0} \frac{h}{h}$$

$$= 1 \qquad (\because \ h \neq 0)$$

Hence, if $f(x) = x$ then $f'(x) = 1$.

Illustration 2 : Obtain derivative of $f(x) = x^3$ with the help of definition.

Here, $f(x) = x^3$

$$\therefore \ f(x+h) = (x+h)^3$$

$$= x^3 + 3x^2h + 3xh^2 + h^3$$

Now, $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

$$= \lim_{h \to 0} \frac{\left(x^3 + 3x^2h + 3xh^2 + h^3\right) - x^3}{h}$$

$$= \lim_{h \to 0} \frac{3x^2h + 3xh^2 + h^3}{h}$$

$$= \lim_{h \to 0} \frac{h\left(3x^2 + 3xh + h^2\right)}{h}$$

$$= \lim_{h \to 0} 3x^2 + 3xh + h^2 \qquad (\because \ h \neq 0)$$

$$= 3x^2 + 3x(0) + (0)^2$$

$$= 3x^2$$

Hence, if $f(x) = x^3$ then $f'(x) = 3x^2$

Illustration 3 : Obtain derivative of $f(x) = x^n$ with the help of definition.

Here, $f(x) = x^n$

$$\therefore \ f(x+h) = (x+h)^n$$

Now, $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

$$= \lim_{h \to 0} \frac{(x+h)^n - x^n}{h}$$

(Taking $x + h = t$, when $h \to 0$ then $t \to x$)

$$= \lim_{t \to x} \frac{t^n - x^n}{t - x} \qquad (\because \ x + h = t)$$

$$= nx^{n-1} \quad \left(\because \lim_{x \to a} \frac{x^n - a^n}{x - a} = na^{n-1}\right)$$

Hence, if $f(x) = x^n$ then $f'(x) = nx^{n-1}$

Illustration 4 : Obtain derivative of $f(x) = \sqrt{x}$ with the help of definition.

Here, $f(x) = \sqrt{x}$

$\therefore \ f(x + h) = \sqrt{x + h}$

Now, $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

$$= \lim_{h \to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h}$$

(Multiplying numerator and denominator by $\sqrt{x+h} + \sqrt{x}$)

$$= \lim_{h \to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \times \frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}}$$

$$= \lim_{h \to 0} \frac{\left(\sqrt{x+h}\right)^2 - \left(\sqrt{x}\right)^2}{h\left(\sqrt{x+h} + \sqrt{x}\right)}$$

$$= \lim_{h \to 0} \frac{x + h - x}{h\left(\sqrt{x+h} + \sqrt{x}\right)}$$

$$= \lim_{h \to 0} \frac{h}{h\left(\sqrt{x+h} + \sqrt{x}\right)}$$

$$= \lim_{h \to 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} \qquad (\because \ h \neq 0)$$

$$= \frac{1}{\sqrt{x+0} + \sqrt{x}}$$

$$= \frac{1}{2\sqrt{x}}$$

Hence, if $f(x) = \sqrt{x}$ then $f'(x) = \frac{1}{2\sqrt{x}}$

Illustration 5 : Obtain derivative of $f(x) = \frac{1}{x}$ with the help of definition.

Here, $f(x) = \frac{1}{x}$

$$\therefore \ f(x+h) = \frac{1}{x+h}$$

Now, $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

$$= \lim_{h \to 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h}$$

$$= \lim_{h \to 0} \frac{x - (x+h)}{hx(x+h)}$$

$$= \lim_{h \to 0} \frac{x - x - h}{hx(x+h)}$$

$$= \lim_{h \to 0} \frac{-h}{hx(x+h)}$$

$$= \lim_{h \to 0} \frac{-1}{x(x+h)} \qquad (\because \ h \neq 0)$$

$$= \frac{-1}{x(x+0)}$$

$$= \frac{-1}{x^2}$$

Hence, if $f(x) = \frac{1}{x}$ then $f'(x) = \frac{-1}{x^2}$

Illustration 6 : Obtain derivative of $f(x) = k$ (k is constant) with the help of definition.

Here, $f(x) = k$

$$\therefore \ f(x+h) = k$$

Now, $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

$$= \lim_{h \to 0} \frac{k - k}{h}$$

$$= \lim_{h \to 0} \frac{0}{h}$$

$$= 0$$

Hence, if $f(x) = k$ then $f'(x) = 0$

EXERCISE 5.1

Obtain the derivatives of the following functions with the help of definition :

1. $f(x) = 2x + 3$

2. $f(x) = x^2$

3. $f(x) = x^7$

4. $f(x) = \frac{1}{x+1}, \quad x \neq -1$

5. $f(x) = \sqrt[3]{x}$

6. $f(x) = \frac{2}{3x-4}, \quad x \neq \frac{4}{3}$

7. $f(x) = 10$

*

5.3 Some Standard Derivatives

We shall use derivatives of following functions.

1. If $y = x^n$ (where $n \in R$ and $x \in R^+$)

then $\frac{dy}{dx} = nx^{n-1}$

2. If $y = k$ (where k is constant)

then $\frac{dy}{dx} = 0$

5.4 Working Rules for Differentiation

We shall accept certain working rules for differentiation without proof.

If u and v are differentiable functions of x then,

Rule 1 : If $y = u \pm v$ then

$$\frac{dy}{dx} = \frac{du}{dx} \pm \frac{dv}{dx}$$

Rule 2 : If $y = u \cdot v$ then

$$\frac{dy}{dx} = u \cdot \frac{dv}{dx} + v \cdot \frac{du}{dx}$$

Rule 3 : If $y = \frac{u}{v}$, $v \neq 0$ then

$$\frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

Rule 4 : **(Chain Rule)**

If y is a differentiable function of u and u is a differentiable function of x then

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

We shall see some illustrations explaining the use of working rules for differentiation mentioned above.

Illustration 7 : Find $\frac{dy}{dx}$ for $y = x^4 - 3x^2 + 2x - 3$.

$$y = x^4 - 3x^2 + 2x - 3$$

$$\therefore \quad \frac{dy}{dx} = \frac{d}{dx}\left(x^4 - 3x^2 + 2x - 3\right)$$

$$= \frac{d}{dx}\left(x^4\right) - \frac{d}{dx}\left(3x^2\right) + \frac{d}{dx}\left(2x\right) - \frac{d}{dx}\left(3\right)$$

$$= \frac{d}{dx}\left(x^4\right) - 3\frac{d}{dx}\left(x^2\right) + 2\frac{d}{dx}\left(x\right) - \frac{d}{dx}\left(3\right)$$

$$= 4x^3 - 3(2x) + 2(1) - (0)$$

$$= 4x^3 - 6x + 2$$

Illustration 8 : Find $\frac{dy}{dx}$ for $y = x^3 + \sqrt{x} - \frac{4}{x} + \frac{1}{\sqrt[3]{x}} + \frac{1}{4}$.

$$y = x^3 + \sqrt{x} - \frac{4}{x} + \frac{1}{\sqrt[3]{x}} + \frac{1}{4}$$

$$= x^3 + x^{\frac{1}{2}} - 4x^{-1} + x^{-\frac{1}{3}} + \frac{1}{4}$$

$$\therefore \quad \frac{dy}{dx} = \frac{d}{dx}\left(x^3\right) + \frac{d}{dx}\left(x^{\frac{1}{2}}\right) - 4\frac{d}{dx}\left(x^{-1}\right) + \frac{d}{dx}\left(x^{-\frac{1}{3}}\right) + \frac{d}{dx}\left(\frac{1}{4}\right)$$

$$= 3x^2 + \frac{1}{2}x^{\frac{1}{2}-1} - 4\left(-1\,x^{-1-1}\right) + \left(\frac{-1}{3}\right)x^{\frac{-1}{3}-1} + 0$$

$$= 3x^2 + \frac{1}{2}x^{-\frac{1}{2}} + 4x^{-2} - \frac{1}{3}x^{-\frac{4}{3}}$$

$$= 3x^2 + \frac{1}{2x^{\frac{1}{2}}} + \frac{4}{x^2} - \frac{1}{3x^{\frac{4}{3}}}$$

Illustration 9 : If $y = \left(2x^2 + 3\right)\left(3x - 2\right)$ then find derivative of y with respect to x.

$$y = \left(2x^2 + 3\right)\left(3x - 2\right)$$

Take, $u = 2x^2 + 3$ and $v = 3x - 2$.

$$\therefore \quad \frac{du}{dx} = 4x \quad \text{and} \quad \frac{dv}{dx} = 3$$

Now, $y = u \cdot v$.

$$\therefore \quad \frac{dy}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}$$

$$= \left(2x^2 + 3\right)(3) + (3x - 2)(4x)$$

$$= 6x^2 + 9 + 12x^2 - 8x$$

$$= 18x^2 - 8x + 9$$

Note : Illustration 9 can also be solved using working rule 1 by simplifying y i.e. multiplying two terms of y.

