Required Probability $=\frac{2}{6}=\frac{1}{3}$

Example 8. Find the probability of drawing a face card in a single random draw from a well shuffled pack of 52 cards.

Solution: There are 52 cards in a pack of cards.

Total number of cases = 52

Number of favourable cases (face cards include the Jack, Queen and King in each) = 12

Required Probability $=\frac{12}{52}=\frac{3}{13}$

Example 9. A card is drawn from an ordinary pack of playing cards and a person bets that it is a spade or an ace. What are odds against his winning this bet?

Solution: Total number of cases = 52

Since there are 13 spades and 3 aces (one ace is also present in spades), Therefore the favourable cases = 13 + 3 = 16

The probability of winning the bet $=\frac{16}{52}=\frac{4}{13}$

The probability of losing the bet $=1-\frac{4}{13}=\frac{9}{13}$

Hence, odds against winning the bet $=\frac{9}{13}:\frac{4}{13}=9:4$

Example 10. A single letter is selected at random from the word 'PROBABILITY'. What is the probability that it is a vowel?

Solution: There are 11 letters in the word 'PROBABILITY' out of which 1 is be selected.

∴ Total No. of words = 11

There are four vowels viz. O, A, I, I. Therefore favourable number of cases = 4

Hence, the required probability $=\frac{4}{11}$

Example 11. Find the probability of drawing an ace from a set of 52 cards.

Solution: Number of exhaustive cases (n) = 52

There are 4 ace cards in an ordinary pack.

∴ Favourable cases (n) = 4

∴ Probability of getting an ace $=\frac{4}{52}=\frac{1}{13}$

Example 12. What is the probability that a leap year selected at random will contain 53 Sundays?

Solution: Total number of days in a leap year = 366

Number of weeks in a year $=\frac{366}{7}=52\frac{2}{7}$

= 52 weeks and 2 days

Following may be the 7 possible combinations of these two extra days:

(*i*) Monday and Tuesday (*ii*) Tuesday and Wednesday

(*iii*) Wednesday and Thursday (*iv*) Thursday and Friday

(*v*) Friday and Saturday (*vi*) Saturday and Sunday

(*vii*) Sunday and Monday

A selected leap year can have 53 Sundays if these two extra days happen to be a Sunday

Total possible outcomes of 2 days = $n = 7$

Number of cases having Sundays = $m = 2$

∴ The required probability $=\frac{2}{7}$

Example 13. Two dice are tossed. Find the probability that the sum of dots on the faces that turn up is (*i*) 8 (*ii*) 7 (*iii*) 11.

Solution: There are 36 likely chances of throwing of two dice which are given below:

(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)

6×6=36

(*i*) Total number of possible outcomes = 36

Number of outcomes favourable to 8 are

(6, 2) (5, 3) (4, 4) (3, 5) and (2, 6)

i.e., the number of outcomes favourable to 8 = 5

∴ P(Sum of dots is 8) $=\frac{5}{36}$

(*ii*) The number of outcomes favourable to 7 are

(6, 1) (5, 2) (4, 3) (3, 4) (2, 5) (1, 6)

i.e., the number of outcomes favourable to 7 = 6

∴ P(Sum of dots is 7) $=\frac{6}{36}=\frac{1}{6}$

(*iii*) The number of outcomes favourable to 11 are (6, 5) (5, 6)

∴ the number of outcomes favourable to 11 = 2

∴ $P(\text{Sum of dots is } 11) = \frac{2}{36} = \frac{1}{18}$

Example 14. The following table gives the distribution of wages:

Wages per day in Rs.:	30—40	40—50	50—60	60—70	70—80	80—90
No. of wage earners:	20	45	68	35	20	12

An individual is selected at random from the above group. Find the probability that (*i*) his wages were under Rs. 50, (*ii*) his wages were Rs. 60 or over and (*iii*) his wages were either between Rs. 30—40 or 70—80.

Solution: Total number of wages earners are -

$$n = 20 + 45 + 68 + 35 + 20 + 12 = 200$$

(*i*) Number of wage earners having wages under Rs. 50

$$m = 20 + 45 = 65$$

∴ $\text{Required Probability} = \frac{\text{No. of cases favourable}}{\text{Total No. of cases}} = \frac{m}{n} = \frac{65}{200} = \frac{13}{40}$

(*ii*) Number of wage earners having wages 60 or over

$$m = 35 + 20 + 12 = 67$$

∴ $\text{Required Probability} = \frac{m}{n} = \frac{67}{200}$

(*iii*) Number of wage earners with wages between Rs. 30—40 or 70—80

$$m = 20 + 20 = 40$$

$\text{Required Probability} = \frac{\text{Favourable cases}}{\text{Total No. of cases}} = \frac{m}{n} = \frac{40}{200} = \frac{1}{5}$

Example 15. Suppose an ideal die is tossed twice. What is the probability of getting a sum 10 in two tosses?

Solution: First die can be thrown in 6 ways

Second die can be thrown in 6 ways

Total probable ways of throwing of a die twice = 6 × 6 = 36

36 possible outcomes are shown below:

(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)

Number of outcomes favourable to sum 10 are

(6, 4) (5, 5) (4, 6) = 3

∴ $P(\text{Sum } 10) = \frac{3}{36} = \frac{1}{12}$

Example 16. In a single throw of 3 dice, find the probability of getting the same number on each of them.

Solution: Total number of cases = 6 × 6 × 6 = 216

The number of favourable cases = 6

[(1,1,1), (2,2,2), (3,3,3) ..... (6,6,6)]

∴ The required probability $= \frac{6}{216} = \frac{1}{36}$.

o Use of Combinations in Theory of Probability

The concept of combination is very useful in understanding the theory of probability. It is not always possible that the number of cases favourable to the happening of an event is easily determinable. In such cases, the concept of combination is used. The different selections(or groups) that can be made out of a given set of things taking some or all of them at a time, are called combinations. The combinations of n things, taking r at a time is denoted by ^nC_r.

Symbolically,

$$^nC_r = \frac{n!}{\overline{n-r}!\ r!}$$

For example, if three letters A, B and C are to be arranged in two's, the number of combinations will be: AB, AC and BC. In terms of formula,

$$^3C_2 = \frac{3!}{\overline{3-2}!\ 2!} = \frac{3 \times 2 \times 1}{1 \times 2 \times 1} = 3 \text{ ways}$$

Example 17. A bag contains 5 black, 3 white and 2 red balls. In how many ways can (*i*) 3 balls be drawn and (*ii*) 3 black balls be drawn?

Solution: Total number of balls in a bag = 5 + 3 + 2 = 10

(*i*) Total number of balls drawn = 3

∴ Total number of ways is which 3 balls can be drawn out of 10

$$= {^{10}C_3}$$

$$= \frac{10!}{(10-3)!\ 3!} = \frac{10}{7!\ 3!} = \frac{10 \times 9 \times 8 \times 7!}{7!\ 3!} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120$$

(*ii*) Total number of ways in which 3 black balls from 5 black balls can be drawn

$$= {^5C_3}$$

$$= \frac{5!}{\overline{5-3}!\ 3!} = \frac{5!}{2!\ 3!} = \frac{5 \times 4 \times 3!}{2 \times 1 \times 3!} = 10$$

Example 18. A bag contains 5 green and 7 red balls. Two balls are drawn. What is the probability that one is green and the other is red?

Solution: Total number of balls in the bag = 7 + 5 = 12

Two balls can be drawn from 12 balls in ${}^{12}C_2$ ways.

There are 5 green and 7 red balls in the bag.

A green ball can be drawn in ${}^{5}C_1$ ways.

A red ball can be drawn in ${}^{7}C_1$ ways.

$$\therefore \text{ The required probability is } = \frac{{}^{5}C_1 \cdot {}^{7}C_1}{{}^{12}C_2}$$

$$= \frac{5\times 7}{66} = \frac{35}{66}$$

Example 19. From a pack of 52 cards, two cards are drawn at random. Find the probability that one a king and the other is queen.

Solution: Two cards can be drawn from 52 in ${}^{52}C_2$ ways.

There are 4 kings and 4 queens in a pack of cards.

A king can be drawn in ${}^{4}C_1$ ways.

A queen can be drawn in ${}^{4}C_1$ ways.

$$\therefore \text{ The required probability} = \frac{{}^{4}C_1 \cdot {}^{4}C_1}{{}^{52}C_2}$$

$$= \frac{4\times 4\times 2}{52\times 51} = \frac{8}{663}$$

Example 20. A bag contain 9 red balls, 7 white balls, and 4 green balls. Three balls are drawn randomly without replacement. Find the probability of getting:

(*i*) one ball of each colour

(*ii*) only two red balls

(*iii*) no white balls.

Solution: (*i*) Required Probability $= \frac{{}^{9}C_1 \cdot {}^{7}C_1 \cdot {}^{4}C_1}{{}^{20}C_3} = \frac{9\times 7\times 4}{1140} = 0.221$

(*ii*) Required Probability $= \frac{{}^{9}C_2 \times {}^{11}C_1}{{}^{20}C_3} = \frac{36\times 11}{1140} = 0.347$

(*iii*) Required Probability $= 1 - \frac{{}^{7}C_3}{{}^{20}C_3} = 1 - \frac{35}{1140} = 0.969$

Example 21. What is the probability of getting 9 cards of the same suit in one hand at a game of bridge?

Solution: In a game of bridge, each player has 13 cards with him. If the particular player is to have 9 cards of the same suit, he must have remaining 4 cards from the other 39 cards. Since there are 4 suits in a pack of cards, total number of favourable cases will be:

$$= 4\times {}^{13}C_9 \times {}^{39}C_4$$

$$\text{Required Probability} = \frac{4\times {}^{13}C_9 \times {}^{39}C_4}{{}^{52}C_{13}}$$

EXERCISE 7.1

1. An urn contains two blue balls and three white balls. Find the probability of a blind man obtaining one blue ball in a single draw. **[Ans. 2/5]**
2. What is the probability that a vowel selected at random in any English book is an '*i*'? **[Ans. 1/5]**
3. Find the probability of drawing a black card in single random draw from a well-shuffled pack of ordinary playing cards. **[Ans. 1/2]**
4. Find the probability of drawing (*i*) a spade, (*ii*) an ace, (*iii*) an ace of spade from a well-shuffled pack of ordinary playing cards. **[Ans. (*i*) 1/4, (*ii*) 1/13, (*iii*) 1/52]**
5. Find the probability of having at least one son in a family if there are two children in a family on an average. [Hint: SS, SD, DS, DD] **[Ans. 3/4]**
6. What is the probability that a non-leap year should have 53 Sundays. **[Ans. 1/7]**
7. If two dice are thrown—
   (*i*) What is the probability of total of 7?
   (*ii*) What is the probability of total of 8? **[Ans. (*i*) 1/6, (*ii*) 5/36]**
8. From a well shuffled pack of cards, a card is drawn. What is the probability that it is:
   (*i*) a card of spade
   (*ii*) a king of heart
   (*iii*) a queen of diamond
   (*iv*) a ace of club **[Ans. (*i*) 1/4, (*ii*) 1/52, (*iii*) 1/52, (*iv*) 1/52]**
9. Find the probability of a ball being green when it is drawn out of a bag containing 7 green and 7 white balls. **[Ans. 1/2]**
10. What is the probability of drawing a court card (king, queen and knave) from a deck of 52 playing-cards? **[Ans. 3/13]**
11. Two unbiased dice are thrown. Find the probability that both dice show the same number. **[Ans. 1/6]**
12. Tickets are numbered from 1 to 100. These are well shuffled and a ticket is drawn at random. What is the probability that the drawn ticket has (*i*) a number which is greater than 75 and (*ii*) a perfect square. **[Ans. (*i*) 1/4, (*ii*) 1/10]**

13. If three dice are thrown simultaneously, find the probability of getting a sum of (*i*) 5 and (*ii*) at the most 5 and (*iii*) at least 5.
[**Hint:** See Example 119] **[Ans.** (*i*) $\frac{1}{36}$ (*ii*) $\frac{5}{108}$ (*iii*) $\frac{53}{54}$]

14. Find the probability of not getting a sum 7 in a single throw with a pair of dice. **[Ans.** 5/6]

15. Five balls are drawn from a bag containing 6 white and 4 black balls. What is the probability that 3 white and 2 black balls are drawn. **[Ans.** $\frac{10}{21}$]

16. Four cards are drawn from a pack of cards. Find the probability that (*i*) All are diamonds (*ii*) There are two spades and two hearts. **[Ans.** (*i*) $\frac{11}{4165}$ (*ii*) $\frac{468}{20825}$]

17. Three light bulbs are selected at random from 20 bulbs of which 5 are defective. What is the probability that (*i*) none of the bulbs is defective and (*ii*) exactly one is defective. **[Ans.** (*i*) 91/228, (*ii*) 105/228]

18. Six cards are drawn at random from a pack of cards. What is the probability that 3 will be red and 3 black? **[Ans.** $\frac{13000}{39151}$]

THEOREMS OF PROBABILITY

There are mainly three theorems of probability which are given below:

(1) Addition Theorem

(2) Multiplication Theorem

(3) Bayes' Theorem

Let us discuss them in detail.