Illustration 10 : Find $\frac{dy}{dx}$, $y = \frac{2x+3}{3x-2}$.

$$y = \frac{2x+3}{3x-2}$$

Take $u = 2x + 3$ and $v = 3x - 2$.

$$\therefore \quad \frac{du}{dx} = 2 \text{ and } \frac{dv}{dx} = 3$$

Now, $y = \frac{u}{v}$

$$\therefore \quad \frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

$$= \frac{(3x-2)(2) - (2x+3)(3)}{(3x-2)^2}$$

$$= \frac{(6x-4) - (6x+9)}{(3x-2)^2}$$

$$= \frac{6x - 4 - 6x - 9}{(3x-2)^2}$$

$$= \frac{-13}{(3x-2)^2}$$

Illustration 11 : If $y = \frac{3}{4x+5}$ then differentiate y with respect to x.

$$y = \frac{3}{4x+5}$$

Take $u = 3$ and $v = 4x + 5$.

$$\therefore \quad \frac{du}{dx} = 0 \text{ and } \frac{dv}{dx} = 4$$

Now, $y = \frac{u}{v}$

$$\therefore \quad \frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

$$= \frac{(4x+5)(0) - 3(4)}{(4x+5)^2}$$

$$= \frac{0 - 12}{(4x+5)^2}$$

$$= \frac{-12}{(4x+5)^2}$$

Illustration 12 : If $y = \frac{2x^2 + 3x + 4}{x^2 + 5}$ then find $\frac{dy}{dx}$.

$$y = \frac{2x^2 + 3x + 4}{x^2 + 5}$$

Take $u = 2x^2 + 3x + 4$ and $v = x^2 + 5$.

$$\therefore \quad \frac{du}{dx} = 4x + 3 \text{ and } \frac{dv}{dx} = 2x$$

Now, $y = \frac{u}{v}$.

$$\therefore \quad \frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

$$= \frac{\left(x^2 + 5\right)(4x + 3) - \left(2x^2 + 3x + 4\right)(2x)}{\left(x^2 + 5\right)^2}$$

$$= \frac{\left(4x^3 + 20x + 3x^2 + 15\right) - \left(4x^3 + 6x^2 + 8x\right)}{\left(x^2 + 5\right)^2}$$

$$= \frac{4x^3 + 20x + 3x^2 + 15 - 4x^3 - 6x^2 - 8x}{\left(x^2 + 5\right)^2}$$

$$= \frac{-3x^2 + 12x + 15}{\left(x^2 + 5\right)^2}$$

Illustration 13 : Differentiate $y = (3x + 7)^8$ with respect to x.

$$y = (3x + 7)^8$$

Taking $u = 3x + 7$, $y = u^8$

$$\therefore \quad \frac{du}{dx} = 3 \text{ and } \frac{dy}{du} = 8u^7$$

Now, $\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$

$$= \left(8u^7\right)(3)$$

$$= 24u^7$$

Putting value of u,

$$\frac{dy}{dx} = 24(3x + 7)^7$$

Illustration 14 : Find $\frac{dy}{dx}$, $y = \sqrt{x^2 + 3}$.

$$y = \sqrt{x^2 + 3}$$

Taking $u = x^2 + 3$, $y = \sqrt{u}$

$$\therefore \quad \frac{du}{dx} = 2x \text{ and } \frac{dy}{du} = \frac{1}{2\sqrt{u}}.$$

Now, $\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$

$$= \left(\frac{1}{2\sqrt{u}}\right)(2x)$$

$$= \frac{x}{\sqrt{u}}$$

Putting value of u,

$$\frac{dy}{dx} = \frac{x}{\sqrt{x^2 + 3}}$$

Illustration 15 : Obtain derivative of $y = 1 + \frac{2}{3 + \frac{4}{x}}$ with respect to x.

$$y = 1 + \frac{2}{3 + \frac{4}{x}}$$

$$= 1 + \frac{2x}{3x + 4}$$

$$= \frac{(3x + 4) + 2x}{3x + 4}$$

$$\therefore \quad y = \frac{5x + 4}{3x + 4}$$

Here, take $u = 5x + 4$ and $v = 3x + 4$

$$\therefore \quad \frac{du}{dx} = 5 \text{ and } \frac{dv}{dx} = 3$$

Now, $y = \frac{u}{v}$

$$\therefore \quad \frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

$$= \frac{(3x + 4)(5) - (5x + 4)(3)}{(3x + 4)^2}$$

$$= \frac{(15x + 20) - (15x + 12)}{(3x + 4)^2}$$

$$= \frac{15x + 20 - 15x - 12}{(3x+4)^2}$$

$$= \frac{8}{(3x+4)^2}$$

Illustration 16 : If $2xy + 3x + y - 4 = 0$ then find $\frac{dy}{dx}$.

$$2xy + 3x + y - 4 = 0$$

$$\therefore \quad 2xy + y = 4 - 3x$$

$$\therefore \quad y(2x+1) = 4 - 3x$$

$$\therefore \quad y = \frac{4 - 3x}{2x + 1}$$

Here, take $u = 4 - 3x$ and $v = 2x + 1$.

$$\therefore \quad \frac{du}{dx} = -3 \text{ and } \frac{dv}{dx} = 2$$

Now, $y = \frac{u}{v}$

$$\therefore \quad \frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

$$= \frac{(2x+1)(-3) - (4-3x)(2)}{(2x+1)^2}$$

$$= \frac{-6x - 3 - 8 + 6x}{(2x+1)^2}$$

$$= \frac{-11}{(2x+1)^2}$$

Illustration 17 : If $y = 2 + 3x + 4x^2 + \frac{5}{6-7x}$ then find $\frac{dy}{dx}$.

$$y = 2 + 3x + 4x^2 + \frac{5}{6-7x}$$

$$\therefore \quad \frac{dy}{dx} = \frac{d}{dx}\left[2 + 3x + 4x^2 + \frac{5}{6-7x}\right]$$

$$= 0 + 3(1) + 4(2x) + \frac{d}{dx}\left(\frac{5}{6-7x}\right)$$

$$= 3 + 8x + \frac{(6-7x)(0) - 5(-7)}{(6-7x)^2} \qquad [\because \text{ Division rule}]$$

$$= 3 + 8x + \frac{35}{(6-7x)^2}$$

Illustration 18 : If $y = \left(x + \frac{6}{x+5}\right)\left(\frac{3x+2}{x^2+5x+6}\right)$ **then find** $\frac{dy}{dx}$**.**

$$y = \left(x + \frac{6}{x+5}\right)\left(\frac{3x+2}{x^2+5x+6}\right)$$

$$= \left[\frac{x(x+5)+6}{x+5}\right]\left(\frac{3x+2}{x^2+5x+6}\right)$$

$$= \left(\frac{x^2+5x+6}{x+5}\right)\left(\frac{3x+2}{x^2+5x+6}\right)$$

$$= \frac{3x+2}{x+5}$$

Here, take $u = 3x + 2$ and $v = x + 5$.

$$\therefore \quad \frac{du}{dx} = 3 \text{ and } \frac{dv}{dx} = 1$$

Now, $y = \frac{u}{v}$

$$\therefore \quad \frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$$

$$= \frac{(x+5)(3) - (3x+2)(1)}{(x+5)^2}$$

$$= \frac{(3x+15) - (3x+2)}{(x+5)^2}$$

$$= \frac{3x+15-3x-2}{(x+5)^2}$$

$$= \frac{13}{(x+5)^2}$$

Illustration 19 : If $f(x) = 3x^2 + 2x + 1$ **then find** $f'(x)$ **and hence obtain** $f'(1)$**.**

Here, $f(x) = 3x^2 + 2x + 1$

$$\therefore f'(x) = 6x + 2$$

$$\therefore f'(1) = 6(1) + 2$$

$$= 8$$

Illustration 20 : If $f(x) = x^2 - x + 3$ then for which value of x, $f'(x) = 0$?

Here, $f(x) = x^2 - x + 3$

$\therefore \; f'(x) = 2x - 1 + 0$

Now, $f'(x) = 0$ is given

$\therefore \; 2x - 1 = 0$

$\therefore \; 2x = 1$

$\therefore \; x = \frac{1}{2}$

5.5 Second Order Differentiation

As seen in many of the previous illustrations, the derivative of a function of x is generally also a function of x. The derivative of $y = f(x)$ is denoted by $\frac{dy}{dx}$ or $f'(x)$. This derivative is called the first order derivative of the function. The second order derivative of the function means the derivative of the first order derivative of the function. It is denoted by $\frac{d^2y}{dx^2}$ or $f''(x)$. Second order derivative along with the first order derivative can be useful in maximization or minimization of a function. This can be applied to minimize cost function, maximize revenue function and maximize profit function.

We shall now see the method of obtaining second order derivative with few illustrations.

Illustration 21 : Obtain $\frac{dy}{dx}$ for $y = 3x^4 - 2x^3 + x^2 - 8x + 7$. Also obtain its value at $x = 1$.

$y = 3x^4 - 2x^3 + x^2 - 8x + 7$

$$\therefore \; \frac{dy}{dx} = 12x^3 - 6x^2 + 2x - 8$$

$$\therefore \; \frac{d^2y}{dx^2} = \frac{d}{dx}\left[\frac{dy}{dx}\right]$$

$$= \frac{d}{dx}\left[12x^3 - 6x^2 + 2x - 8\right]$$

$$= 36x^2 - 12x + 2$$

Putting $x = 1$,

$$\frac{d^2y}{dx^2} = 36(1)^2 - 12(1) + 2$$

$$= 36 - 12 + 2$$

$$= 26$$

Illustration 22 : If $f(x) = 4x^3 + 2x^2 + 7x + 9$ then for which value of x, $f''(x) = 52$?

$$f(x) = 4x^3 + 2x^2 + 7x + 9$$

$$\therefore\ f'(x) = 12x^2 + 4x + 7$$

$$\therefore\ f''(x) = 24x + 4$$

Now, $f''(x) = 52$

$$\therefore\ 24x + 4 = 52$$

$$\therefore\ 24x = 48$$

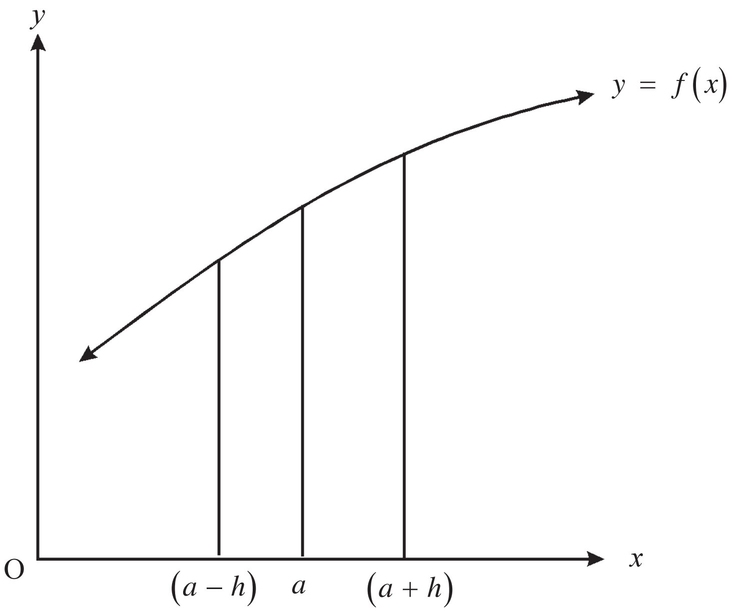
$$\therefore\ x = 2$$

5.6 Increasing Function and Decreasing Function

Increasing function

In the adjacent figure, the curve of the function $y = f(x)$ is drawn. The value of the function at $x = a$ is $y = f(a)$. If h is a very small positive number and if $f(a+h) > f(a)$ and also $f(a) > f(a-h)$ then $f(x)$ is said to be an increasing function at $x = a$.