(1) Addition Theorem

Addition theorem of probability is studied under two headings:

Addition Theorem for Mutually Exclusive Events

Addition theorem states that if A and B are two mutually exclusive events, then the probability of occurrence of either A or B is the sum of the individual probabilities of A and B. Symbolically,

$$\mathbf{P(A \text{ or } B) = P(A) + P(B)}$$

Or

$$\mathbf{P(A + B) = P(A) + P(B)}$$

Proof of the Theorem: Let n be the total number of exhaustive and equally likely cases of an experiment. Further let m_1 and m_2 be the number of cases favourable to the happening of the event A and B respectively. Then,

$$P(A) = \frac{m_1}{n}$$

$$P(B) = \frac{m_2}{n}$$

Since, the events A and B are mutually exclusive, the total number of ways in which event A or B can happen is $m_1 + m_2$, then

$$P(A \text{ or } B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B)$$

$\therefore$ $$\mathbf{P(A \text{ or } B) = P(A) + P(B)}$$

Hence, the theorem is proved.

Generalisation

The theorem can be extended to three or more mutually exclusive events. If A, B and C are three mutually exclusive events, then

$$\mathbf{P(A + B + C) = P(A) + P(B) + P(C)}$$

The following examples would illustrate the applications of addition theorem.

Example 22. A card is drawn from a pack of 52 cards. What is the probability of getting either a king or queen?

Solution: There are 4 kings and 4 queens in a pack of 52 cards.

The probability of drawing a king card is $P(K) = \frac{4}{52}$

and the probability of drawing a queen card is $P(Q) = \frac{4}{52}$

Since, both the events are mutually exclusive, the probability that the card drawn either a king or queen is

$$P(K \text{ or } Q) = P(K) + P(Q)$$
$$= \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$$

Example 23. A perfect die is tossed. What is the probability of throwing 3 or 5?

Solution: There are 6 possible outcomes.

The probability of throwing 3 is $P(A) = \frac{1}{6}$

The probability of throwing 5 is $P(B) = \frac{1}{6}$

Since, both the events are mutually exclusive, so the probability of throwing either 3 or 5 is:

$$P(A \text{ or } B) = P(A) + P(B)$$
$$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

Example 24. An investment consultant predicts that the odds against the price of a certain stock will go up during the next week are 2 : 1 and the odds in favour of the price remaining the same are 1 : 3. What is the probability that the price of the stock will go down during the next week?

Solution: Let A denote the event 'stock price will go up', and B be the event 'stock price will remain same'.

Then $P(A)=\frac{1}{3}$, and $P(B)=\frac{1}{4}$

$\therefore$ P (stock price will either go up or remain same) $= P(A\cup B)$

$$= P(A)+P(B)=\frac{1}{3}+\frac{1}{4}=\frac{7}{12}$$

Now, P (stock price will go down) $= 1-P(A\cup B)$

$$=1-\frac{7}{12}=\frac{5}{12}$$

Example 25. Three horses A, B and C are in a race, A is twice as likely to win as B and B is twice as likely to win as C. What are the respective probabilities of winning ?

Solution: Here, $P(B)=2P(C),\ P(A)=2P(B)$

Since, A, B and C are mutually exclusive and exhaustive events,

$$P(A+B+C)=P(A)+P(B)+P(C)=1$$

Or $2P(B)+P(B)+\frac{1}{2}P(B)=1$ $\therefore$ $P(B)=\frac{2}{7}$

Hence, $P(A)=\frac{4}{7}$ and $P(C)=\frac{1}{7}$

Example 26. Among 3 events A, B and C only one event can take place. The odds against A are 3 : 2, against B are 4 : 3. Find the odds against event C.

Solution: Probability of happening of A event is $P(A)=\frac{2}{5}$

Probability of happening of B event is $P(B)=\frac{3}{7}$

Since, events are mutually exclusive,

$$P(A \text{ or } B \text{ or } C)=P(A)+P(B)+P(C)=1$$

By substituting values,

$$1=\frac{2}{5}+\frac{3}{7}+P(C)$$

$$1-\left(\frac{2}{5}+\frac{3}{7}\right)=P(C)$$

$\Rightarrow$ $P(C)=\frac{6}{35}$

$\frac{6}{35}$ probability implies 6 in favour out of 35 chances.

Thus, odds against event C are $35-6=29$

Thus, odds against event C are $= 29 : 6$

Example 27. A card is drawn at random from a pack of cards. Find the probability that the drawn card is either a club or an ace of diamond.

Solution: The probability of drawing a card of club $P(A)=\frac{13}{52}$

The probability of drawing an ace of diamond $P(B)=\frac{1}{52}$

Since the events are mutually exclusive, the probability of the drawn card being a club or an ace of diamond is:

$$P(A \text{ or } B)=P(A)+P(B)$$

$$=\frac{13}{52}+\frac{1}{52}=\frac{14}{52}=\frac{7}{26}$$

Example 28. A bag contains 30 balls numbered from 1 to 30. One ball is drawn at random. Find the probability that the number of the ball is a multiple of 5 or 8.

Solution: The probability of the number being a multiple of 5 within 30 is given by

(5, 10, 15, 20, 25, 30) is $P(A)=\frac{6}{30}$

The probability of the number being a multiple of 8 within 30 is given by

(8, 16, 24) is $P(B)=\frac{3}{30}$

Since the events are mutually exclusive, the probability that the ball drawn bears a number which is a multiple of 5 or 8 is:

$$P(A \text{ or } B)=P(A)+P(B)$$

$$=\frac{6}{30}+\frac{3}{30}=\frac{9}{30}$$

Example 29. In a single throw of 2 dice, determine the probability of getting a total 7 or 9.

Solution: In a throw of 2 dice, there are $6\times 6=36$ possible outcomes as follows:

(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)

$6\times 6=36$

A total of 7 can come in the following 6 ways:-

(6, 1) (5, 2) (4, 3) (3, 4) (2, 5) (1, 6)

A total of 9 can come in the following 4 ways:-

(6, 3) (5, 4) (4, 5) (3, 6)

$\therefore$ The probability of getting a total of 7 is $P(A) = \frac{6}{36}$

The probability of getting a total of 9 is $P(B) = \frac{4}{36}$

Since the events are mutually exclusive, the probability of getting either a total of 7 or 9 is:

$$P(A \text{ or } B) = P(A) + P(B)$$
$$= \frac{6}{36} + \frac{4}{36} = \frac{10}{36} = \frac{5}{18}$$

Example 30. In a given race the odds in favour of three horses A, B and C are 1 : 3, 2 : 3 and 2 : 5 respectively. Assuming that a dead heat (in which all the three horses win) is impossible, find the chance that one of them will win the race.

Solution: Odds in favour of horse A = 1 : 3

$\therefore$ The probability that A wins $P(A) = \frac{1}{1+3} = \frac{1}{4}$

Odds in favour of horse $B = 2 : 3$

Similarly, the probability that B wins $P(B) = \frac{2}{2+3} = \frac{2}{5}$

Odds in favour of horse $C = 2 : 5$

The probability that C wins $P(C) = \frac{2}{2+5} = \frac{2}{7}$

Since the dead heat is impossible, therefore the events are mutually exclusive.

The probability that one of them wins the race is:

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$$
$$= \frac{1}{4} + \frac{2}{5} + \frac{2}{7} = \frac{131}{140}$$

Example 31. An urn contains 4 red, 5 black, 3 yellow and 11 green balls. A ball is drawn at random. Find the probability that it is (*i*) either red, black or a yellow ball (*ii*) either a red, black, yellow or green.

Solution: Total number of balls are: 4R + 5B + 3Y + 11G = 23

Probability of getting a red ball $P(A) = \frac{4}{23}$

Probability of getting a black ball $P(B) = \frac{5}{23}$

Probability of getting a yellow ball $P(C) = \frac{3}{23}$

Probability of getting a green ball $P(D) = \frac{11}{23}$

(*i*) Since the events are mutually exclusive, the probability of the drawn ball being R, B and Y is

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$$
$$= \frac{4}{23} + \frac{5}{23} + \frac{3}{23} = \frac{12}{23}$$

(*ii*) The probability of the drawn ball being R, B, Y or G is

$$P(A \text{ or } B \text{ or } C \text{ or } D) = P(A) + P(B) + P(C) + P(D)$$
$$= \frac{4}{23} + \frac{5}{23} + \frac{3}{23} + \frac{11}{23} = \frac{23}{23} = 1$$

Example 32. If a pair of dice is thrown, find the probability that (*i*) the sum is neither 7 nor 11. (*ii*) neither a doublet nor a total of 9 will appear.

Solution: There are 36 possible outcomes, we write them as follows:

(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)

$6 \times 6 = 36$

(*i*) A total of 7 can come in 6 ways:

(6, 1) (5, 2) (4, 3) (3, 4) (2, 5) (1, 6)

A total of 11 can come in 2 ways:

(6, 5) and (5, 6)

The probability of getting a total of 7 is $P(A) = \frac{6}{36}$

The probability of getting a total of 11 is $P(B) = \frac{2}{36}$

Since, the events are mutually exclusive, the probability of getting either 7 or 11 is

$$P(A \text{ or } B) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9}$$

$\therefore$ The probability that the sum is neither 7 or 11 is

$$P\text{ (neither 7 or 11)} = 1 - P\text{ (either 7 or 11)}$$
$$= 1 - \frac{2}{9} = \frac{7}{9}$$

(ii) A doublet can come in 6 ways:

(1, 1) (2, 2) (3, 3) (4, 4) (5, 5) (6, 6)

A total of 9 can come in 4 ways:

(6, 3) (5, 4) (4, 5) (3, 6)

The probability of getting a doublet $P(A) = \frac{6}{36}$

The probability of getting a total of 9 is $P(B) = \frac{4}{36}$

Since, the events are mutually exclusive, the probability of getting a doublet or a total of 9 is:

$$P(A \text{ or } B) = \frac{6}{36} + \frac{4}{36} = \frac{10}{36} = \frac{5}{18}$$

∴ The probability that neither a doublet nor a total of 9 will appear is:

$$P(\text{Neither a doublet nor } 11) = 1 - \frac{5}{18} = \frac{13}{18}$$

Example 33. There are 11 red and 14 white balls in a bag. Two balls are drawn. Find the probability that both of them are of the same colour.

Solution: Total number of ways in which 2 balls can be drawn out of 25 balls = ${}^{25}C_2$

Total number of ways in which 2 red balls can be drawn out of 11 red balls = ${}^{11}C_2$

Total number of ways in which 2 white balls can be drawn out of 14 white balls = ${}^{14}C_2$

There are two cases:

(i) Both balls are red,

$$\text{Probability of getting two red balls} = \frac{{}^{11}C_2}{{}^{25}C_2}$$

(ii) Both balls are white,

$$\text{Probability of getting two white balls} = \frac{{}^{14}C_2}{{}^{25}C_2}$$

Since the (i) and (ii) cases are mutually exclusive, therefore,

$$\begin{aligned} P(\text{Both balls are of the same colour}) &= P(\text{Either 2R or 2W}) \\ &= P(2R) + P(2W) \\ &= \frac{{}^{11}C_2}{{}^{25}C_2} + \frac{{}^{14}C_2}{{}^{25}C_2} \\ &= \frac{11}{60} + \frac{91}{300} = \frac{55+91}{300} = \frac{146}{300} = \frac{73}{150} \end{aligned}$$

Example 34. A and B are mutually exclusive events for which $P(A) = 0.3$, $P(B) = p$ and $P(A + B) = 0.5$. Find the value of p.