If the function is increasing at $x = a$ then $f'(a) > 0$



Decreasing function

In the adjacent figure, the curve of the function $y = f(x)$ is drawn. The value of the function at $x = a$ is $y = f(a)$. If h is a very small positive number and if $f(a+h) < f(a)$ and also $f(a) < f(a-h)$, then $f(x)$ is said to be a decreasing function of $x = a$.

If the function is decreasing at $x = a$ then $f'(a) < 0$

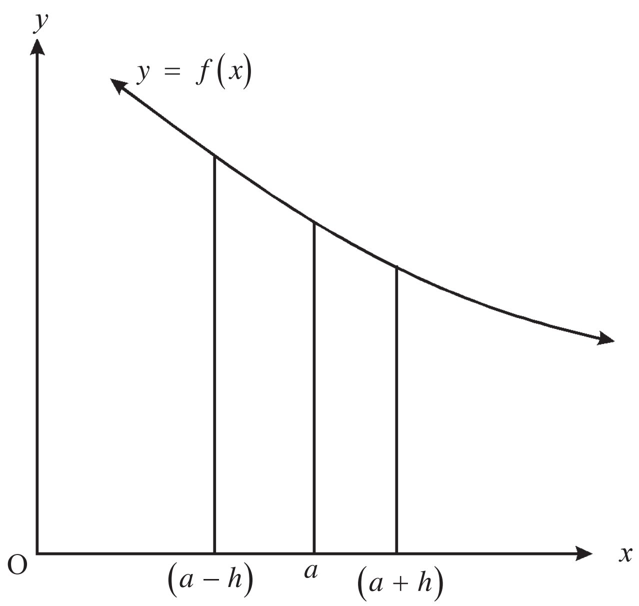


Illustration 23 : If $f(x) = x^2 - 4x$ then decide whether the function is increasing or decreasing at $x = -1$, $x = 0$ and $x = 3$.

$$f(x) = x^2 - 4x$$

$$\therefore \quad f'(x) = 2x - 4$$

At $x = -1$

$$f'(-1) = 2(-1) - 4$$

$$= -6 < 0$$

$\therefore$ Function is decreasing at $x = -1$.

At $x = 0$

$$f'(0) = 2(0) - 4$$

$$= -4 < 0$$

$\therefore$ Function is decreasing at $x = 0$.

At $x = 3$

$$f'(3) = 2(3) - 4$$

$$= 2 > 0$$

$\therefore$ Function is increasing at $x = 3$.

Illustration 24 : Decide whether the function $y = x^3 - 3x^2 + 7$ is increasing or decreasing at $x = 1$ and $x = 3$.

$$y = x^3 - 3x^2 + 7$$

$$\therefore \quad \frac{dy}{dx} = 3x^2 - 6x$$

At $x = 1$

$$\frac{dy}{dx} = 3(1)^2 - 6(1)$$

$$= 3 - 6$$

$$= -3 < 0$$

$\therefore$ Function is decreasing at $x = 1$.

At $x = 3$

$$\frac{dy}{dx} = 3(3)^2 - 6(3)$$

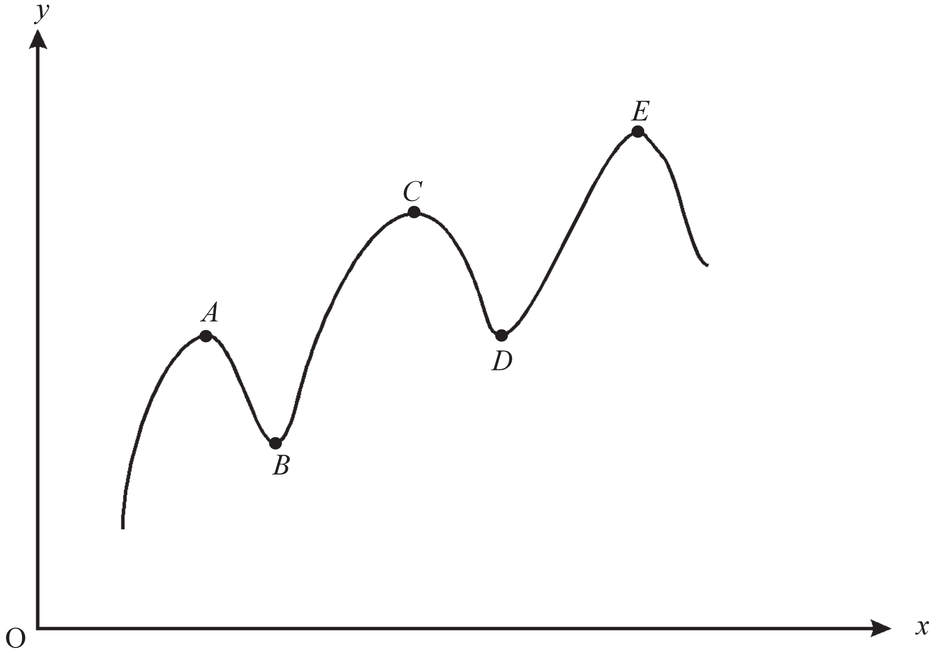
$$= 27 - 18$$

$$= 9 > 0$$

$\therefore$ Function is increasing at $x = 3$.

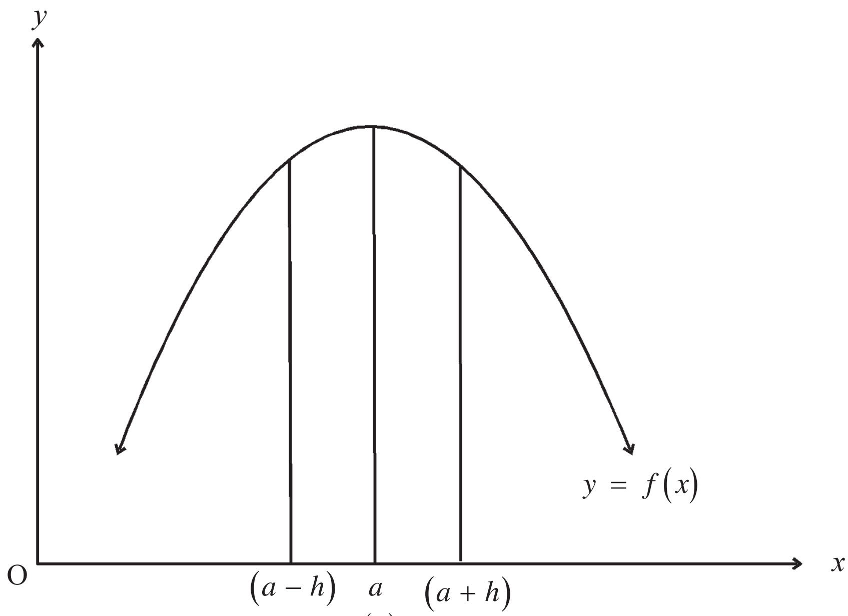
5.7 Maximum and Minimum Values of a Function

We discussed about increasing and decreasing function. Now, we shall study the method of obtaining maximum and minimum value of a function. Suppose the graph of a function $y = f(x)$ is obtained as follows.



It can be seen that the curve obtains maximum values at points *A*, *C* and *E* while its values are minimum at points *B* and *D*. Thus, the function may have more than one maximum or minimum values.

Maximum Value :

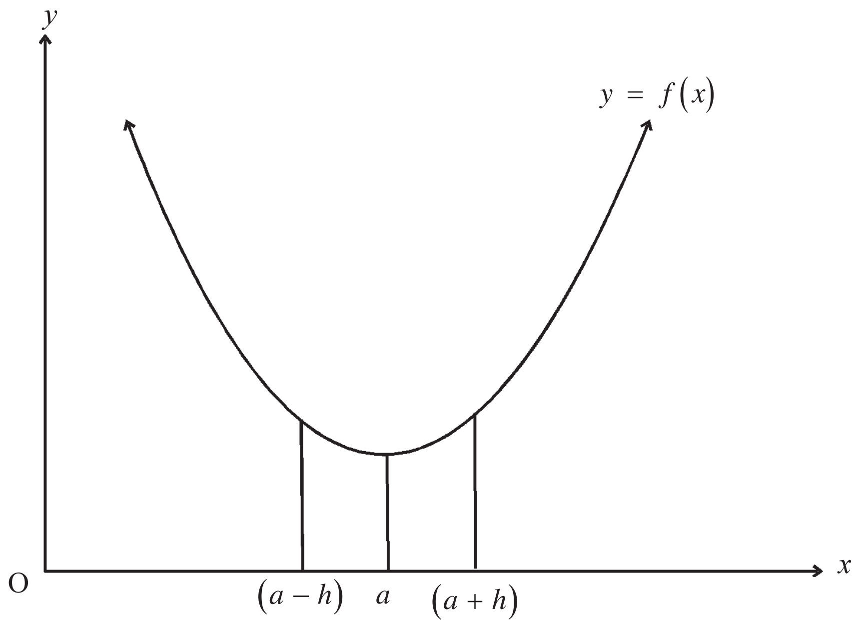


In the figure, curve of the function $y = f(x)$ is drawn. The value of the function at $x = a$ is $y = f(a)$. If h is a small positive number and if $f(a) > f(a+h)$ and also $f(a) > f(a-h)$ then $f(x)$ is said to be maximum at $x = a$.

The necessary and sufficient conditions for a function to be maximum at $x = a$ are as follows :

(i) $f'(a) = 0$ (ii) $f''(a) < 0$

Minimum Value :



In the figure, curve of the function $y = f(x)$ is drawn. The value of the function at $x = a$ is $y = f(a)$. If h is a small positive number and if $f(a) < f(a+h)$ and $f(a) < f(a-h)$ then $f(x)$ is said to be minimum at $x = a$.

The necessary and sufficient conditions for a function to be minimum at $x = a$ are as follows :

(i) $f'(a) = 0$ (ii) $f''(a) > 0$

The maximum and minimum values of a function are known as stationary maximum and stationary minimum values of function.

Maximum or minimum values do not mean the largest or the smallest value of a function. The function is maximum of $x = a$ only means that the value of the function is maximum in a small interval around $x = a$. Similarly, the function is minimum at $x = b$ only means that the value of the function is minimum in a small interval around $x = b$. The points where maximum or minimum values occur are known as **stationary points**. The necessary condition to obtain a stationary value is $\frac{dy}{dx} = 0$.

Method of obtaining maximum and minimum values of a function :

- Find the first derivative $\frac{dy}{dx} = f'(x)$ of the given function.
- Putting $\frac{dy}{dx} = 0$, solve the equation and obtain the values of x. These values of x give the stationary points.
- Find the second order derivative and put these values of x alternatively in the second derivative.
- The value of x at the stationary points for which the second order derivative is negative gives the maximum value of the function while the value of x at the stationary points for which the second order derivative is positive gives the minimum value of the function.
- The maximum and minimum values of a function are obtained by putting these values of x in the given function.

We shall now see the method of obtaining the maximum and minimum values of a function with a few illustrations.

Illustration 25 : Find the maximum and minimum values of $f(x) = 2x^3 + 3x^2 - 12x - 4$.

Here, $f(x) = 2x^3 + 3x^2 - 12x - 4$

$\therefore \quad f'(x) = 6x^2 + 6x - 12$

For stationary values, $f'(x) = 0$

$\therefore \quad 6x^2 + 6x - 12 = 0$

$\therefore \quad x^2 + x - 2 = 0$

$\therefore \quad (x+2)(x-1) = 0$

$\therefore \quad x = -2 \quad \text{or} \quad x = 1$

Now, $f''(x) = 12x + 6$

At $x = -2$

$f''(-2) = 12(-2) + 6$

$= -18 < 0$

$\therefore$ We get the maximum value of the function at $x = -2$.

At $x = 1$

$f''(1) = 12(1) + 6$

$= 18 > 0$

$\therefore$ We get the minimum value of the function at $x = 1$.

Minimum value of $f(x)$

Putting $x = 1$ in the function $f(x)$,

$f(1) = 2(1)^3 + 3(1)^2 - 12(1) - 4$

$= 2 + 3 - 12 - 4$

$= -11$

Maximum value of $f(x)$

Putting $x = -2$ in the function $f(x)$,

$f(-2) = 2(-2)^3 + 3(-2)^2 - 12(-2) - 4$

$= -16 + 12 + 24 - 4$

$= 16$

Thus, the maximum value of $f(x)$ is 16 and the minimum value is -11.

Illustration 26 : Find the maximum and minimum values of $y = x^3 - 2x^2 - 4x - 1$.

Here, $y = x^3 - 2x^2 - 4x - 1$

$$\therefore \quad \frac{dy}{dx} = 3x^2 - 4x - 4$$

For stationary values, $\frac{dy}{dx} = 0$

$$\therefore \quad 3x^2 - 4x - 4 = 0$$

$$\therefore \quad 3x^2 - 6x + 2x - 4 = 0$$

$$\therefore \quad 3x(x-2) + 2(x-2) = 0$$

$$\therefore \quad (x-2)(3x+2) = 0$$

$$\therefore \quad x = 2 \text{ or } x = -\frac{2}{3}$$

Now, $\frac{d^2y}{dx^2} = 6x - 4$

At $x = 2$

$$\frac{d^2y}{dx^2} = 6(2) - 4$$

$$= 8 > 0$$

$\therefore$ Function is minimum at $x = 2$.

At $x = -\frac{2}{3}$

$$\frac{d^2y}{dx^2} = 6\left(\frac{-2}{3}\right) - 4$$

$$= -4 - 4$$

$$= -8 < 0$$

$\therefore$ Function is maximum at $x = -\frac{2}{3}$.

Minimum value of function y

Putting $x = 2$ in the function y,

$$y = (2)^3 - 2(2)^2 - 4(2) - 1$$

$$= 8 - 8 - 8 - 1$$

$$= -9$$

Maximum value of function y

Putting $x = -\frac{2}{3}$ in the function y,

$$y = \left(-\frac{2}{3}\right)^3 - 2\left(-\frac{2}{3}\right)^2 - 4\left(-\frac{2}{3}\right) - 1$$

$$= \frac{-8}{27} - \frac{8}{9} + \frac{8}{3} - 1$$

$$= \frac{13}{27}$$

Thus, the maximum value of y is $\frac{13}{27}$ and the minimum value is –9.

5.8 Marginal Income and Marginal Cost

The differentiation is used to obtain solutions of economic and business problems. We have seen that the first and second order derivatives can be used to obtain the maximum and minimum values of a function.

First order derivative of a function can also be used to obtain marginal income and marginal cost.

In study of economics, the relation between price and demand of a commodity can be represented as a function. If we denote the price of a commodity by p and its demand by x then, we get the relation $x = f(p)$, which is called the **demand function**. If the income or revenue obtained by selling x units of a commodity is denoted by R then,

$$R = xp$$

Thus, revenue R is a function of demand x.

The change in revenue due to small change in demand is called **marginal revenue**.

Marginal revenue can be obtained by taking the derivative of revenue function with respect to x. Thus, when the demand is x then

$$\text{Marginal revenue} = \frac{dR}{dx}$$

If we denote the cost of producing x units by C then C can also be represented as function of x.

The change in cost due to small change in production is called **marginal cost**.

Marginal cost can be obtained by taking the derivative of cost function with respect to x. Thus, when the production is x then

$$\text{Marginal cost} = \frac{dC}{dx}$$

Illustration 27 : If the demand function of pizza is $p = 150 - 4x$ then find the marginal revenue when demand is of 3 pizzas and interpret it.

Here, demand function $p = 150 - 4x$

Now, revenue function $R = p \cdot x$

$$= (150 - 4x)\, x$$

$$\therefore \quad R = 150x - 4x^2$$

Marginal revenue $\frac{dR}{dx} = 150 - 8x$

When demand of pizza is $x = 3$ then

Marginal revenue $\frac{dR}{dx} = 150 - 8(3)$

$$= 126$$

Interpretation : Revenue for selling the 4th pizza is approximately ₹ 126.

Illustration 28 : If the demand function of a commodity is $x = \frac{50 - p}{2}$ then find the marginal revenue when price is ₹ 30.

Demand function $x = \frac{50 - p}{2}$

$\therefore \quad 2x = 50 - p$

$\therefore \quad p = 50 - 2x$

Now, revenue function $R = p \cdot x$

$= (50 - 2x)x$

$\therefore \quad R = 50x - 2x^2$

Marginal revenue $\frac{dR}{dx} = 50 - 4x$

When price $p = 30$ then

$x = \frac{50 - 30}{2}$

$\therefore \quad x = 10$

When demand $x = 10$ then

Marginal Revenue $= \frac{dR}{dx} = 50 - 4(10)$

$= 10$

Interpretation : Revenue for selling the 11th unit is approximately ₹ 10.

Illustration 29 : The cost function of a commodity is $C = 5x^2 + 6x + 2000$, where x is the number of units produced. Find marginal cost when production is 50 units.

Cost function $C = 5x^2 + 6x + 2000$

$\therefore$ Marginal Cost $\frac{dC}{dx} = 10x + 6$

When $x = 50$ then

Marginal Cost $\frac{dC}{dx} = 10(50) + 6$

$= 506$

Interpretation : The cost of producing the 51st unit is approximately ₹ 506.

5.9 Elasticity of Demand

Generally, a change in price of a commodity results in change in its demand in opposite direction. When the price of a commodity increases, its demand decreases and when the price of a commodity decreases, its demand increases. But these changes are not equal for all the commodities. For example, a sudden increase in price of luxury commodities results in a major decrease in its demand. While increase in the price of necessary commodities does not result in a major decrease in its demand. The change in demand for a commodity due to change in its price can be studied using elasticity of demand.

Definition : The ratio of percentage change in the demand of a commodity due to percentage change in its price is called elasticity of demand.

i.e.

$$\text{Elasticity of demand} = -\frac{\text{Percentage change in demand}}{\text{Percentage change in price}}$$

The ratio is negative as the changes in price and demand of a commodity is in opposite direction. For convenience, the value of elasticity of demand is taken positive and hence the negative sign is taken in the formula. If we denote the demand as x and price as p and the demand function $x = f(p)$ is given then

$$\text{Elasticity of demand} = -\frac{p}{x} \cdot \frac{dx}{dp}.$$

Illustration 30 : The demand function of a commodity is $x = 50 - 4p$. Find elasticity of demand when price is $p = 5$ and interpret it.

Demand function $x = 50 - 4p$

$$\therefore \quad \frac{dx}{dp} = 0 - 4(1)$$

$$= -4$$

$$\text{Now, elasticity of demand} = -\frac{p}{x} \cdot \frac{dx}{dp}$$

$$= \frac{-p}{(50 - 4p)} \times (-4)$$

$$= \frac{4p}{50 - 4p}$$

When price $p = 5$ then

$$\text{Elasticity of demand} = \frac{4(5)}{50 - 4(5)}$$

$$= \frac{20}{50 - 20}$$

$$= \frac{20}{30}$$

$$= 0.67$$

Interpretation : When the price changes by 1 percent, demand changes by 0.67 percent (in opposite direction) when the price is 5.

Illustration 31 : The demand function of a commodity is $p = 12 - \sqrt{x}$. Find the elasticity of demand when the price is 9 units and interpret it.

Demand function $p = 12 - \sqrt{x}$

$$\therefore \quad \frac{dp}{dx} = 0 - \frac{1}{2\sqrt{x}}$$

$$= -\frac{1}{2\sqrt{x}}$$

$$\therefore \quad \frac{dx}{dp} = -2\sqrt{x} \qquad \left[\because \frac{dx}{dp} = \frac{1}{\frac{dp}{dx}}\right]$$

$$\text{Now, elasticity of demand} = -\frac{p}{x} \cdot \frac{dx}{dp}$$

$$= \frac{-\left(12 - \sqrt{x}\right)}{x} \times \left(-2\sqrt{x}\right)$$

$$= \frac{\left(12 - \sqrt{x}\right)\left(2\sqrt{x}\right)}{x}$$

When demand is 9 units then

$$\text{Elasticity of demand} = \frac{\left(12 - \sqrt{9}\right)\left(2\sqrt{9}\right)}{9}$$

$$= \frac{(12 - 3)(2 \times 3)}{9}$$

$$= \frac{9 \times 6}{9}$$

$$= 6$$

Interpretation : When price changes by 1 percent, demand changes by 6 percent (in opposite direction) when demand is 9 units.