Solution: Since, A and B are mutually exclusive events, then

$$P(A + B) = P(A) + P(B)$$

Substituting the values, we get

$$0.5 = 0.3 + p$$

$$\Rightarrow \quad p = 0.5 - 0.3 = 0.2$$

EXERCISE 7.2

1. A card is drawn from a pack of 52 cards. What is the probability of getting either a heart or queen of spade? **[Ans. 14/52]**
2. There are three events A, B, C; one and only one of which must occur. The odds are 8 to 3 against A and 5 : 2 against B. What are the odds against C? **[Ans. 43 : 34]**
3. A, B and C are bidding for a contract. It is believed that A has exactly half the chance that B has, B in turn, is 4/5th as likely as C to win the contract. What is the probability for each to win the contract? **[Ans. 2/11, 4/11, 5/11]**
4. In a class of 25 students with role numbers 1 to 25, a student is picked up at random to answer the question. Find the probability that roll number of the selected student is either a multiple 5 or 7. **[Ans. 8/25]**
5. A bag contains 3 red, 6 white, 4 blue and 7 yellow balls. A ball is drawn. What is the probability that the ball will be either white or yellow? **[Ans. 13/20]**
6. In a given race the odds in favour of three horses A, B and C are 1:4, 2:5 and 3:6 respectively. Assuming a dead heat is impossible, find the chance that one of them will win the race. **[Ans. 86/105]**
7. In a single throw with two dice, find the chance of throwing at least 8 (or more than 7). **[Ans. 5/12]**
8. Set up a sample space for the single toss of a pair of fair dice. From the sample space, determine the probability that the sum is either 7 or 11. **[Ans. 2/9]**
9. In a single throw of three dice, find the probability of getting a total of 17 or 18. **[Ans. 1/54]**
10. In a single throw of 2 dice, find the probability of obtaining a total 9 or 11. **[Ans. 1/6]**

Addition Theorem for Not Mutually Exclusive Events

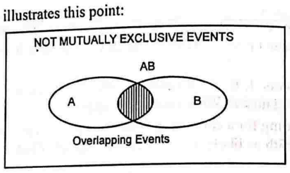
The addition theorem discussed above is not applicable when the events are not mutually exclusive. For example, if the probability of drawing a card of spade is 13/52 and that of drawing a card of king is 4/52, we cannot calculate the probability of drawing a card of either spade or king by adding the two probabilities because the events are not mutually exclusive. The card could very well be a spade card as well as king. When the events are not mutually exclusive, the addition theorem has to be modified.

Modified Addition Theorem states that if A and B are not mutually exclusive events, the probability of the occurrence of either A or B or both is equal to the probability that event A occurs, plus the probability that event B occurs minus the probability that events common to both A and B occur. Symbolically,

$$\boxed{P(A \text{ or } B \text{ or Both}) = P(A) + P(B) - P(AB)}$$

In this formula, we subtract P(A and B), namely the probability of the events which are counted twice from P(A) + P(B). The theorem is thus modified in such a way as to render A and B mutually exclusive.

The following figure illustrates this point:



Generalisation: The theorem can be extended to three or more events. If A, B and C are not mutually exclusive events, then

$$\boxed{P(\text{Either A or B or C}) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)}$$

The following examples illustrate the application of the modified addition theorem:

Example 35. A card is drawn at random from a well shuffled pack of cards. What is the probability that it is either a spade or a king?

Solution: The probability of drawing a spade $P(A) = \frac{13}{52}$

The probability of drawing a king $P(B) = \frac{4}{52}$

Because one of the kings can belong to spade, therefore the events are not mutually exclusive.

The probability of drawing a king of spade $P(AB) = \frac{1}{52}$

So, the probability of drawing a spade or king is:

$$P(A \text{ or } B \text{ or Both}) = P(A) + P(B) - P(AB)$$

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$$

$$= \frac{16}{52} = \frac{4}{13}$$

Example 36. A bag contains 30 balls numbered from 1 to 30. One ball is drawn at random. Find the probability that the number of ball is a multiple of 5 or 6.

Solution: The probability of the ball being multiple of 5 is:

(5, 10, 15, 20, 25, 30); $P(A) = \frac{6}{30}$

The probability of the ball being multiple of 6 is

(6, 12, 18, 24, 30); $P(B) = \frac{5}{30}$

Since, 30 is a multiple of 5 as well as 6, therefore the events are not mutually exclusive.

$P(A \text{ and } B) = \frac{1}{30}$ (common multiple 30)

So, the probability of getting a ball being multiple of 5 or 6 is:

$$P(A \text{ or } B) = P(A) + P(B) - P(AB)$$

$$= \frac{6}{30} + \frac{5}{30} - \frac{1}{30} = \frac{10}{30} = \frac{1}{3}$$

Example 37. One number is drawn from numbers 1 to 150. Find the probability that it is either divisible by 3 or 5.

Solution: The probability that the number being divisible by 3 is

(3, 6, 9, ... 147, 150) $P(A) = \frac{50}{150}$

The probability that the number being divisible by 5 is

(5, 10, 15, ... 145, 150) $P(B) = \frac{30}{150}$

Since, the numbers (15, 30, 45.........135, 150) = 10 are common to both, therefore, the events are not mutually exclusive.

$\therefore$ $P(A \text{ and } B) = \frac{10}{150}$ (common multiple)

So, the probability of getting either divisible by 3 or 5 is:

$$P(A \text{ or } B) = P(A) + P(B) - P(AB)$$

$$= \frac{50}{150} + \frac{30}{150} - \frac{10}{150} = \frac{70}{150} = \frac{7}{15}$$

Example 38. A number was drawn at random from the number 1 to 50. What is the probability that it will be a multiple of 2 or 3 or 10.

Solution: Probability of getting a multiple of 2 :

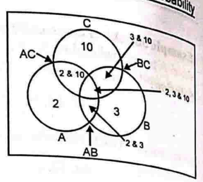
$$P(A) = \frac{25}{50}$$

Probability of getting a multiple of 3 :

$$P(B) = \frac{16}{50}$$

Probability of getting a multiple of 10 :

$$P(C) = \frac{5}{50}$$



Common Probability of getting a multiple of 2 and 3

$$P(A \text{ and } B) = \frac{8}{50} \quad \left[\begin{array}{l}\text{Common multiple of 2 and 3} \\ = 6, 12, 18, 24, 30, 36, 42, 48 = 8\end{array}\right]$$

Common Probability of getting a multiple of 3 and 10

$$P(B \text{ and } C) = \frac{1}{50} \quad \left[\begin{array}{l}\text{Common multiple of 3 and 10} \\ = 30 = 1\end{array}\right]$$

Common Probability of getting a multiple of 2 and 10

$$P(A \text{ and } C) = \frac{5}{50} \quad \left[\begin{array}{l}\text{Common multiple of 2 and 10} \\ = 10, 20, 30, 40, 50 = 5\end{array}\right]$$

Common Probability of getting a multiple of 2, 3 and 10

$$P(A \text{ and } B \text{ and } C) = \frac{1}{50} \quad \left[\begin{array}{l}\text{Common multiple of 2,3 and 10} \\ = 30 = 1\end{array}\right]$$

Probability that it is a multiple of 2 or 3 or 10 =

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) - P(A \text{ and } B) - P(B \text{ and } C) - P(A \text{ and } C) + P(A \text{ and } B \text{ and } C)$$

$$= \frac{25}{50} + \frac{16}{50} + \frac{5}{50} - \frac{8}{50} - \frac{1}{50} - \frac{5}{50} + \frac{1}{50} = \frac{33}{50}$$

Example 39. A card is drawn at random from standard pack of cards. What is the probability that (*i*) it is either a king or queen, (*ii*) it is either a king or a black card?

Solution: (*i*) The probability of drawing a king card $P(K) = \frac{4}{52}$

The probability of drawing a queen card $P(Q) = \frac{4}{52}$

Since, both the events are mutually exclusive, the probability that the card drawn is either a king or queen is

$$P(K \text{ or } Q) = P(K) + P(Q)$$

$$= \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{4}{26} = \frac{2}{13}$$

(*ii*) The probability of drawing a king card $P(K) = \frac{4}{52}$

The probability of drawing a black card $P(B) = \frac{26}{52}$

Since, black kings are common to both, the events are not mutually exclusive.

$$\therefore \quad P(\text{Black Kings}) = \frac{2}{52}$$

Thus, the probability that card drawn is either a king or a black card is

$$P(\text{a king or black}) = P(\text{a king}) + P(\text{a black card}) - P(\text{a black king})$$

$$= \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52}$$

Example 40. A chartered accountant applied for a job in two firms X and Y. He estimated that the probability of his being selected in a firm X is $\frac{7}{10}$ and being rejected in Y is $\frac{5}{10}$ and the probability that he will be selected in both the firm is $\frac{4}{10}$. What is the probability that he will be selected in one of the firms?

Solution: P (Chartered accountant is selcted in firm X) $= \frac{7}{10}$

P (he is selected in firm Y) = 1 – P (he is being rejected in firm Y)

$$= 1 - \frac{5}{10} = \frac{5}{10}.$$

P (he is selected in both X and Y firms) $= \frac{4}{10}$

P (he will be selected in one of the firm) = P(X) + P(Y) – P(X and Y)

$$= \frac{7}{10} + \frac{5}{10} - \frac{4}{10} = \frac{8}{10} = \frac{4}{5}$$

Example 41. A card is drawn out of a pack of cards. Find the probability that the card is an ace, a king, a queen or a card of clubs.

Solution: The probability of drawing a card of ace, a king and a queen = $P(A) = \frac{12}{52}$

The probability of drawing a card of club = $P(B) = \frac{13}{52}$

Because the cards of ace, king and a queen can belong to club, therefore the events are not mutually exclusive.

The probability of drawing a card of an ace, a king and a queen of clubs

$$= P(AB) = \frac{3}{52}$$

So, the probability of drawing a card of an ace, a king, a queen or a card of club is

$$P(\text{A or B or both}) = P(A) + P(B) - P(AB)$$

$$= \frac{12}{52} + \frac{13}{52} - \frac{3}{52} = \frac{22}{52} = \frac{11}{26}$$

EXERCISE 7.3

1. What is the probability of drawing a heart or a king card from a pack of cards? [Ans. 4/13]
2. A bag contains 50 balls numbered from 1 to 50. One ball is drawn at random. Find the probability that a drawn ball is a multiple of 5 or 7. [Ans. 8/25]
3. A card is drawn at random from a standard pack of cards. What is the probability that (*i*) it is either a king or queen (*ii*) it is a king or a red card. [Ans. (*i*) 2/13 (*ii*) 28/52]
4. A bag contains 30 balls numbered from 1 to 30. One ball is drawn at random. Find the probability that the number of drawn ball will be a multiple of (*i*) 5 or 9 (*ii*) 3 or 5. [Ans. (*i*) 3/10 (*ii*) 14/30]
5. A bag contains 50 balls numbered from 1 to 50. One ball is drawn at random. Find the probability that it is a multiple of 2 or 3 or 10. [Ans. 33/50]
6. A book contains 100 pages numbered from 1 to 100. A page is opened at random and is selected, find the probability that the opened page is a multiple of 6 or 10. [Ans. 23/100]
7. A card is drawn out of a pack of cards. Find the probability that the card is a club or an honour card? (Cards of ace, king and queen are the honour cards). [Ans. 11/26]
8. A card is drawn out of a pack of cards. Find the probability that the card is a spade or a face card? (Cards of king, queen and Jackel are the face cards). [Ans. 11/26]
9. A die is thrown. Find the probability of getting a number which is a multiple of 2 or 3. [Ans. 2/3]
10. What is the probability that a leap year selected at random will contain (*i*) 53 Sundays (*ii*) either 53 Sundays or 53 Mondays (*iii*) either 53 Sundays or 53 Fridays? [Ans. (*i*) 2/7 (*ii*) 3/7 (*iii*) 4/7]
11. Probability that an electric bulb will last for 150 days or more is 0.7 and that it will last at the most 160 days is 0.8. Find the probability that it will last between 150 to 160 days? [Ans. 0.5]
[**Hint:** $0.7 + 0.8 - P(AB) = 1$]
12. There are 10 boys and 20 girls in a class, in which half boys and half girls have blue eyes. One representative is selected at random from the class. What is the probability that he is a boy or his eyes are blue colour? [Ans. 2/3]
[**Hint:** See Example 127]
13. The probability that a contractor will get a plumbing contract is 2/3 and probability that he will not get an electric contract is 5/9. If the probability of getting at least one contract is 4/5, what is probability that he will get both? [Ans. 14/15]
14. A student applies for a job in two firms X and Y. The probability of his being selected in a firm X is 0.7 and being rejected in the firm Y is 0.5. The probability that his application is being rejected in both the firms is 0.6. What is the probability that he will selected in one of the firms? [Ans. 0.8]

15. The result of an examination given to a class on 3 papers A, B and C are given. It is estimated that 40% failed in paper A, 30% failed in paper B, 25% failed in paper C, 15% failed in paper A and B both, 12% failed in paper B and C both, 10% failed in paper A and C both and 3% failed in all the papers. What is the probability of a randomly selected candidate passing in at least one of three papers.
[**Hint:** See Example 126] [Ans. 0.39]

(2) Multiplication Theorem

Multiplication theorem of probability is studied under two headings:

Multiplication Theorem for Independent Events

Multiplication theorem states that if A and B are two independent events, then the probability of the simultaneous occurrence of A and B is equal to the product of their individual probabilities. Symbolically,

$$\boxed{P(AB) = P(A) \times P(B)}$$

Proof of the Theorem: Let m_1 be the number of cases favourable to the happening of the event A out of n_1 exhaustive and equally likely cases.

$$\therefore \quad P(A) = \frac{m_1}{n_1}$$

Let m_2 be the number of cases favourable to the happening of the event B out of n_2 exhaustive and equally likely cases.

$$\therefore \quad P(B) = \frac{m_2}{n_2}$$

Now, by the Fundamental Principle of counting, the number of cases favourable to the happening of the event AB is $m_1 m_2$ out of $n_1 n_2$

$$\therefore \quad P(AB) = \frac{m_1 m_2}{n_1 n_2} = \left(\frac{m_1}{n_1}\right) \cdot \left(\frac{m_2}{n_2}\right)$$

$$= P(A) \, . \, P(B)$$

$$\therefore \quad P(AB) = P(A) . P(B)$$

Hence the theorem is proved.