5.10 Minimization of cost function and maximization of Revenue function and Profit function

In practice, problems of minimizing the production cost of an item, maximizing the revenue by selling produced items and maximizing profits are to be solved. We know that the production cost C or revenue R by selling produced items and profit P can be represented as functions of x. Using the derivatives, we can decide when it will be maximum or minimum.

The conditions for minimizing the production cost function C are

$$\frac{dC}{dx} = 0 \text{ and } \frac{d^2C}{dx^2} > 0.$$

Similarly, the conditions for maximizing the revenue function R are

$$\frac{dR}{dx} = 0 \text{ and } \frac{d^2R}{dx^2} < 0.$$

And conditions for maximizing the profit function P are

$$\frac{dP}{dx} = 0 \text{ and } \frac{d^2P}{dx^2} < 0.$$

We shall now see the method of obtaining minimum cost, maximum revenue and maximum profit with few illustrations.

Illustration 32 : The daily cost of production for x tons of a commodity is $10x^2 - 1000x + 50000$. How many units should be produced for the minimum cost ? Also find the minimum cost.

Production cost function $C = 10x^2 - 1000x + 50000$

$$\therefore \quad \frac{dC}{dx} = 20x - 1000$$

Putting $\frac{dC}{dx} = 0$,

$$20x - 1000 = 0$$

$$\therefore \quad 20x = 1000$$

$$\therefore \quad x = 50$$

Now $\frac{d^2C}{dx^2} = 20$

Here, putting $x = 50$ in $\frac{d^2C}{dx^2}$,

$$\frac{d^2C}{dx^2} = 20 > 0$$

$\therefore$ Production cost is minimum at $x = 50$.

To find minimum cost, put $x = 50$ in the production cost function,

$$\text{Minimum Cost} = 10(50)^2 - 1000(50) + 50000$$

$$= 10(2500) - 50000 + 50000$$

$$= 25000$$

Illustration 33 : A factory produces x units and its production capacity is 60,000 units per day. Its daily total production cost is $C = 250000 + 0.08x + \frac{200000000}{x}$. How many units should be produced for minimum production cost ?

Production cost function $C = 250000 + 0.08x + \frac{200000000}{x}$

$$\therefore \quad \frac{dC}{dx} = 0.08 - \frac{200000000}{x^2}$$

Putting $\frac{dC}{dx} = 0$

$$0.08 - \frac{200000000}{x^2} = 0$$

$$\therefore \quad 0.08 = \frac{200000000}{x^2}$$

$$\therefore \quad 0.08\ x^2 = 200000000$$

$$\therefore \quad x^2 = 2500000000$$

$$\therefore \quad x = 50000 \quad \text{or} \quad x = -50000$$

Production cannot be negative, so we will take $x = 50000$.

Now $\frac{d^2C}{dx^2} = \frac{400000000}{x^3}$

Here, putting $x = 50000$ in $\frac{d^2C}{dx^2}$,

$$\frac{d^2C}{dx^2} = \frac{400000000}{(50000)^3} > 0$$

$\therefore$ Production cost is minimum at $x = 50000$.

Thus, 50,000 units should be produced so that the production cost is minimum.

Illustration 34 : The demand function of a watch is $p = 6000 - 2x$. Find the demand which maximizes the revenue and also find the corresponding price.

Demand function $p = 6000 - 2x$

Now, revenue function $R = p \cdot x$

$$= (6000 - 2x)x$$

$$\therefore \quad R = 6000x - 2x^2$$

$$\therefore \quad \frac{dR}{dx} = 6000 - 4x$$

Putting $\frac{dR}{dx} = 0$,

$$6000 - 4x = 0$$

$$\therefore \quad 6000 = 4x$$

$$\therefore \quad x = 1500$$

Now $\frac{d^2R}{dx^2} = 0 - 4$

$= -4$

Here, putting $x = 1500$ in $\frac{d^2R}{dx^2}$,

$\frac{d^2R}{dx^2} = -4 < 0$

$\therefore$ Revenue is maximum at $x = 1500$.

Now we shall find the corresponding price.

Putting $x = 1500$ in demand function $p = 6000 - 2x$,

Price $p = 6000 - 2(1500)$

$= 6000 - 3000$

$p = 3000$

Illustration 35 : If the production cost function for a producer is $C = 100 + 0.015x^2$ and revenue function is $R = 3x$ then find the profit function. How many units should be produced by the producer for maximum profit ?

Production cost function $C = 100 + 0.015x^2$ and revenue function $R = 3x$

Now, profit function $P = R - C$

$= 3x - (100 + 0.015x^2)$

$\therefore \quad P = 3x - 100 - 0.015x^2$

$\therefore \quad \frac{dP}{dx} = 3 - 0.015(2x)$

$= 3 - 0.03x$

Putting $\frac{dP}{dx} = 0$

$3 - 0.03x = 0$

$\therefore \quad 3 = 0.03x$

$\therefore \quad x = \frac{3}{0.03}$

$x = 100$

Now $\frac{d^2P}{dx^2} = 0 - 0.03\ (1)$

$= -0.03$

Here putting $x = 100$ in $\frac{d^2P}{dx^2}$,

$\frac{d^2P}{dx^2} = -0.03 < 0$

$\therefore$ At $x = 100$, profit is maximum.

Summary

- Derivative $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$
- If $y = x^n$, $\frac{dy}{dx} = nx^{n-1}$
- If $y = k$ (constant), $\frac{dy}{dx} = 0$
- If u and v are differentiable functions of x then,

  (1) If $y = u \pm v$ then $\frac{dy}{dx} = \frac{du}{dx} \pm \frac{dv}{dx}$

  (2) If $y = u \cdot v$ then $\frac{dy}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}$

  (3) If $y = \frac{u}{v}$ then $\frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$

  (4) Chain Rule : $\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$
- If the function $f(x)$ is increasing at $x = a$ then $f'(a) > 0$.
- If the function $f(x)$ is decreasing at $x = a$ then $f'(a) < 0$.
- The necessary and sufficient conditions for a function to be maximum at $x = a$: $f'(a) = 0$ and $f''(a) < 0$.
- The necessary and sufficient conditions for a function to be minimum at $x = a$: $f'(a) = 0$ and $f''(a) > 0$.
- Marginal Cost $= \frac{dC}{dx}$
- Marginal Revenue $= \frac{dR}{dx}$
- Elasticity of demand $= -\frac{p}{x} \cdot \frac{dx}{dp}$
- The necessary and sufficient conditions for minimizing the production cost function C :

  $\frac{dC}{dx} = 0$ and $\frac{d^2C}{dx^2} > 0$.
- The necessary and sufficient conditions for maximizing the revenue function R :

  $\frac{dR}{dx} = 0$ and $\frac{d^2R}{dx^2} < 0$.
- The necessary and sufficient conditions for maximizing the profit function P :

  $\frac{dP}{dx} = 0$ and $\frac{d^2P}{dx^2} < 0$.

EXERCISE 5

Section A

Choose the correct option for the following multiple choice questions :

1. What is the formula for derivative of function $f(x)$?

(a) $\lim_{h \to x} \frac{f(x+h) - f(x)}{h}$ (b) $\lim_{h \to 0} \frac{f(x+h) + f(x)}{h}$

(c) $\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$ (d) $\lim_{h \to x} \frac{f(x) - f(x+h)}{h}$

2. What is $\frac{dy}{dx}$ if $y = ax^n$, a is a constant ?

(a) nx^{n-1} (b) $an\,x^{n-1}$ (c) 0 (d) $an\,x^{n+1}$

3. If $y = ax + b$, a and b are constant then what will be $\frac{dy}{dx}$?

(a) a (b) b (c) $a + b$ (d) 0

4. What is the derivative of $f(x) = \frac{4}{x^2}$?

(a) $\frac{4}{2x}$ (b) $-\frac{8}{x^3}$ (c) $\frac{8}{x^3}$ (d) 0

5. If u and v are two functions of x then what is the formula of derivative of their product ?

(a) $u\frac{du}{dx} + v\frac{dv}{dx}$ (b) $u\frac{dv}{dx} - v\frac{du}{dx}$ (c) $\frac{du}{dx} \times \frac{dv}{dx}$ (d) $u\frac{dv}{dx} + v\frac{du}{dx}$

6. If u and v are functions of x then what is the formula for derivative of $\frac{v}{u}$?

(a) $\frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$ (b) $\frac{v\frac{du}{dx} + u\frac{dv}{dx}}{v^2}$ (c) $\frac{u\frac{dv}{dx} + v\frac{du}{dx}}{u^2}$ (d) $\frac{u\frac{dv}{dx} - v\frac{du}{dx}}{u^2}$

7. If the function $f(x)$ is increasing at $x = a$ then which is the correct option from the following ?

(a) $f'(a) < 0$ (b) $f'(a) > 0$ (c) $f'(a) = 0$ (d) $f''(a) > 0$

8. What are the necessary and sufficient conditions for a function to be minimum at $x = a$?

(a) $f'(a) = 0,\ f''(a) < 0$ (b) $f'(a) > 0,\ f''(a) > 0$

(c) $f'(a) = 0,\ f''(a) > 0$ (d) $f'(a) < 0,\ f''(a) > 0$

9. What is the formula for elasticity of demand ?

(a) $-\frac{p}{x} \cdot \frac{dx}{dp}$ (b) $\frac{p}{x} \cdot \frac{dx}{dp}$ (c) $-\frac{x}{p} \cdot \frac{dp}{dx}$ (d) $-\frac{p}{x} \cdot \frac{dp}{dx}$

10. What are the conditions of revenue function R to be maximum ?

(a) $\frac{dR}{dx} = 0, \frac{d^2R}{dx^2} < 0$ (b) $\frac{dR}{dx} = 0, \frac{d^2R}{dx^2} > 0$

(c) $\frac{dR}{dx} > 0, \frac{d^2R}{dx^2} < 0$ (d) $\frac{dR}{dx} > 0, \frac{d^2R}{dx^2} > 0$

Section B

Answer the following questions in one sentence:

1. Define differentiation.
2. Find $f'(x)$ for the function $f(x) = 50$.
3. Find $\frac{dy}{dx}$ if $y = a^n$, a is constant.
4. State the rule for derivative for product of two functions of x.
5. How will be the first order derivative of a function at $x = a$ if function is decreasing at $x = a$?
6. How will be the second order derivative of a function at $x = a$ if function is maximum at $x = a$?
7. What are the stationary points of a function ?
8. What is marginal revenue ?
9. Define marginal cost.
10. State the formula of elasticity of demand.
11. Find $f'(x)$ if $f(x) = 7x^2 - 6x + 5$.
12. Find $\frac{dy}{dx}$ if $y = 6x^3 + \frac{7}{2}x^2 + \frac{6}{5}x - 8$.

Section C

Answer the following questions :

1. Define derivative.
2. State the division rule of derivative.
3. State necessary and sufficient conditions for a function to be maximum at $x = a$.
4. Explain marginal cost and give its formula.
5. Define elasticity of demand.
6. What are the conditions for profit function P to be maximum ?
7. State the conditions for production cost function C to be minimum.
8. Find $f''(x)$ if $f(x) = \sqrt[4]{x}$.
9. Write the chain rule of differentiation.
10. Find $f''(0)$ if $f(x) = x^4 - 4x^3 + 3x^2 + x + 1$.
11. Find marginal revenue if revenue function is $90x - \frac{x^2}{2}$.

12. What is maximum value of a function ?

13. When can it be said that a function is decreasing at a point ?

14. Determine whether the function $y = 12 + 4x - 7x^2$ is increasing or decreasing at $x = 2$.

15. Find the derivative of $y = 4x^2 + 4x + 8$. For which value of x will the derivative be zero ?

16. $f(x) = x^3 + 5x^2 + 3x + 7$, prove that $f'(2) = 35$.

17. If $f(x) = 3x^2 + 3$ then for which value of x, $f'(x) = f(x)$?

18. Find $\frac{d^2y}{dx^2}$ if $y = 2x^3 + 5x^2 - 3 + \frac{4}{x^2} - \frac{5}{x^3}$.

19. Find $\frac{d^2y}{dx^2}$ if $y = \sqrt{x} + \frac{1}{\sqrt{x}}$.

20. Obtain marginal cost if the production cost function is $C = 0.0012x^2 - 0.18x + 25$.

Section D

Answer the following questions :

1. Find derivative of $y = ax + b$ (a and b are constants) using definition.

2. Find derivative of $f(x) = x^{10}$ using definition.

3. Find derivative of $\frac{2}{3 + 4x}$ using definition.

4. $y = x^3 - 3x^2 - 3x + 80$. For which value of x, $\frac{dy}{dx} = -6$?

5. Find $f'(2)$ if $f(x) = \frac{4x^5 + 3x^3 + 2x^2 + 24}{x^2}$.

6. Find the derivative of $y = (3x^2 + 4x - 2)(3x + 2)$ with respect to x.

7. Find $\frac{dy}{dx}$ if $y = \frac{ax + b}{bx + a}$ (a and b are constants).

8. Find the derivative of $y = 1 + \frac{1}{1 + \frac{1}{x}}$ with respect to x.

9. Find $\frac{dy}{dx}$ if $(2x + 3)(y + 2) = 15$.

10. Find $\frac{dy}{dx}$ if $y = 5 + \frac{6}{7x + 8}$.

11. Find $f'(x)$ if $f(x) = \sqrt{x^2 + 5}$.

12. Find the derivative of $(3x^3 - 2x^2 + 1)^{\frac{5}{2}}$ with respect to x.

13. Find $f'(x)$ if $f(x) = \left(x^2 + 3x + 4\right)^7$.

14. If $f(x) = 3x^2 + 4x + 5$ then for which value of x, $f'(x) = f''(x)$?

15. Find marginal revenue if demand function is $p = \frac{2500 - x^2}{100}$.

16. Determine whether the function $y = 3x^2 - 10x + 7$ is increasing or decreasing at $x = 1$ and $x = 2$.

17. Determine whether the function $y = 2x^3 - 7x^2 - 11x + 5$ is increasing or decreasing at $x = \frac{1}{2}$ and $x = 3$.

18. Determine whether the function $y = 3 + 2x - 7x^2$ is increasing or decreasing at $x = -4$ and $x = 4$.

19. Production cost of a factory producing sugar is $C = \frac{x^2}{10} + 5x + 200$. Find the marginal cost if the production is 100 units and interpret it.

20. The cost function of producing x units of a commodity is $C = 50 + 2x + \sqrt{x}$. Find the marginal cost if the production is 100 units and interpret it.

21. State the method of obtaining maximum or minimum value of a function.

Section E

Answer the following questions :

1. Give working rules for differentiation.
2. How can it be decided using derivative that the function is increasing or decreasing at a point ?
3. What is maximum value of a function ? State the conditions for maximum value.
4. What is minimum value of a function ? State the conditions for minimum value.
5. In a factory, production cost per hundred tons of steel is $\frac{1}{10}x^3 - 4x^2 + 50x + 300$. Determine the production for minimum cost.
6. The cost of producing x units of an item is $C = 1000 + 8x + \frac{5000}{x}$. What should be the production for minimum cost ? Also find the minimum cost.
7. Production cost function of a commodity is $C = 1500 + 0.05x - 2\sqrt{x}$. Prove that production is minimum when 400 units are produced.
8. The demand function of an item is $p = 30 - \frac{x^2}{10}$. Find the demand and price for maximum revenue.
9. In a market, demand law of rice is $x = 3(60 - p)$. Find the demand for maximum revenue. Also find the price and revenue for that demand.
10. If the demand function is $p = 75 - \frac{x^2}{2500}$ then at which demand is revenue maximum ? Also find the price for maximum revenue.

11. The profit function of a producer is $40x + 10000 - 0.1x^2$. At what production is the profit maximum ? Also find this maximum profit.

12. The profit function of a merchant is $5x - 100 - 0.01x^2$. How many units should be produced for maximum profit ?

Section F

Solve the following :

1. Find the values of x which maximize or minimize $y = 2x^3 - 15x^2 + 36x + 12$. Also find the maximum and minimum values of y.

2. Find the values of x which maximize or minimize $f(x) = 2x^3 + 3x^2 - 36x + 10$. Also find the maximum and minimum values of $f(x)$.

3. Find the maximum and minimum values of $f(x) = x^3 - x^2 - x + 2$.

4. A producer produces x units at cost $200x + 15x^2$. The demand function is $p = 1200 - 10x$. Find the profit function and how many units should be produced for maximum profit ?

5. The selling price of a refrigerator as determined by the company is ₹ 10,000. The total cost of the production for x refrigerator is $C = 0.1x^2 + 9000x + 100$ rupees. How many refrigerators should be manufactured for maximum profit ?

6. A toy is sold at ₹ 20. Total cost of producing x such toys is $C = 1000 + 16.5x + 0.001x^2$ rupees. How many toys should be produced for maximum profit ?

Gottfried Wilhelm Leibniz
(1646 - 1716)

Gottfried Leibniz was a German polymath and philosopher who occupies a prominent place in the history of mathematics and the history of philosophy, having developed differential and integral calculus independently of Isaac Newton. It was only in the 20th century that his Law of Continuity and Transcendental Law of Homogeneity found mathematical implementation (by means of non-standard analysis). He became one of the most prolific inventors in the field of mechanical calculators.

Leibniz made major contributions to physics and technology, and anticipated notions that surfaced much later in philosophy, probability theory, biology, medicine, geology, psychology, linguistics, and computer science.

Answers

Exercise 1.1

1. (1) $U = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

(2) $U = \{(1, H), (2, H), (3, H), (4, H), (5, H), (6, H),$
$(1, T), (2, T), (3, T), (4, T), (5, T), (6, T)\}$

(3) $U = \{(a, b), (a, c), (a, d), (a, e), (b, c),$
$(b, d), (b, e), (c, d), (c, e), (d, e)\}$

2. (1) U = {0, 1, 2, ..., 100}, No. of sample points = 101

3. Denoting four persons by a, b, c, d

U = {(a, b), (a, c), (a, d), (b, c), (b, d), (c, d), (b, a), (c, a), (d, a), (c, b), (d, b), (d, c)}

The first place in the bracket shows minister and the second place shows deputy minister.

4. U = {H, TH, TTH, TTTH, ...}, infinite sample space

5. U = {(1, 2, 3), (1, 2, 4), (1, 2, 5), (1, 3, 4), (1, 3, 5), (1, 4, 5), (2, 3, 4), (2, 3, 5), (2, 4, 5), (3, 4, 5) }

6. (1) A = {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}

(2) B = {3, 6, 9, 12, 15, 18}

(3) C = {2, 3, 4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20}

7. U= {BB, BG, GB, GG}

(1) A_1 = {BG, GB}

(2) A_2 = {BG, GB, GG}

8. U= {(i, j); i, j = 1, 2, 3, 4, 5, 6}

(1) A_1 = {(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)}

(2) A_2 = {(1, 1), (1, 2), (2, 1),}

(3) A_3 = {(1, 2), (1, 5), (2, 1), (2, 4), (3, 3), (3, 6), (4, 2), (4, 5), (5, 1), (5, 4), (6, 3), (6, 6)}

(4) A_4 = { }

9. (1) U = {(1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5)}

(2) A = {(1, 5), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5)}

(3) B = {(1, 3), (1, 5), (2, 4), (3, 5)}

(4) $A \cup B$ = {(1, 3), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5)}