Generalisation

The theorem can be extended to three or more independent events. If A, B and C are three independent events, then

$$\boxed{P(ABC) = P(A) \times P(B) \times P(C)}$$

The following examples illustrate the application of the multiplication theorem.

Example 42. A coin is tossed 3 times. What is the probability of getting all the 3 heads?

Solution: Probability of head in the first toss $P(A) = \frac{1}{2}$

Probability of head on the second toss $P(B) = \frac{1}{2}$

Probability of head on the third toss $P(C) = \frac{1}{2}$

Since, the events are independent, the probability of getting all heads in three tosses is:

$$P(ABC) = P(A) \times P(B) \times P(C)$$
$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

Example 43. From a pack of 52 cards, two cards are drawn at random one after the another with replacement. What is the probability that both cards are kings?

Solution: The probability of drawing a king $P(A) = \frac{4}{52}$

The probability of drawing again a king after replacement $P(B) = \frac{4}{52}$

Since, the events are independent, the probability of drawing two kings is:

$$P(AB) = P(A) \times P(B)$$
$$= \frac{4}{52} \times \frac{4}{52} = \frac{1}{169}$$

Example 44. A man wants to marry a girl having qualities: white complexion—the probability of getting such a girl is one in twenty; handsome dowry - the probability of getting this one in fifty; westernised manners and etiquettes - the probability here is one in hundred. Find out the probability of his getting married to such a girl when the possession of these attributes is independent.

Solution: Probability of getting a girl with white complexion $= P(A) = \frac{1}{20} = 0.05$

Probability of getting a girl with handsome dowry $= P(B) = \frac{1}{50} = 0.02$

Probability of getting a girl with westernised manner $= P(C) = \frac{1}{100} = 0.01$

Since, the events are independent, the probability of the simultaneous occurrence of all these qualities is:

$$P(ABC) = P(A) \times P(B) \times P(C)$$
$$= \frac{1}{20} \times \frac{1}{50} \times \frac{1}{100}$$
$$= 0.05 \times 0.02 \times 0.01$$
$$= 0.00001$$

Example 45. A class consists of 100 students. Out of these 25 are girls and 75 are boys. 10 of them are rich and remaining poor. 20 of them are fair complexioned. What is the probability of selecting a fair complexioned rich girl?

Solution: Probability of selecting a girl student $= P(A) = \frac{25}{100} = \frac{1}{4}$

Probability of selecting a rich student $= P(B) = \frac{10}{100} = \frac{1}{10}$

Probability of selecting a fair complexioned student $= P(C) = \frac{20}{100} = \frac{1}{5}$

Now, ABC is the event of selecting a rich, fair complexioned girl student. The events A, B, and C are independent.

The probability of selecting a fair complexioned rich girl is:

$$P(ABC) = P(A) \times P(B) \times P(C)$$
$$= \frac{1}{4} \times \frac{1}{10} \times \frac{1}{5} = \frac{1}{200} = 0.005$$

Example 46. A bag containing 5 white and 3 black balls. Two balls are drawn at random one after another with replacement. Find the probability that both the balls drawn are black.

Solution: Probability of drawing black ball in the first draw $= P(A) = \frac{3}{8}$

Probability of drawing black ball in the second draw $= P(B) = \frac{3}{8}$

Since, the events are independent, the probability that both the balls are black:

$\therefore$ $P(2\text{ Black}) = P(1\text{st Black}) \times P(2\text{nd Black})$
$$= \frac{3}{8} \times \frac{3}{8} = \frac{9}{64}$$

Example 47. A bag contains 4 red balls, 3 white balls and 5 black balls. Two balls are drawn one after the other with replacement. Find the probability that first is red and the second is black.

Solution: Probability of a red ball in the first draw $= \frac{4}{12}$

The probability of a black ball in the second draw $= \frac{5}{12}$

Since, the events are independent, the probability that first is red and the second is black will be:

$$= P(1R).\ P(1B)$$
$$= \frac{4}{12} \times \frac{5}{12} = \frac{20}{144} = \frac{10}{72} = \frac{5}{36}$$

Example 48. Suppose that it is 11 to 5 against a person A who is now 38 years of age living till he is 73 and 5 to 3 against a person B now 43 living till he is 78 years. Find the probability that both will die 35 years, hence

Solution: The probability that A will die within 35 years is:

$$P(A)=\frac{11}{11+5}=\frac{11}{16}$$

The probability that B will die within 35 years is:

$$P(B)=\frac{5}{5+3}=\frac{5}{8}$$

Since, the events are independent, the probability that both will die is

$$P(AB)=P(A).P(B)$$

$$=\frac{11}{16}\times\frac{5}{8}=\frac{55}{128}$$

Example 49. An electronic device is made of three components A, B and C. The probability of failure of the component A is 0.01, that of B is 0.02 and that of C is 0.05 in some fixed period of time. Find the probability that the device will work satisfactorily during the period of time assuming that the three components work independently of one another.

Solution: Let the three failure components are denoted by $\overline{A}$, $\overline{B}$ and $\overline{C}$ respectively.

$P(\overline{A})=0.01, P(\overline{B})=0.02, P(\overline{C})=0.05$

∴ Probability that these components do not fail

$$P(A)=1-P(\overline{A})=1-0.01=0.99$$

$$P(A)=1-P(\overline{B})=1-0.02=0.98$$

$$P(C)=1-P(\overline{C})=1-0.05=0.95$$

Probability that the device will work satisfactorily is

$$P(ABC)=P(A).\ P(B).\ P(C)$$

$$=0.99\times0.98\times0.95$$

$$=0.92169=0.92\text{ (approx.)}$$

EXERCISE 7.4

1. Find the probability of getting 3 tails in 3 tosses of a coin. **[Ans. 1/8]**
2. Three aeroplanes fly from Bombay to London. Odds in favour of their arriving safely are 2:1, 3:1 and 4:1. Find the probability that they all arrive safely. **[Ans. 2/5]**
3. Two balls are drawn one after the other at random with replacement from an urn containing 4 red, 3 black and 5 white balls. Find the following probabilities, (*i*) E_1, both are red (*ii*) E_2, first is red and second is black. (*iii*) E_3, the first is black and the second is white. **[Ans. (*i*) $\frac{1}{9}$, (*ii*) $\frac{1}{12}$, (*iii*) $\frac{5}{48}$]**

4. A husband and a wife appear in an interview for 2 vacancies for the same post. The probability of selection of husband is 4/5 and that of wife is 3/4. Find the probability that (*i*) both of them will be selected (*ii*) none of them will be selected and (*iii*) only wife will be selected. **[Ans. (*i*) 3/5 (*ii*) 1/20 (*iii*) 3/20]**
5. The odds in favour of passing driving test by Mohan is 3:5 and odds in favour of passing the same test by Ram is 3:2. What is the probability that both will pass the test? **[Ans. 9/40]**
6. A university has to appoint examiners to evaluate paper in Statistics. Out of a panel of 40 examiners, 10 are women, 30 of them knowing Hindi and 5 of them are Ph.D. Find the probability of selecting a Hindi knowing Ph.D. women teacher to evaluate the papers. **[Ans. 3/128]**

o Probability of happening of at least one event in case of *n* independent events

If we are given n independent events $A_1, A_2, A_3 \dots A_n$ with respective probabilities of happening as $p_1, p_2, p_3 \cdots p_n$, then the probability of happening of at least one of independent events $A_1, A_2, A_3 \dots A_n$ is given by:

P (happening of at least one of the events) = 1 – P (happening of none of the events)

$$= 1-[(1-p_1).(1-p_2).(1-p_3).(1-p_n)]$$

The following examples illustrate the applications of this theorem:

Example 50. A problem in statistics is given to three students A, B and C whose chances of solving it are 1/2, 1/3 and 1/4. What is the probability that the problem will be solved?

Solution: Probability that A will solve the problem = $P(A)=\frac{1}{2}$

Probability that B will solve the problem = $P(B)=\frac{1}{3}$

Probability that C will solve the problem = $P(C)=\frac{1}{4}$

∴ Probability that A will not solve the problem = $P(\overline{A})=1-\frac{1}{2}=\frac{1}{2}$

Probability that B will not solve the problem = $P(\overline{B})=1-\frac{1}{3}=\frac{2}{3}$

Probability that C will not solve the problem = $P(\overline{C})=1-\frac{1}{4}=\frac{3}{4}$

Since, all the events are independent, so

∴ P (that none will solve the problem) = $P(\overline{A}).P(\overline{B}).P(\overline{C})$

$$=\frac{1}{2}\times\frac{2}{3}\times\frac{3}{4}=\frac{1}{4}$$

∴ P (that problem will be solved) = 1 – P (that none will solve)

$$=1-\frac{1}{4}=\frac{3}{4}$$

Example 51. A candidate (Mr. X) is interviewed for 3 posts. For the first post, there are 3 candidates, for the second post, there are 4 and for third there are 2. What are the chances of Mr. X being getting selected?

Solution: Probability of selection for 1st post = $P(A) = \frac{1}{3}$

Probability of selection for 2nd post = $P(B) = \frac{1}{4}$

Probability of selection for 3rd post = $P(C) = \frac{1}{2}$

$\therefore$ Probability of not selecting on 1st post = $P(\bar{A}) = 1 - \frac{1}{3} = \frac{2}{3}$

Probability of not selecting on 2nd post = $P(\bar{B}) = 1 - \frac{1}{4} = \frac{3}{4}$

Probability of not selecting on 3rd post = $P(\bar{C}) = 1 - \frac{1}{2} = \frac{1}{2}$

Since, the events are independent, the probability that Mr. X is not selected for three posts is:

$$P(\bar{A}\ \bar{B}\ \bar{C}) = P(\bar{A}).P(\bar{B}).P(\bar{C})$$
$$= \frac{2}{3} \times \frac{3}{4} \times \frac{1}{2} = \frac{1}{4}$$

$\therefore$ Probability of selection for at least 1 post

$= 1 - P$ (not selected at all)

$$= 1 - \frac{1}{4} = \frac{3}{4}$$

Example 52. It is 9:7 against person A who is now 40 years of age living till he is 60 and 3:2 against person B now 50 years living till he is 70. Find the probability that at least one of them will be alive 20 years hence.

Solution: The probability that person A is not alive 20 years hence = $P(\bar{A}) = \frac{9}{16}$

The probability that person B is not alive 20 years hence = $P(\bar{B}) = \frac{3}{5}$

Since, the events are independent, the probability that both the persons are not alive 20 years hence is:

$$P(\bar{A}) \times P(\bar{B}) = \frac{9}{16} \times \frac{3}{5} = \frac{27}{80}$$

The probability that, at least one of them will be alive 20 years hence is:

$$= 1 - \frac{27}{80} = \frac{53}{80}$$

Example 53. A person is known to hit the target in 3 out of 4 shots whereas another person is known to hit he target in 2 out of 3 shots. Find the probability of the target being hit at all when they both try.

Solution: Probability of hitting the target by A is $P(A) = \frac{3}{4}$

Probability of hitting the target by B is $P(B) = \frac{2}{3}$

Probability that target is not hit by A is $P(\bar{A}) = 1 - \frac{3}{4} = \frac{1}{4}$

Probability that target is not hit by B is $P(\bar{B}) = 1 - \frac{2}{3} = \frac{1}{3}$

Since, the events are independent, the probability of not hitting by both A and B is

$$P(\bar{A}\ \bar{B}) = P(\bar{A}).P(\bar{B}) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

Hence, probability of hitting at all $= 1 - P$ (not hitting at all)

$$= 1 - \frac{1}{12} = \frac{11}{12}$$

Example 54. The chances of survival after 25 years are 3 out of 10 for a man and 4 out of 10 for a woman. Find the probability that (*i*) both will be alive after 25 years, (*ii*) at least one of them will be alive after 25 years.

Solution: (*i*) The probability that a man will survive after 25 years $P(M) = \frac{3}{10}$

The probability that a woman will survive after 25 years $P(W) = \frac{4}{10}$

Since, the events are independent, the probability that both will be alive after 25 years is

$$P(MW) = P(M) \times P(W)$$
$$= \frac{3}{10} \times \frac{4}{10} = \frac{12}{100} = \frac{3}{25}$$

(*ii*) The probability that man will not alive after 25 years

$$P(\bar{M}) = 1 - \frac{3}{10} = \frac{7}{10}$$

The probability that woman will not alive after 25 years $P(\bar{W}) = 1 - \frac{4}{10} = \frac{6}{10}$

Since, the events are independent, the probability that man and woman will not alive is

$$P(\overline{MW}) = P(\bar{M}).P(\bar{W}) = \frac{7}{10} \times \frac{6}{10}$$

$\therefore$ The probability that at least one of them will be alive is

$$=1-\frac{7}{10}\times\frac{6}{10}$$

$$=1-\frac{42}{100}=\frac{58}{100}=\frac{29}{50}$$

Example 55. Find the probability of throwing 6 at least once in six throws with a single die.