(5) $A \cap B$ = {(1, 5), (2, 4), (3, 5)}

(6) A' = {(1, 2), (1, 3), (1, 4), (2, 3)}

(7) $A - B$ = {(2, 5), (3, 4), (4, 5)}

(8) $A' \cap B$ = {(1, 3)}

(9) Events A and B are not mutually exclusive as $A \cap B \neq \phi$

(10) No. of sample points = 10

10. Denoting three females by *a, b, c* and two males by *x, y*.

(1) U = {*a, b, c, x, y*}

(2) A = {*a, b, c*}

(3) B = {*x, y*}

(4) $A \cup B$ = {*a, b, c, x, y*}

(5) $A \cap B$ = { }

(6) $A' \cap B$ = {*x, y*}

(7) Events A and B are mutually exclusive as $A \cap B = \phi$

(8) Events A and B are exhaustive as $A \cup B = U$

11. (1) U = $\{S_A, S_2, S_3, \ldots\ldots\ldots, S_K, D_A, D_2, D_3, \ldots\ldots\ldots, D_K,$
$C_A, C_2, C_3, \ldots\ldots\ldots, C_K, H_A, H_2, H_3, \ldots\ldots\ldots, H_K\}$

(2) A = $\{S_A, S_2, S_3, \ldots\ldots\ldots, S_K\}$

(3) B = $\{S_A, S_2, S_3, \ldots\ldots\ldots, S_{10}, D_A, D_2, D_3, \ldots\ldots\ldots, D_{10},$
$C_A, C_2, C_3, \ldots\ldots\ldots, C_{10}, H_A, H_2, H_3, \ldots\ldots\ldots, H_{10}\}$

(4) $A \cup B = \{S_A, S_2, S_3, \ldots\ldots\ldots, S_k, D_A, D_2, D_3, \ldots\ldots\ldots, D_{10},$
$C_A, C_2, C_3, \ldots\ldots\ldots, C_{10}, H_A, H_2, H_3, \ldots\ldots\ldots, H_{10}\}$

(5) $A \cap B = \{S_1, S_2, S_3, \ldots\ldots\ldots, S_{10}\}$

(6) $B' = \{S_J, S_Q, S_K, D_J, D_Q, D_K, C_J, C_Q, C_K, H_J, H_Q, H_K,\}$

12. $A_1 \cup A_2 = \{x \mid 0 \le x < 5\}$

$A_1 \cap A_2 = \{x \mid x = 1, 2\}$

13. $A_1 \cup A_2 = \{x \mid 2 \le x \le 8, x \in N\}$

$A_1 \cap A_2 = \{x \mid x = 4, 5\}$

14. $A' = \{x \mid x = 0, 1, 3, 5, 7, 8, 9, 10\}$

15. $A' = \{x \mid 0 < x < \frac{1}{2}\}$

Exercise 1.2

1. (1) $\frac{1}{8}$ (2) $\frac{1}{8}$ (3) $\frac{7}{8}$ (4) $\frac{1}{2}$ (5) $\frac{1}{2}$

(6) $\frac{1}{2}$ (7) $\frac{1}{4}$ (8) $\frac{1}{2}$ **2.** (1) $\frac{5}{36}$ (2) $\frac{11}{12}$

(3) $\frac{1}{3}$ (4) $\frac{1}{9}$ **3.** (1) $\frac{1}{2}$ (2) $\frac{3}{4}$ **4.** $\frac{7}{50}$

5. (1) $\frac{1}{3}$ (2) $\frac{2}{3}$ (3) $\frac{1}{4}$ (4) $\frac{3}{4}$ (5) $\frac{1}{12}$

6. $\frac{1}{30}$ **7.** $\frac{1}{20}$ **8.** $\frac{2}{5}$ **9.** $\frac{1}{7}$ **10.** $\frac{1}{7}$

11. $\frac{1}{7}$ **12.** $\frac{3}{7}$ **13.** (1) $\frac{1}{7}$ (2) $\frac{4}{7}$ (3) $\frac{3}{7}$

14. (1) $\frac{3}{28}$ (2) $\frac{5}{14}$ (3) $\frac{15}{28}$ **15.** (1) $\frac{26}{51}$ (2) $\frac{11}{221}$

(3) $\frac{32}{221}$ **16.** $\frac{14}{15}$ **17.** (1) 0.4 (2) 0.35 (3) 0.45

(4) 0.05 (5) 0.85 **18.** $P(A-B) = 0.2,\ P(B-A) = 0.5$

Exercise 1.3

1. (1) $\frac{4}{17}$ (2) $\frac{25}{51}$ **2.** $\frac{3}{7}$ **3.** (1) $\frac{4}{13}$ (2) $\frac{9}{13}$

4. $\frac{47}{100}$ **5.** $\frac{2}{3}$ **6.** 0.9 **7.** 0.97 **8.** 0.59

9. $P(A) = \frac{2}{3}$ **10.** $P(A \cup B) = \frac{35}{47}$, $P(B \cup C) = \frac{32}{47}$ **11.** $P(A \cup B \cup C) = 0.8$

12. $P(A) = 0.4$, $P(B) = 0.4$

Exercise 1.4

1. $\frac{1}{2}$ **2.** $\frac{1}{5}$ **3.** (1) $\frac{5}{8}$ (2) $\frac{5}{6}$ **4.** $\frac{4}{5}$

5. $P(A/B) = \frac{5}{6}$ **6.** $P(A \cap M) = \frac{1}{20}$, $P(A \cap F) = \frac{1}{4}$ **7.** $\frac{11}{24}$ **8.** $\frac{8}{15}$

9. $\frac{9}{100}$ **10.** $\frac{4}{15}$ **11.** (1) $\frac{11}{36}$ (2) $\frac{1}{6}$ **12.** $\frac{23}{24}$

13. $\frac{14}{15}$ **14.** 0.26 **15.** 0.72

Exercise 1.5

1. (1) $\frac{29}{357}$ (2) $\frac{125}{357}$ (3) $\frac{275}{357}$ **2.** (1) $\frac{1319}{2536}$ (2) $\frac{1319}{2437}$

Exercise 1

Section A

1. (d)	**2.** (b)	**3.** (c)	**4.** (a)	**5.** (a)
6. (c)	**7.** (c)	**8.** (b)	**9.** (a)	**10.** (b)
11. (d)	**12.** (b)	**13.** (a)	**14.** (a)	**15.** (c)
16. (c)	**17.** (b)			

Section B

13. $P(A \cap B) = 0$, $P(A \cup B) = 1$ **15.** $A \cap B = \{x \mid \frac{1}{4} \leq x < 1\}$ **16.** 0.15

17. 0.55 **18.** 0.1 **19.** 2K **20.** 0.45 **21.** $\frac{2}{3}$

22. 0.98 **23.** $2^5 = 32$ **24.** $2 \times 6 \times 6 = 72$ **25.** Not possible. As $P(A \cup B) < P(A)$

26. 2704 **27.** $\frac{2}{5}$ **28.** $\frac{1}{1000}$

Section C

12. $\frac{1}{6}$ **13.** 0.08 **14.** 0.0024 **15.** $\frac{19}{20}$ **16.** 0.58

17. $A \cup B = \{x \mid \frac{1}{2} \le x < 3\}$, $A \cap B = \{x \mid 1 < x < 2\}$ **18.** $\frac{1}{20}$ **19.** $\frac{1}{6}$

20. (1) $\frac{1}{6}$ (2) $\frac{3}{10}$ **21.** (1) 0.44 (2) 0.09 **22.** $\frac{17}{20}$

23. 0.976

Section D

1. $\frac{31}{80}$ **2.** $\frac{4}{25}$ **3.** $\frac{1}{2}$ **4.** $p(3-3p+p^2)$

5. (1) $\frac{9}{10}$ (2) $\frac{4}{5}$

●

Exercise 2.1

1. Given distribution is a probability distribution of variable x.

2. $k = 30$ **3.** $k = \frac{24}{17}$, P(1 < x < 4) = $\frac{5}{17}$ **4.** $k = \frac{1}{4}$, Mean = $\frac{3}{4}$

5. Mean = $\frac{-1}{8}$, Variance = $\frac{135}{64}$

6. Probability distribution of sum

x	2	3	4	5	6	7	8	9	10	11	12	કુલ
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Expected value of sum = 7

7. Probability distribution of x

x	1	2	3	Total
$p(x)$	$\frac{4}{20}$	$\frac{12}{20}$	$\frac{4}{20}$	1

Expected number of red balls = 2

8. Probability distribution of x

x	1	2	3	4	5	Total
$p(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	1

Mean = $\frac{31}{16}$, Variance = $\frac{367}{256}$

9. Expected value of the prize = ₹ $\frac{20}{3}$

Exercise 2.2

1. $p(X \le 1) = \frac{9}{256}$ 2. $n = 6, \quad p = \frac{5}{6}; \frac{15625}{46656}$ 3. 0.6912 4. 0.0729

5. 0.1382

Exercise 2

Section A

1. (d) 2. (c) 3. (c) 4. (a) 5. (d)

6. (d) 7. (b) 8. (b) 9. (b) 10. (b)

Section B

6. 14 7. $\frac{12}{5}$ 8. $p + q = 1$ 9. Mean > Variance 10. 0.4

Section C

1. $C = 0.1$ 2. $\frac{10}{3}$ 3. 2.8 4. $\frac{1}{16}$ 6. $\frac{8}{9}$

7. $\frac{4}{3}$ 8. $n = 8, p = \frac{1}{2}$ 9. 2 10. 0.96

Section D

1. $k = \frac{1}{5}; \frac{2}{5}$ 2. $c = \frac{1}{10}$ 3. $k = \frac{1}{326}$, મધ્યક $= \frac{1305}{326}$ 6. $\frac{54}{125}$, Mean $= \frac{9}{5}$

7. $\frac{1620}{16807}$ 8. $p(1) = \frac{162}{625}$, $p(2) = \frac{216}{625}$ 9. 0.0146, Variance = 0.54

Section E

1. Expected demand = 3.62, Variance = 2.2156

2.

x	0	1	2	Total
$p(x)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	1

3. (i) 0.9510 (ii) 0.0490 4. 0.1631 5. 0.3446 6. 0.5443

7. (i) 0.3599 (ii) 0.1066

Section F

(1)

x	0	1	2	3	Total
$p(x)$	$\frac{84}{220}$	$\frac{108}{220}$	$\frac{27}{220}$	$\frac{1}{220}$	1