Solution: The probability of throwing 6 at least once = 1 – Probability that 6 is not thrown at all

Probability that 6 is not thrown in the 1st throw $=\frac{5}{6}$

Probability that 6 is not thrown in the 2nd throw $=\frac{5}{6}$

Probability that 6 is not thrown in the 3rd throw $=\frac{5}{6}$

Probability that 6 is not thrown in the 4th throw $=\frac{5}{6}$

Probability that 6 is not thrown in the 5th throw $=\frac{5}{6}$

Probability that 6 is not thrown in the 6th throw $=\frac{5}{6}$

Since, the events are independent, the probability that 6 is not thrown in any throw

$$=P(\bar{I}).P(\overline{II}).P(\overline{III}).P(\overline{IV}).P(\bar{V}).P(\overline{VI})$$

$$=\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}=\left(\frac{5}{6}\right)^6$$

Hence, the probability of throwing 6 at least once

$$=1-\left(\frac{5}{6}\right)^6$$

Example 56. A and B decide to meet at Hanuman Temple between 5 to 6 p.m. with the condition that no one would want wait for the other for more than 15 minutes. What is the probability that they meet?

Solution: The probability that A will meet B $=\frac{15}{60}=\frac{1}{4}$

The probability that A will not meet B $=1-\frac{1}{4}=\frac{3}{4}$

The probability that B will meet A $=\frac{15}{60}=\frac{1}{4}$

The probability that B will not meet A $=1-\frac{1}{4}=\frac{3}{4}$

Since, the events are independent, the probability that they fail to meet $=\frac{3}{4}\times\frac{3}{4}=\frac{9}{16}$

The probability that they will meet $=1-\frac{9}{16}=\frac{7}{16}$

Example 57. A husband and wife appear in an interview for two vacancies for the same post. The probability of husband's selection is 1/7 and that of wife selection is 1/5. What is the probability that:

(*i*) at least one of them will be selected.

(*ii*) both of them will be selected.

(*iii*) None of them will be selected.

Solution: Let P(H) and P(W) denote probability that husband and wife are selected respectively. Then

$$P(H)=1/7, P(W)=1/5$$

$$P(\bar{H})=1-1/7=6/7,\ P(\bar{W})=1-1/5=4/5$$

(*i*) Now the probability that at least one of them is selected:

$$=1-P(\bar{H}).P(\bar{W})=1-6/7\times4/5=11/35$$

(*ii*) The probability that both husband and wife are selected

$$=P(H).P(W)=1/7\times1/5=1/35$$

(*iii*) Probability that none of them is selected

$$=P(\bar{H}).P(\bar{W})=\frac{6}{7}\times\frac{4}{5}=\frac{24}{35}$$

Example 58. Let p be the probability that a man aged x year dies in a year. Find the probability that out of n men $A_1, A_2, A_3 \ldots A_n$ each aged x, A_1 will die and be the first to die.

Solution: The probability that a man aged x year dies in a year $=p$

The probability that a man aged x year does not die in a year $=1-p$

The probability that out of n men none dies in that year:

$$=(1-p)(1-p)(1-p)\ldots n \text{ times}=(1-p)^n$$

The probability that at least one man dies in that year

$$=1-P\text{ (none dies in that year)}$$

$$=[1-(1-p)^n]$$

Also the probability that out of n men, A_1 will die is $\frac{1}{n}$.

Thus, required probability $=\frac{1}{n}[1-(1-p)^n]$

Example 59. A and B are two independent witnesses. The probability that A will speak truth is x and the probability that B will speak the truth is y. A and B agree in a certain statement. Find the probability that the statement is true.

Solution: Given, $P(A)=x$, $P(B)=y$

$P(\bar{A})=1-x$, $P(\bar{B})=1-y$

A and B both agree when (*i*) either of them speaking the truth or (*ii*) making false statements.

$\therefore$ The probability that both A and B speaks the truth $= P(A).P(B)=xy$

$\therefore$ The probability that both A and B makes false statements

$$= P(\bar{A}).P(\bar{B})$$
$$= (1-x)(1-y)$$

Thus, the total number of cases agreeing both $= xy+(1-x)(1-y)$

$$\therefore \text{P(the statement is true)} = \frac{\text{No. of cases speaking the truth}}{\text{Total no. of cases}}$$

$$= \frac{xy}{xy+(1-x)(1-y)}$$

EXERCISE 7.5

1. A problem in Statistics is given to four students. Their chances of solving it are 1/2, 1/3, 1/4 and 1/5 respectively. What is the probability that the problem is solved? [Ans. 4/5]
2. A and B decide to meet at Durga Temple between 5 to 7 p.m. with the condition that no one would wait for the other for more than 30 minutes. What is the probability that they meet? [Ans. 7/16]
3. The probability that a boy will get a scholarship is 0.90 and that a girl will get is 0.80. What is the probability that at least one of them will get the scholarship? [Ans. 0.98]
4. A can solve 75% of the problems in Statistics and B can solve 30%. What is the probability that at least one of them will solve the problem? [Ans. 33/40]
5. A can hit a target 3 times in 5 shots, B 2 times in 5 shots and C 3 times in 4 shots. They fire a volley. What is the probability of the target being hit at all? [Ans. 47/50]
6. Find the probability of throwing 6 at least once in three tosses of a die. [Ans. 91/216]
7. A candidate is called for interview by three companies. For the first company there are 12 candidates, for the second there are 15 candidates and for the third there are 10 candidates. What is the probability of his getting selected at least at one of the companies? [Ans. 23/100]
8. Suppose it is 11 to 5 against a person who is now 38 years of age living till he is 73 and 5 to 3 against B now 43 living till he is 78, find the chance that at least one of these persons will be alive 35 years hence. [Ans. 73/128]
9. The probability that India wins a cricket test match against England is given to be 1/3. If India and England play three test matches, what is the probability that:
   (*i*) India will lose all the three matches;
   (*ii*) India will win at least one test match. [Ans. (*i*) 8/27; (*ii*) 19/27]

10. The odds against A solving a sum are 7 : 6 and odds in favour of B solving the same are 11 : 8. What is the probability that the sum is solved if both A and B try it? [Ans. 181/247]
11. Find the probability of having at least one head is 5 throws with a coin. [Ans. 31/32]
12. Three dice are thrown. What is the probability that at least one of the numbers turning up being greater than 4? [Ans. 19/27]
13. Let p be the probability that a man aged y years will die in a year. Find the probability that out of 4 men A, B, C and D each aged y, A will die in the year and will be the first to die. [Ans. $\frac{1}{4}[1-(1-p)^4]$]

Conditional Probability

The multiplication theorem discussed above is not applicable in case of dependent events. Dependent events are those in which the occurrence of one event affects the probability of other events. The probability of the event B given that A has occurred is called the conditional probability of B. It is denoted by P(B/A). Similarly, the conditional probability of A given that B has occured is denoted by P(A/B).

Definition of Conditional Probability

If A and B are two dependent events, then the conditional probability of B given A is defined and given by:

$$P(B/A) = \frac{P(AB)}{P(A)} \qquad \text{provided } P(A) > 0$$

Similarly, the conditional probability of A given B is defined and given by:

$$P(A/B) = \frac{P(AB)}{P(B)} \qquad \text{provided } P(B) > 0$$

Multiplication Theorem for Dependent Events
Or
Multiplication Theorem in Case of Conditional Probability

When the events are not independent, *i.e.*, they are dependent events, then the multiplication theorem has to be modified. **The Modified Multiplication Theorem sates that if A and B are two dependent events, then the probability of their simultaneous occurrence is equal to the probability of one event multiplied by the conditional probability of the other.** Symbolically,

$$P(AB) = P(A) . P(B/A)$$
$$\text{or } P(AB) = P(B) . P(A/B)$$

Where, P(B/A) = Conditional Probability of B given A

P(A/B) = Conditional Probability of A given B.

The following examples will illustrate the application of the modified multiplication theorem:

Example 60. A bag contains 10 white and 5 black balls. Two balls are drawn at random one after the other without replacement. Find the probability that both balls drawn are black.

Solution: The probability of drawing a black ball in the first attempt is:

$$P(A) = \frac{5}{10+5} = \frac{5}{15}$$

The probability of drawing the second black ball given that the first drawn is black and not replaced is:

$$P(B/A) = \frac{4}{10+4} = \frac{4}{14}$$

Since, the events are dependent, so the probability that both balls drawn are black is:

$$P(AB) = P(A) \,.\, P(B/A)$$

$$= \frac{5}{15} \times \frac{4}{14} = \frac{2}{21}$$

Example 61. Find the probability of drawing a king, a queen and a knave in that order from a pack of cards in three consecutive draws, the cards drawn not being replaced.

Solution: The probability of drawing a king $= P(A) = \frac{4}{52}$

The probability of drawing a queen after a king has been drawn

$$P(B/A) = \frac{4}{51}.$$

The probability of drawing a knave after a king and a queen have been drawn

$$P(C/AB) = \frac{4}{50}.$$

Since, the events are dependent, the required probability of drawing a king, a queen and ace in that order is:

$$P(ABC) = \frac{4}{52} \times \frac{4}{51} \times \frac{4}{50} = \frac{8}{16{,}575}$$

Example 62. Find the probability of drawing two kings from a pack of cards in two consecutive draws, the card drawn not being replaced.

Solution: The probability of drawing a king in the 1st draw $= P(A) = \frac{4}{52}$

The probability of drawing a king in the 2nd draw, given that the first king has already been drawn and not replaced $= P(B/A) = \frac{3}{51}$

Since, the events are dependent, so the required probability of 2 kings is:

$$= \frac{4}{52} \times \frac{3}{51} = \frac{1}{221}$$

Example 63. Four cards are drawn without replacement. What is the probability that they are all aces?

Solution: Probability of drawing an ace in the first attempt $= \frac{4}{52}$

Probability of drawing 2nd ace after the I ace has been drawn $= \frac{3}{51}$.

Probability of drawing 3rd ace after the I and II aces have been drawn $= \frac{2}{50}$.

Probability of drawing 4th ace after I, II and III aces have been drawn $= \frac{1}{49}$

Since, the events are dependent, the required probability is:

P (1st Ace × 2nd Ace × 3rd Ace × 4th Ace)

$$= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} = \frac{1}{270725}$$

Example 64. Two cards are drawn from a well shuffled pack of 52 cards. Find the probability that they are both aces if the first card is (*i*) replaced (*ii*) not replaced.

Solution: (*i*) **When the first card is replaced**

Probability of drawing 1st Ace = P(A) = 4/52

Probability of drawing 2nd Ace = P(B) = 4/52

Since, the events are independent, the required probability is:

$$P(AB) = \frac{4}{52} \times \frac{4}{52} = \frac{1}{13} \times \frac{1}{13} = \frac{1}{169}$$

(*ii*) **When the first card is not replaced**

Probability of drawing 1st Ace in the first attempt $= P(A) = \frac{4}{52}$

Probability of drawing 2nd Ace after the first ace has been drawn $= P(B/A) = \frac{3}{51}$

Since, the events are dependent, the required probability is:

$$P(AB) = P(A) \,.\, P(B/A)$$

$$= \frac{4}{52} \times \frac{3}{51} = \frac{1}{221}$$

Example 65. A bag contains 5 white and 8 red balls. Two successive drawings of 3 balls are made such that (*i*) the balls are replaced before the second trial, and (*ii*) the balls are not replaced before the second trial. Find the probability that the first drawing will give 3 white and the second 3 red balls in each case.

Solution: (*i*) **When balls are replaced**

Total balls in a bag = 8 + 5 = 13

3 balls can be drawn out of 13 balls in ${}^{13}C_3$ ways.

3 white balls can be drawn out of 5 white balls in ${}^5C_3$ ways.

Probability of 3 white balls $= P(3W) = \frac{{}^5C_3}{{}^{13}C_3}$

Since, the balls are replaced after the first draw so again there are 13 balls in the bag 3 red balls can be drawn out of 8 red balls in ${}^8C_3$ ways.

Probability of 3 red balls $= P(3R) = \frac{{}^8C_3}{{}^{13}C_3}$

Since, the events are independent, the required probability is:

$$P(3W \text{ and } 3R) = P(3W) \times P(3R)$$
$$= \frac{{}^5C_3}{{}^{13}C_3} \times \frac{{}^8C_3}{{}^{13}C_3} = \frac{10}{286} \times \frac{56}{286} = \frac{140}{20{,}449}$$

(ii) When the balls are not replaced before second draw

Total balls in a bag = 8 + 5 = 13

3 balls can be drawn out of 13 balls in ${}^{13}C_3$ ways

3 white balls can be drawn out of 5 white balls in ${}^5C_3$ ways.