Expected value $= \frac{165}{220}$, Variance $= 0.4330$

(2) 56

●

Exercise 3

Section A

1. (c) **2.** (d) **3.** (b) **4.** (a) **5.** (c)
6. (b) **7.** (d) **8.** (c) **9.** (a) **10.** (b)
11. (c) **12.** (d) **13.** (a) **14.** (b) **15.** (c)

Section B

2. 0 **4.** 0 **5.** Yes **6.** $z = 0$ **7.** Mean
8. 95.45 % **9.** 20 **10.** 4 **11.** 10 **12.** 18
13. (–0.675, 0.675) **14.** 25 **15.** No **16.** 1.5 **17.** 50

Section C

7. $Q_1 = 6.63$ **8.** 12 **9.** $Q_1 = 40$ **10.** (90, 110) **11.** ± 2.575

Section D

4. (1) 0.3413 (2) 0.6826 **5.** (1) 84.13 % (2) 97.72 % **6.** (1) 409 approximately
(2) 11 approximately **7.** (1) $Z_1 = 2.445$ (2) $Z_1 = 1.96$ **8.** (1) $Z_1 = -1.035$ (2) $Z_1 = -0.675$

9. (1) $Z_1 = -0.5$ (2) $Z_1 = 1.08$ **10.** $\mu = 2000$ $\sigma = 400$

11. $Q_1 = 95.95$, $Q_3 = 104.05$ **12.** (5, 95) **13.** $\mu = 15.75$

14. (1) $Q_1 = 193.25$, $Q_3 = 206.75$ (2) 6.67 (3) 8

Section E

1. (1) 0.3785 (2) 0.2426

2. No. of fat persons = 33, No. of healthy persons = 192, No. of physically weak persons = 11

3. (1) 6.06 % (2) 65.54 % (3) 78.81 %

4. ₹ 8320 and ₹ 12,560

5. (49.44, 54.56)

6. $x_1 = 55$, 8 weeks (approximately)

7. $\sigma^2 = 1362.25$, 0.2148

8. $\mu = 21.15$ mm, 80.27 %

9. $D_4 = 392.35$, $P_{90} = 438.4$

10. (1) 150 (2) 140

Section F

1. (1) 250 students (2) 30.54 % (3) 96 marks (approximately)

2. (i) 30.13 years (ii) 33.79 years (iii) 47.68 years

3. (a) 456 students (b) 1846 (c) 16362 (d) 1336

4. N = 5051, ₹ 8350

5. $\mu = 62.12$, $\sigma = 17.28$, $Q_3 = 73.79$

6. $\mu = 4300$, $\sigma^2 = 2500$, (3320, 5280)

7. (1) $x_2 = 157$ (2) $x_1 = 68$ (3) 0.1401

8. (1) 50 (2) $Q_1 = 43.25$, $Q_3 = 56.75$ (3) $\frac{20}{3}$ (4) 8

●

Exercise 4.1

1. (1) Modulus form : $|x-4| < 0.4$

Interval form : (3.6, 4.4)

(2) Modulus form : $|x-2| < 0.02$

Interval form : (1.98, 2.02)

(3) Modulus form : $|x| < 0.05$

Interval form : (–0.05, 0.05)

(4) Modulus form : $|x+1| < 0.001$

Interval form : (–1.001, –0.999)

2. (1) Interval form : (1.99, 2.01)

Neighbourhood form : N (2, 0.01)

(2) Interval form : (–5.1, –4.9)

Neighbourhood form : N (–5, 0.1)

(3) Interval form : $\left(-\frac{1}{3}, \frac{1}{3}\right)$

Neighbourhood form : N $\left(0, \frac{1}{3}\right)$

(4) Interval form : (–3.15, –2.85)

Neighbourhood form : N (–3, 0.15)

3. (1) Modulus form : $|x - 4.3| < 0.5$

Neighbourhood form : N (4.3, 0.5)

(2) Modulus form : $|x - 2| < 0.05$

Neighbourhood form : N (2, 0.05)

(3) Modulus form : $|x - 0.5| < 0.9$

Neighbourhood form : N (0.5, 0.9)

(4) Modulus form : $|x - 2| < 0.002$

Neighbourhood form : N (2, 0.002)

4. Interval form : (15.5, 16.5)

Modulus form : $|x - 16| < 0.5$

5. $b = 0.05$, $k = 3.05$

6. $K_1 = 0.01$, $K_2 = 9.99$

Exercise 4.2

1. (1) 3 (2) 4 (3) 11 (4) –3 (5) 2

Exercise 4

Section A

1. (b)	**2.** (c)	**3.** (a)	**4.** (b)	**5.** (d)
6. (c)	**7.** (d)	**8.** (a)	**9.** (a)	**10.** (d)
11. (b)	**12.** (c)			

Section B

1. (–0.09, 0.09) **2.** $|x+5|<0.001$ **3.** $N(10, \frac{1}{10})$ **4.** $\left(-\frac{1}{4}, \frac{1}{4}\right)$

5. $|x-50|<0.8$ **6.** $a = 7$ **7.** $k = -4.04$ **8.** 20 **9.** 2

10. 2 **11.** 80 **12.** ma^{m-1} **13.** $k = 10$ **14.** $k = 5$

Section C

4. $|x|<0.5$ **5.** N (–8, 0.75) **6.** $K_1 = 20,\ K_2 = 20.5$

7. Neighbourhood form : $N\left(-\frac{1}{3}, \frac{2}{3}\right)$ Interval form : $\left(-1, \frac{1}{3}\right)$ **8.** $A_1 = 4$ $A_2 = 3.91$

Section D

1. 2 **2.** $\frac{1}{9}$ **3.** –8 **4.** 2 **5.** $\frac{7}{4}$

6. $\frac{3}{5}$ **7.** $-\frac{1}{3}$ **8.** $\frac{31}{3}$ **9.** $-\frac{1}{7}$ **10.** 1

11. 1 **12.** $-\frac{4}{3}p$ **13.** $\frac{27}{2}$ **14.** –64 **15.** $-\frac{2017}{2018}$

16. $\frac{7}{3}$ **17.** $\frac{2}{3}$

Section E

II. (1) 7 (2) 7 (3) –3 (4) –1

III. (1) $7x^6$ (2) $\frac{1}{10}$ (3) n (4) 1 (5) $3x^2$

(6) $7x^6$ (7) $\frac{1}{6}$ (8) 8 (9) 4 (10) 5

●

Exercise 5.1

1. 2 **2.** $2x$ **3.** $7x^6$ **4.** $\frac{-1}{(x+1)^2}$ **5.** $\frac{1}{3x^{\frac{2}{3}}}$

6. $\frac{-6}{(3x-4)^2}$ **7.** 0

Exercise 5

Section A

1. (c) **2.** (b) **3.** (a) **4.** (b) **5.** (d)

6. (d) **7.** (b) **8.** (c) **9.** (a) **10.** (a)

Section B

2. 0 **3.** 0 **5.** Negative **6.** Negative **11.** $14x-6$

12. $18x^2+7x+\frac{6}{5}$

Section C

8. $-\frac{3}{16x^{\frac{7}{4}}}$ **10.** 6 **11.** $90-x$ **14.** Decreasing **15.** $-\frac{1}{2}$

17. 1 **18.** $12x+10+\frac{24}{x^4}-\frac{60}{x^5}$ **19.** $\frac{-1}{4x^{\frac{3}{2}}}+\frac{3}{4x^{\frac{5}{2}}}$

20. $0.0024x - 0.18$

Section D

1. a **2.** $10x^9$ **3.** $\frac{-8}{(3+4x)^2}$ **4.** 1 **5.** 45

6. $27x^2+36x+2$ **7.** $\frac{a^2-b^2}{(bx+a)^2}$ **8.** $\frac{1}{(x+1)^2}$ **9.** $\frac{-30}{(2x+3)^2}$

10. $\frac{-42}{(7x+8)^2}$ **11.** $\frac{x}{\sqrt{x^2+5}}$ **12.** $\frac{5}{2}(3x^3-2x^2+1)^{\frac{3}{2}}(9x^2-4x)$

13. $7(x^2+3x+4)^6(2x+3)$ **14.** $\frac{1}{3}$ **15.** $25-\frac{3x^2}{100}$

16. Function is decreasing at $x=1$

Function is decreasing at $x=2$

17. Function is decreasing at $x=\frac{1}{2}$

Function is decreasing at $x=3$

18. Function is increasing at $x = -4$

Function is decreasing at $x = 4$

19. Marginal cost $= \frac{x}{5} + 5$

$x = 100$, Marginal cost $= 25$

20. Marginal cost $= 2 + \frac{1}{2\sqrt{x}}$

$x = 100$ Marginal cost $= 2.05$

Section E

5. $\frac{50}{3}$ hundred tons **6.** $x = 25$, Minimum cost $= 1400$ **8.** $x = 10,\ p = 20$

9. $p = 30,\ x = 90,\ R = 2700$ **10.** $x = 250,\ p = 50$ **11.** $x = 200$, Maximum profit $= 14{,}000$

12. $x = 250$

Section F

1. y is maximum at $x = 2$, Maximum value of y is 40

y is minimum at $x = 3$, Minimum value of y is 39

2. $f(x)$ is maximum at $x = -3$, Maximum value of $f(x)$ is 91

$f(x)$ is minimum at $x = 2$, Minimum value of $f(x)$ is -34

3. $f(x)$ is maximum at $x = \frac{-1}{3}$, Maximum value of $f(x)$ is $\frac{59}{27}$

$f(x)$ is minimum at $x = 1$, Minimum value of $f(x)$ is 1

4. Profit function $= 1000x - 25x^2$

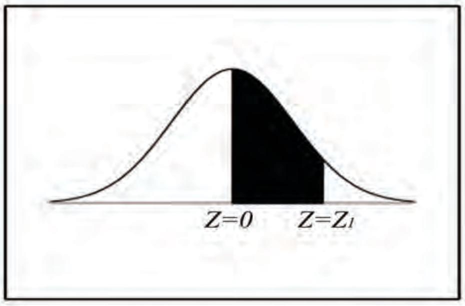
Profit is maximum at $x = 20$

5. 5000 refrigerators

6. 1750 toys

● ● ●

Table of Standard Normal Curve



Area Under the Standard Normal Curve

$Z = 0$ to $Z = Z_1$, z being standard normal variate

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4762	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998