The probability of drawing 3 white balls $= P(3W) = \frac{{}^5C_3}{{}^{13}C_3}$

After the first draw, balls left are 10. 3 balls can be drawn out of 10 balls in ${}^{10}C_3$ ways.

3 red balls can be drawn out of 8 balls in ${}^8C_3$ ways.

Probability of drawing 3 red balls $= \frac{{}^8C_3}{{}^{10}C_3}$

Since, both the events are dependent, the required probability is:

$$P(3W \text{ and } 3R) = \frac{{}^5C_3}{{}^{13}C_3} \times \frac{{}^8C_3}{{}^{10}C_3} = \frac{5}{143} \times \frac{7}{15} = \frac{7}{429}$$

Example 66. A bag contains 5 white and 3 red balls and four balls are successively drawn out and not replaced. What is the chance that (*i*) white and red balls appear alternatively and (*ii*) red and white balls appear alternatively?

Solution: (*i*) The probability of drawing a white ball $= \frac{5}{8}$

The probability of drawing a red ball $= \frac{3}{7}$

The probability of drawing a white ball $= \frac{4}{6}$

The probability of drawing a red ball $= \frac{2}{5}$

Since, the events are dependent, therefore the required probability is:

$$P(1W\ 1R\ 1W\ 1R) = \frac{5}{8} \times \frac{3}{7} \times \frac{4}{6} \times \frac{2}{5} = \frac{1}{14}$$

(*ii*) The probability of drawing red ball $= \frac{3}{8}$

The probability of drawing a white ball $= \frac{5}{7}$

The probability of drawing a red ball $= \frac{2}{6}$

The probability of drawing a white ball $= \frac{4}{5}$

Since, the events are dependent, the required probability is:

$$P(1R\ 1W\ 1R\ 1W) = \frac{3}{8} \times \frac{5}{7} \times \frac{2}{6} \times \frac{4}{5} = \frac{1}{14}$$

Example 67. Four cards are drawn without replacement. What is the probability that:

(*i*) They are of same suit, and

(*ii*) They are of different suits?

Solution: (*i*) Since the events are dependent, the required probability is:

$$= \frac{52}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} = \frac{44}{4165}$$

(*ii*) Since the events are dependent, the required probability is:

$$= \frac{52}{52} \times \frac{39}{51} \times \frac{26}{50} \times \frac{13}{49} = \frac{2{,}197}{20{,}825}$$

EXERCISE 7.6

1. A bag contains 6 white and 4 black balls. Two balls are drawn at random one after another without replacement. Find the probability that both drawn balls are white. **[Ans. 1/3]**
2. Find the probability of drawing a king, a knave and an ace in that order from a pack of cards in 3 consecutive draws, the cards drawn not being replaced. **[Ans. 8/16575]**
3. A bag contains 7 red, 5 white and 4 blue balls. Three balls are drawn successively. Find the probability that these are drawn, in order of red, white and blue if the drawn ball is not replaced. **[Ans. 1/24]**
4. Find the probability of drawing a king and an ace in this order from a pack of cards in two successive draws assuming that first card drawn is not replaced. **[Ans. 4/663]**
5. Two cards are drawn from a pack of cards. Find the probability that they are both queens if the first card is (*i*) replaced, (*ii*) not replaced. **[Ans. (*i*) 1/169 (*ii*) 1/221]**

6. A bag contains 10 gold and 8 silver coins. Two successive drawings of 3 coins are made such that (*i*) coins are replaced before the second drawings (*ii*) the coins are not replaced before the second drawing. In each case, find the probability that the first drawing will give 3 gold and the second 3 silver coins. **[Ans.** (*i*) 35/3468 (*ii*) 28/1547]
7. The probability that a trainee will remain with a company is 0.8. The probability that an employee earns more than Rs. 20,000 per year is 0.4. The probability that an employee who was a trainee and remained with the company or who earns more than Rs. 20,000 per year is 0.9. What is the probability that an employee earns more than Rs. 20,000 per year given that he is a trainee who stayed with the company?
[**Hint:** See Similar Example 76] **[Ans.** 3/8]
8. A box contains 8 tickets bearing the following numbers:
1, 2, 3, 4, 5, 6, 8 and 10
One ticket is drawn at random and kept side. Then a second ticket is drawn. Find the probability that both the tickets show even numbers. **[Ans.** 5/14]
9. A bag contains 5 white and 4 black balls and 4 balls are successively drawn out and not replaced. What is the chance that white and black balls appear alternatively? **[Ans.** 5/63]

Combined Use of Addition and Multiplication Theorem

Under probability, there are certain problems where both addition and multiplication theorems are to be used simultaneously. In such cases, we first apply multiplication theorem and then ultimately we apply addition theorem.

The following examples illustrate the combined use of addition and multiplication theorems.

Example 68. A speaks truth in 80% cases, B in 90% cases. In what percentage of cases are they likely to contradict each other in stating the same fact.

Solution: Let P(A) and P(B) denote the probability that A and B speak the truth. Then,

$$P(A)=\frac{80}{100}=\frac{4}{5}, \quad P(\bar{A})=1-P(A)=1-\frac{4}{5}=\frac{1}{5}$$

$$P(B)=\frac{90}{100}=\frac{9}{10}, \quad P(\bar{B})=1-P(B)=1-\frac{9}{10}=\frac{1}{10}$$

They will contradict each other only when one of them speaks the truth and the other speaks a lie.

Thus, there are two possibilities:

(*i*) A speaks the truth and B tells a lie

(*ii*) B speaks the truth and A tells a lie.

Since, the events are independent, so by using multiplication theorem, we have

(*i*) Probability in the 1st case $=\frac{4}{5}\times\frac{1}{10}=\frac{4}{50}$

(*ii*) Probability in the 2nd case $=\frac{9}{10}\times\frac{1}{5}=\frac{9}{50}$

Since these cases are mutually exclusive, so by using addition theorem, we have

Required Probability $=\frac{4}{50}+\frac{9}{50}=\frac{13}{50}=26\%$

Example 69. A bag contains 5 white and 4 black balls. A ball is drawn from this bag and is replaced and then second draw of a ball is made. What is the probability that two balls are of different colours (*i.e.*, one is white and one is black).

Solution: There are two possibilities:

(*i*) Ist ball drawn is white and the second drawn is black.

(*ii*) Ist ball drawn is black and the second drawn is white.

Since the events are independent, so by using multiplication theorem, we have

(*i*) Probability in the 1st case $=\frac{5}{9}\times\frac{4}{9}=\frac{20}{81}$

(*ii*) Probability in the 2nd case $=\frac{4}{9}\times\frac{5}{9}=\frac{20}{81}$

Since, these possibilities are mutually exclusive, so by using addition theorem, we have

Required Probability $=\frac{20}{81}+\frac{20}{81}=\frac{40}{81}$

Example 70. The chances of survival after 25 years are 3 out of 10 for a man and 4 out of 10 for a woman. Find the probability that only one of them will be alive after 25 years.

Solution: Let P(A) and P(B) denote the probability that man and woman will survive. Then,

$$P(A)=\frac{3}{10}, \quad P(\bar{A})=1-\frac{3}{10}=\frac{7}{10}$$

$$P(B)=\frac{4}{10}, \quad P(\bar{B})=1-\frac{4}{10}=\frac{6}{10}$$

There are two possibilities:

(*i*) Man is alive and woman is not alive

(*ii*) Woman is alive and man is not alive.

Since the events are independent, so by using multiplication theorem, we have

(*i*) Probability in the 1st case $=\frac{3}{10}\times\frac{6}{10}$

(*ii*) Probability in the 2nd case $=\frac{4}{10}\times\frac{7}{10}$

Since, both the cases are mutually exclusive, so by using addition theorem, we have

$$\text{Required Probability}=\frac{3}{10}\times\frac{6}{10}+\frac{4}{10}\times\frac{7}{10}$$

$$=\frac{18}{100}+\frac{28}{100}=\frac{46}{100}$$

Example 71. A bag contains 5 white and 3 red balls and four balls are successively drawn out and not replaced. What is the chance that they are alternatively of different colours?

Solution: 4 balls of alternative colours can be either white, red, white, red or red, white, red, white.

Beginning with White Ball:

The probability of drawing a white ball $= \frac{5}{8}$

The probability of drawing a red ball $= \frac{3}{7}$

The probability of drawing a white ball $= \frac{4}{6}$

The probability of drawing a red ball $= \frac{2}{5}$

Since, the events are dependent, so by using multiplication theorem, we have

$$P(1W\,1R\,1W\,1R) = \frac{5}{8}\times\frac{3}{7}\times\frac{4}{6}\times\frac{2}{5} = \frac{1}{14} \quad ...(i)$$

Beginning with Red Ball:

The probability of drawing a red ball $= \frac{3}{8}$

The probability of drawing a white ball $= \frac{5}{7}$

The probability of drawing a red ball $= \frac{2}{6}$

The probability of drawing a white ball $= \frac{4}{5}$

Since the events are dependent, so by using multiplication theorem, we have

$$P(1R\,1W\,1R\,1W) = \frac{3}{8}\times\frac{5}{7}\times\frac{2}{6}\times\frac{4}{5} = \frac{1}{14} \quad ...(ii)$$

Since, (*i*) and (*ii*) cases are mutually exclusive, so by using addition theorem, we have

$$\text{Required Probability} = \frac{1}{14}+\frac{1}{14} = \frac{2}{14} = \frac{1}{7}$$

Example 72. One bag contains 4 white and 2 black balls. Another bag contains 3 white and 5 black balls. If one ball is drawn from each bag, find the probability that one is white and one is black.

Solution: There are two possibilities:

(*i*) either 1st ball is white from 1st bag and the 2nd ball is black from the 2nd bag.

(*ii*) or 1st ball is black from 1st bag and the 2nd ball is white from the 2nd bag.

Since, the events are independent, so by using multiplication theorem, we have

(*i*) Probability in the 1st case $= \frac{4}{6}\times\frac{5}{8} = \frac{20}{48}$

(*ii*) Probability in the 2nd case $= \frac{2}{6}\times\frac{3}{8} = \frac{6}{48}$

Since, the possibilities are mutually exclusive, so by using addition theorem, we have:

$$\text{Required Probability} = \frac{20}{48}+\frac{6}{48} = \frac{26}{48} = \frac{13}{24}$$

Example 73. A six faced die is so biased that it is twice as likely to show an even number as odd number when it is thrown twice. What is the probability that the sum of two numbers thrown is even?

Solution: Let p be the probability of getting an even number in a single throw of a die and q be that of an odd number.

Given: Even Number : Odd Number :: 2 : 1

$$\therefore \; p = P(\text{Even}) = \frac{2}{3},\; q = P(\text{Odd}) = \frac{1}{3}$$

There are two mutually exclusive cases in which sum of two numbers may be even:

(*i*) Odd number in the first throw and again an odd number in the second throw.

(*ii*) An even number in the first throw and again an even number in the second throw.

Since, the events are independent, so by using multiplication theorem, we have

(*i*) Probability in the 1st case $= \frac{1}{3}\times\frac{1}{3} = \frac{1}{9}$

(*ii*) Probability in the 2nd case $= \frac{2}{3}\times\frac{2}{3} = \frac{4}{9}$

Since, these possibilities are mutually exclusive, so by using addition theorem, we have:

$$\text{Required Probability} = \frac{1}{9}+\frac{4}{9} = \frac{5}{9}$$

Example 74. Four persons are chosen at random from a group containing 3 men, 2 women and 4 children. Find the probability that exactly 2 of them will be children.

Solution: There are three possibilities in this case:

(*i*) 2 children, 1 man and 1 woman

(*ii*) 2 children and 2 men

(*iii*) 2 children and 2 women.

(*i*) Probability in the first case $= \frac{{}^4C_2 \times {}^3C_1 \times {}^2C_1}{{}^9C_4} = \frac{6\times3\times2}{126} = \frac{36}{126}$

(*ii*) Probability in the second case $= \frac{{}^4C_2 \times {}^3C_2}{{}^9C_4} = \frac{6\times3}{126} = \frac{18}{126}$

(*iii*) Probability in the third case $= \frac{{}^4C_2 \cdot {}^2C_2}{{}^9C_4} = \frac{6\times1}{126} = \frac{6}{126}$

Since, the three possibilities are mutually exclusive, so by using addition theorem, we have:

$\therefore$ Required Probability $= \frac{36}{126} + \frac{18}{126} + \frac{6}{126} = \frac{60}{126} = \frac{10}{21}$

Example 75. Three groups of workers contain 3 men and 1 woman, 2 men and 2 women, and 1 man and 3 women. One worker is selected at random from each group. What is probability that the group selected consists of 1 man and 2 women?

Solution: There are three possibilities in this case:

(*i*) 1 man is selected from the first group and 1 woman each from 2nd and 3rd group.

(*ii*) 1 man is selected from the 2nd group and 1 woman each from 1st and 3rd group.

(*iii*) 1 man is selected from the 3rd group and 1 woman each from 1st and 2nd group.

Since, the events are independent, so by using multiplication theorem, we have

(*i*) Probability in the 1st case $= \frac{3}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{18}{64}$.

(*ii*) Probability in the 2nd case $= \frac{2}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{6}{64}$.

(*iii*) Probability in the 3rd case $= \frac{1}{4} \times \frac{1}{4} \times \frac{2}{4} = \frac{2}{64}$.

Since, these possibilities are mutually exclusive, so by using addition theorem, we have:

$\therefore$ Required Probability $= \frac{18}{64} + \frac{6}{64} + \frac{2}{64} = \frac{26}{64} = \frac{13}{32}$.

Example 76. The probability that a management trainee will remain with a company is 0.60. The probability that an employee earns more than Rs. 50,000 per year is 0.50. The probability that an employee is a management trainee who remained with the company or who earns more than Rs. 50,000 per year is 0.70. What is the probability that an employee earns more than Rs. 50,000 per year, given that he is a management trainee who stayed with the company?

Solution: Let A = An employee who earns more than Rs. 50,000 per year

B = A management trainee who stayed with the company

Then, $P(A) = 0.50$; $P(B) = 0.60$; $P(A \cup B) = 0.70$

To get the value of $P(A \cap B)$, we shall use the following formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Substituting the values, we have

$$0.70 = 0.50 + 0.60 - P(A \cap B) \text{ or } P(A \cap B) = 0.40$$

Therefore, $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.40}{0.60} = \frac{2}{3} = 0.67$

Example 77. A can hit a target 4 times in 5 shots. B 3 times in 4 shots and C twice in 3 shots. They fire a volley. What is the probability that at least two shots hit?

Solution: There are four possibilities:

(*i*) A and B hit and C does not hit.

(*ii*) A and C hit and B does not hit.

(*iii*) B and C hit and A does not hit.

(*iv*) A, B and C hit the target.

Since, the events are independent, so by using multiplication theorem, we have

(*i*) Probability in the 1st case $= \frac{4}{5} \times \frac{3}{4} \times \left(1 - \frac{2}{3}\right) = \frac{12}{60}$

(*ii*) Probability in the 2nd case $= \frac{4}{5} \times \frac{2}{3} \times \left(1 - \frac{3}{4}\right) = \frac{8}{60}$

(*iii*) Probability in the 3rd case $= \frac{3}{4} \times \frac{2}{3} \times \left(1 - \frac{4}{5}\right) = \frac{6}{60}$

(*iv*) Probability in the 4th case $= \frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} = \frac{24}{60}$

Since, these possibilities are mutually exclusive, so by using addition theorem, we have:

$\therefore$ Required Probability $= \frac{12}{60} + \frac{8}{60} + \frac{6}{60} + \frac{24}{60} = \frac{50}{60} = \frac{5}{6}$.

Example 78. Two cards are drawn from a pack of 52 cards at random and kept out. Then one card is drawn from the remaining 50 cards. Find the probability that it is an ace.

Solution: There are three cases for first two cards be drawn:

(*i*) two cards drawn are aces, (*ii*) one is an ace, one is not, (*iii*) none of them is an ace.

(*i*) Probability in the first case = (Probability of getting 2 aces) × (Probability of getting third card to be an ace)

$$= \frac{{}^4C_2}{{}^{52}C_2} \times \frac{2}{50} = \frac{1}{5525}$$

(*ii*) Probability in the second case $= \frac{{}^4C_1 \times {}^{48}C_1}{{}^{52}C_2} \times \frac{3}{50}$

$$= \frac{48}{5525}$$

(*iii*) Probability in the third case $= \frac{^{48}C_2}{^{52}C_2} \times \frac{4}{50} = \frac{376}{5525}$

Since, all the three cases are mutually exclusive, so by using addition theorem, we have:

$\therefore$ Required Probability $= \frac{1}{5525} + \frac{48}{5525} + \frac{376}{5525} = \frac{1}{13}$.

Example 79. A purse contains 4 copper coins and 3 silver coins, the second purse contains 6 copper coins and 2 silver coins. A coin is taken from any purse. Find the probability that it is a copper coin.

Solution: There are equal chances of choosing either purse, *i.e.*,

$$P \text{ (Ist Purse)} = P \text{ (2nd Purse)} = \frac{1}{2}$$

The probability that the first purse is chosen and a copper coin is drawn $= \frac{1}{2} \times \frac{4}{7}$

The probability that the second purse is chosen and a copper coin is drawn $= \frac{1}{2} \times \frac{6}{8}$

Since, the events are mutually exclusive, so by using addition theorem, we have:

$\therefore$ Required Probability $= \frac{1}{2} \times \frac{4}{7} + \frac{1}{2} \times \frac{6}{8} = \frac{2}{7} + \frac{3}{8} = \frac{16+21}{56} = \frac{37}{56}$

Example 80. A bag contains 6 white and 4 black balls and a second one 4 white and 8 black balls. One of these bag is chosen at random and a draw of 2 balls is made from it. Find the probability that one is white and the other black.

Solution: There are equal chance of choosing either bag, *i.e.*,

$$P(I) = P(II) = \frac{1}{2}$$

The probability that the first bag was selected and a draw of 2 balls gives one white and one black

$$= \frac{1}{2} \times \frac{^6C_1 \times {^4C_1}}{^{10}C_2} = \frac{4}{15}$$

The probability that the second bag was selected and a draw of 2 balls gives one white and one black

$$= \frac{1}{2} \times \frac{^4C_1 \times {^8C_1}}{^{12}C_2} = \frac{8}{33}$$

Since, the events are mutually exclusive, so by using addition theorem, we have:

$\therefore$ Required Probability $= \frac{4}{15} + \frac{8}{33} = \frac{84}{165} = \frac{28}{55}$

Example 81. A bag contains 1 black and 2 white balls. Another bag contains 2 black and 1 white balls. A ball is drawn from first bag and put it into second bag and then a ball is drawn from the second bag. Find the probability that it is a white ball.

Solution: There are two possibilities–

(*i*) The ball transferred is black one or

(*ii*) It is a white ball

(I) When black ball is transferred:

Probability of drawing a black ball from 1st bag $= \frac{1}{3}$

Now second bag has 3 (2 +1) black and 1 white balls

$\therefore$ Probability of drawing a white ball from 2nd bag $= \frac{1}{(2+1)+1} = \frac{1}{4}$

$\therefore$ Probability of the compound event $= \frac{1}{3} \times \frac{1}{4} = \frac{1}{12}$ ...(*i*)

(II) When white ball is transferred:

Probability of drawing a white ball from 1st bag $= \frac{2}{3}$

Now second bag has 2 (1+1) black and 2 white balls

$\therefore$ Probability of drawing a white ball from 2nd bag $= \frac{1+1}{2+(1+1)} = \frac{2}{4}$.

$\therefore$ Probability of the compound event $= \frac{2}{3} \times \frac{2}{4} = \frac{1}{3}$ ...(*ii*)

Since, (*i*) and (*ii*) possibilities are mutually exclusive, so by using addition theorem, we have:

$\therefore$ Required Probability $= \frac{1}{12} + \frac{1}{3} = \frac{5}{12}$.

Example 82. One bag A contains 10 white and 3 black balls. Another bag B contains 3 white and 5 black balls. Two balls are transferred from bag A and put into the bag B and a ball is drawn from the bag B. Find the probability that the ball drawn is a white ball.

Solution: There are four possibilities— (I) Either both white balls are transferred, or (II) both black balls are transferred, or (III) 1st white and 2nd black balls are transferred, or (IV) 1st black and 2nd white balls are transferred.

(I) When both white balls are transferred:

Probability of drawing 2 white balls from bag A $= \frac{10}{13} \times \frac{9}{12} = \frac{15}{26}$

Now, bag B has 5 white and 5 black balls.

$\therefore$ Probability of drawing a white ball from bag B $= \frac{3+2}{(3+2)+5} = \frac{5}{10}$

$\therefore$ Probability of the compound event $= \frac{15}{26} \times \frac{5}{10} = \frac{75}{260}$. ...(i)

(II) When both black balls are transferred:

Probability of drawing 2 black balls from bag A $= \frac{3}{13} \times \frac{2}{12} = \frac{1}{26}$

Now bag B has 3 white and 7 black balls.

$\therefore$ Probability of drawing a white ball from bag B $= \frac{3}{3+(5+2)} = \frac{3}{10}$

$\therefore$ Probability of the compound event $= \frac{1}{26} \times \frac{3}{10} = \frac{3}{260}$. ...(ii)

(III) When first white and second black balls are transferred:

Probability of drawing first white and second black ball from bag A

$$= \frac{10}{13} \times \frac{3}{12} = \frac{5}{26}$$

Now, bag B has 4 white and 6 black balls.

$\therefore$ Probability of drawing a white ball from bag B $= \frac{3+1}{(3+1)+(5+1)} = \frac{4}{10}$

$\therefore$ The probability of the compound event $= \frac{5}{26} \times \frac{4}{10} = \frac{1}{13}$. ...(iii)

(IV) When first black and second white balls are transferred:

Probability of drawing first black and second white balls from bag A

$$= \frac{3}{13} \times \frac{10}{12} = \frac{5}{26}.$$

Now, bag B has 4 white and 6 black balls.

$\therefore$ Probability of drawing a white ball from bag B $= \frac{3+1}{(3+1)+(5+1)} = \frac{4}{10}$

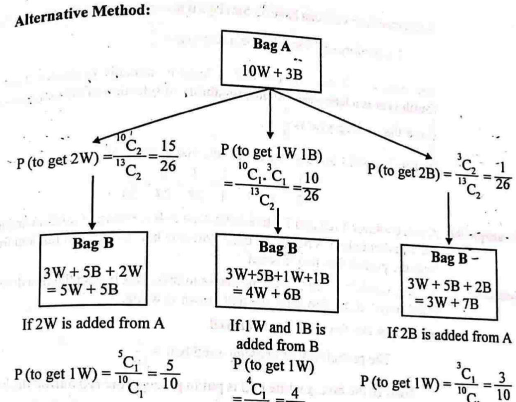
$\therefore$ Probability of the compound event $= \frac{5}{26} \times \frac{4}{10} = \frac{1}{13}$ ...(iv)

Since all the four possibilities are mutually exclusive, so by using addition theorem, we have:

$\therefore$ Required Probability $= \frac{75}{260} + \frac{3}{260} + \frac{1}{13} + \frac{1}{13}$

$$= \frac{118}{260} = \frac{59}{130}.$$

Alternative Method:



Required Probability $= \frac{15}{26} \times \frac{5}{10} + \frac{10}{26} \times \frac{4}{10} + \frac{1}{26} \times \frac{3}{10}$

$$= \frac{75}{260} + \frac{40}{260} + \frac{3}{260} = \frac{118}{260} = \frac{59}{130}$$

Example 83. Find the probability of 53 Sundays in a year selected at random.

Solution: Selection of a year at random means that either it may be normal (non-leap) year or a leap year.

Probability of 53 Sundays in a non-leap (normal) year

In a normal year, there are 365 days, *i.e.*, 52 weeks and 1 day extra. Following are the seven possibilities of this 1 day extra:

$$S, M, T, W, T, F, S = 7$$

A selected normal year can have 53 Sundays if this are extra day happen to be a Sunday.

$\therefore$ Probability of 53 Sundays in a normal year $= \frac{1}{7}$ ...(i)

Probability of 53 Sundays in a leap year

In a leap year, there are 366 days, *i.e.*, 52 weeks and 2 days extra. Following are the seven possibilities of these 2 days extra:

SM, MT, TW, WT, TF, FS, SS = 7

A selected leap year can have 53 Sundays if these 2 extra days happen to be Sundays.

$\therefore$ Probability of 53 Sundays in a leap year $= \frac{2}{7}$ ...(*ii*)

The selection of any year (normal or leap) is mutually exclusive. However, every fourth year is a leap year. So the probability of selection of normal (non-leap) year is $\frac{3}{4}$ and that of leap year is $\frac{1}{4}$.

Probability of 53 Sundays in a year selected at random

$$= \frac{1}{7} \times \frac{3}{4} + \frac{2}{7} \times \frac{1}{4} = \frac{3}{28} + \frac{2}{28} = \frac{5}{28}$$

Example 84. A box contains 3 red and 7 white balls. One ball is drawn at random and in its place a ball of other colour is put in the box. Now one ball is drawn at random from the box. Find the probability that it is red.

Solution: At the second draw red ball can be drawn in two ways if (*i*) at the first draw ball drawn is red or (*ii*) at the first draw the ball drawn is white.

(*i*) When the first ball drawn is red

$\therefore$ The probability of drawing a red ball $= \frac{3}{10}$

Now in the box, a white ball is put in place of the red ball so the box contains 2 red and 8 white balls.

$\therefore$ The probability of drawing a red ball in the second draw $= \frac{2}{10}$

Probability of the compound event $= \frac{3}{10} \times \frac{2}{10}$ ...(*i*)

(*ii*) When the first ball drawn is white

The probability of drawing a white ball $= \frac{7}{10}$

Now in the box, a red ball is put in place of white ball so that the box contains 4 red and 6 white balls in the box.

$\therefore$ The probability of drawing a red ball in the second draw

$$= \frac{4}{10}$$

Probability of the compound event $= \frac{7}{10} \times \frac{4}{10}$ ...(*ii*)

Since, the two possibilities are mutually exclusive, so by using addition theorem, we have

$$\text{Probability of drawing a red ball} = \left(\frac{3}{10} \times \frac{2}{10}\right) + \left(\frac{7}{10} \times \frac{4}{10}\right)$$

$$= \frac{6}{100} + \frac{28}{100} = \frac{34}{100}$$

Example 85. Three coins are tossed simultaneously. What is the probability that they will fall 2 heads and 1 tail?

Solution: Two heads and one tail can came in either of 3 ways:

(*i*) HHT

(*ii*) HTH

(*iii*) THH

Since, the events are independent, so by using multiplication theorem, we have

(*i*) Probability in the 1st case $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

(*ii*) Probability in the 2nd case $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

(*iii*) Probability in the 3rd case $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

But all these three possibilities are mutually exclusive, so by applying addition theorem, we have

$$\text{Required Probability} = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

EXERCISE 7.7

1. The odds that A speaks truth is 3 : 2 and the odds that B speaks the truth is 5 : 3. In what percentage of cases are they like to contradict each other. [Ans. $\frac{19}{40}$]
2. A box contains 10 white and 5 black balls and 4 balls are successively drawn and not replaced. Find the probability that they are alternatively of different colours. [Ans. $\frac{10}{91}$]
3. There are two sections A and B in Statistics at B.Com Examination of Kurukshetra University. The probability that a candidate passes in section A is 0.60 and he passes in section B is 0.50. What is the probability that a particular candidate passes only in one of the two sections? [Ans. 0.50]
4. The odds that a book will be favourably reviewed by three independent critics are 3 : 2, 4 : 3 and 2 : 3 respectively. What is the probability that of the three reviews a majority will be favourably?

   [**Hint:** For majority at least two critics should be favourable.] [Ans. $\frac{94}{175}$]
5. A bag contains 5 red and 3 black balls. Another bag contains 6 red and 4 black balls. If one ball is drawn from each bag, find the probability that (*i*) one is red and the other is black (*ii*) both are red and (*iii*) both are black. [Ans. (*i*) $\frac{19}{40}$ (*ii*) $\frac{3}{8}$ (*iii*) $\frac{3}{20}$]

6. A can hit a target 2 times in 5 shots, B 3 times in 7 shots and C 2 times in 4 shots. What is the probability that only one shot hit the target? [Ans. $\frac{58}{140}$]

7. A bag contains 5 white and 4 red balls and another bag contains 4 white and 6 red balls. One bag is chosen at random and a draw of 2 balls is then made. Find the probability that one is white and the other is red ball. [Ans. $\frac{49}{50}$]

8. A group consists of 4 men, 3 women and 3 children. Three persons are selected from the group at random. Find the probability that (*i*) at least 2 of them are children, and (*ii*) at the most 2 of them are children. [Ans. (*i*) $\frac{11}{60}$, (*ii*) $\frac{119}{120}$]

9. A bag contains 5 red and 4 green balls. Another bag contains 4 red and 6 green balls. A ball is drawn from the first bag and is placed in the second. A ball is then drawn at random from the second bag. What is the probability that it is red? [Ans. $\frac{41}{99}$]

10. A bag contains 4 white and 6 red balls. Two draws of one ball each are made without replacement. What is the probability that one is red and the other white? [Ans. $\frac{4}{15}$]

11. There are two bags. One bag contains 4 white and 2 black balls. The second bag contains 5 white and 4 black balls. Two balls are transferred from first bag to second bag. Then one ball is taken from the second bag. Find the probability that it is white ball. [Ans. $\frac{19}{33}$]

12. A purse contains 2 silver and 4 copper coins. A second purse contains 4 silver and 3 copper coins. If a coin is picked at random from one of the two purses, what is the probability that it is a silver coin? [Ans. $\frac{19}{21}$]

13. From each of the three married couples one of the partners is selected at random. What is the probability of their being all of one sex? [Ans. $\frac{1}{8}+\frac{1}{8}=\frac{1}{4}$]

14. The odds that A speaks the truth are 3 : 2 and the odds that B does so are 5 : 3. In what percentage cases are they likely to contradict each other in stating the same fact. [Ans. 47.5%]

15. In a group of equal number of men and women 10% men and 45% women are unemployed. What is the probability that a person selected at random is employed? [Ans. $\frac{1}{2}\times\frac{9}{10}+\frac{1}{2}\times\frac{11}{20}=\frac{29}{40}$]

16. Three groups of children consist respectively of 3 girls and 1 boy, 2 girls and 2 boys, 1 girl and 3 boys. One child is selected at random from each group. Find the chance that the 3 selected consists of (*i*) 1 girl and 2 boys (*ii*) 2 girls and 1 boy. [Ans. (*i*) $\frac{13}{32}$ (*ii*) $\frac{13}{32}$]

17. A box contains 3 red and 7 white balls. One ball is drawn at random and in its place, a ball of other colour is put in the box. Now one ball is drawn at random from the box. Find the probability that it is red. [Ans. $\frac{3}{10}\times\frac{2}{10}+\frac{7}{10}\times\frac{4}{10}=\frac{17}{50}$]

18. Find the probability of 53 Fridays in a year selected at random. [Ans. $\frac{1}{7}\times\frac{3}{4}+\frac{2}{7}\times\frac{1}{4}=\frac{5}{28}$]

19. One bag contains 5 white and 7 black balls, and other bag contains 7 white and 6 black balls. If one ball is drawn from each bag, find the probability that (*i*) both are of the same colour (*ii*) both are of different colours. [Ans. (*i*) 77/156 (*ii*) 79/156]

20. An investment firm purchases 3 stocks for one week trading purpose. It assesses the probability that the stocks will increase in value over the week are 0.9, 0.7, and 0.6 respectively. What is the chance that:

 (*i*) all the three stocks will increase and

 (*ii*) at least 2 stock will increase?

 Assume that the movement of these stocks are independent [Ans. (*i*) 0.378, (*ii*) 0.834]

21. Three players A, B and C play a game of hitting a target. As per the past experience, A can hit the target accurately in 2 out of 7 shots, B in 3 out of 5 hits and C in 1 out of 3 hits. If they fire the target simultaneously, then what are the chances that: (*i*) it stands hit and (*ii*) it is hit by 2 players? [Ans. (*i*) $\frac{17}{21}$ (*ii*) $\frac{31}{105}$]

Use of Bernoulli's Theorem in Theory of Probability

Bernoulli's theorem is very useful in working out various probability problems. **This theorem states that if the probability of happening of an event in one trial or experiment is known, then the probability of its happening exactly, 1,2,3,...*r* times in *n* trials can be determined by using the formula:**

$$P(r) = {}^nC_r \; p^r \, . \, q^{n-r} \qquad r = 1,2,3,....n$$

where,

$P(r)$ = Probability of r successes in n trials.

p = Probability of success or happening of an event in one trial.

q = Probability of failure or not happening of the event in one trial.

n = Total number of trials.

The following examples illustrate the applications of this theorem:

Example 86. Three coins are tossed simultaneously. What is the probability that there will be exactly two heads?

Solution: Since we have to find the probability of exactly two heads, the use of Bernoulli Theorem will be convenient. According to this theorem:

$$P(r) = {}^nC_r \; p^r . q^{n-r}$$

Given, $n = 3$, $r = 2$, p = probability of head in throw of one coin $= \frac{1}{2}$

$$q = 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(2H) = {}^3C_2\left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^{3-2}$$

$$= \frac{3!}{(3-2)!\ 2!} \times \frac{1}{8} = 3 \times \frac{1}{8} = \frac{3}{8}$$

Example 87. Eight coins are tossed simultaneously. Find the chance of obtaining exactly 6 heads.

Solution: Given, $n = 8$, p = probability of getting head in one coin $= \frac{1}{2}$

$$q = 1 - \frac{1}{2} = \frac{1}{2},\ r = 6$$

$$P(6H) = {}^8C_6\left(\frac{1}{2}\right)^6 \cdot \left(\frac{1}{2}\right)^2 = \frac{8!}{6!\ 2!} \times \frac{1}{256} = \frac{28}{256} = \frac{7}{64}$$

Example 88. In an army battalion 3/5 of the soldiers are known to be married and the remainder 2/5 unmarried. Calculate the probability of getting exactly 4 married soldiers in a row of 5 soldiers.

Solution: Since, we have to find the probability of exactly 4 married soldiers, the use of Bernoulli Theorem will be more convenient. According to this theorem,

$$P(r) = {}^nC_r\ p^r . q^{n-r}$$

Given, $n = 5$, $r = 4$, p = probability of married soliders $= \frac{3}{5}$

$$q = 1 - p = 1 - \frac{3}{5} = \frac{2}{5}$$

$$P\text{ (4 married soldiers)} = {}^5C_4\left(\frac{3}{5}\right)^4 \cdot \left(\frac{2}{5}\right)^1$$

$$= \frac{5!}{4!\ 1!} \cdot \frac{3\times3\times3\times3}{5\times5\times5\times5} \times \frac{2}{5}$$

$$= \frac{5}{1} \times \frac{3\times3\times3\times3}{5\times5\times5\times5} \times \frac{2}{5} = \frac{162}{625}$$

Example 89. If there are three children in a family, find the probability that there is one girl in the family.

Solution: Given, $n = 3, r = 1, p$ = probability of a girl child $= \frac{1}{2}$

$$q = 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(r = 1) = P(1G) = {}^3C_1\left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^2 = \frac{3!}{2!\ 1!} \times \frac{1}{8} = \frac{3}{8}$$

Example 90. The chance that a ship safely reaches a port is 1/5. Find the probability that out of 5 ships expected at least one would arrive safely.

Solution: Given, $n = 5$, $p = \frac{1}{5}$, $q = 1 - \frac{1}{5} = \frac{4}{5}$

P(at least one ship arriving safely) $= 1 - P$(none arriving safely)

$$= 1 - [{}^5C_0 (p)^0 . (q)^5]$$

$$= 1 - \left[{}^5C_0\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^5\right] = 1 - \left(\frac{4}{5}\right)^5$$

$$= 1 - \left(\frac{4}{5} \times \frac{4}{5} \times \frac{4}{5} \times \frac{4}{5} \times \frac{4}{5}\right) = 1 - \frac{1024}{3125} = \frac{2101}{3125}$$

Example 91. Find the probability of throwing 6 at least once in six throws with a single die.

Solution: p = probability of throwing 6 with a single die $= \frac{1}{6}$

$$q = 1 - \frac{1}{6} = \frac{5}{6}$$

$$n = 6, p = \frac{1}{6}, q = \frac{5}{6}$$

P (at least one six) $= 1 - P$[none six in 6 throws]

$$= 1 - \left[{}^6C_0\left(\frac{1}{6}\right)^0 \cdot \left(\frac{5}{6}\right)^6\right] = 1 - \left(\frac{5}{6}\right)^6$$

Example 92. Three dice are thrown. What is the probability that at least one of the numbers turning up being greater than 4?

Solution: p = probability of a number greater than 4 (*i.e.*, 5 and 6) in a throw of one die

$$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$\therefore\ n = 3, p = \frac{1}{3}, q = \frac{2}{3}$$

P(at least one number greater than 4) $= 1 - P$(none of the number greater than 4)

$$= 1 - \left[{}^3C_0\left(\frac{1}{3}\right)^0 \cdot \left(\frac{2}{3}\right)^3\right]$$

$$= 1 - \left(\frac{2}{3}\right)^3 = 1 - \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{19}{27}$$

Example 93. The probability that India wins a cricket test match against England is given to be $\frac{1}{3}$. If India and England play three test matches, find the probability that (*i*) India will lose all the three matches and (*ii*) India will win at least one test match.

Solution: Given, $n=3$,

$$p = \text{probability of wining the match} = \frac{1}{3}$$

$$q = 1-\frac{1}{3} = \frac{2}{3}$$

(*i*) P (losing all matches) $= P(0) = {}^3C_0 \left(\frac{1}{3}\right)^0 \cdot \left(\frac{2}{3}\right)^3 = \frac{8}{27}$

(*ii*) P(at least win one test match) $= 1 - P$(does not win none)

$$= 1 - {}^3C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^3$$

$$= 1 - \frac{8}{27}$$

$$= \frac{19}{27}$$

EXERCISE 7.8

1. The chance that a ship arrives safely at a port is $\frac{9}{10}$. Find the probability that out of 5 ships expected exactly 4 will arrive safely. [Ans. $\frac{59049}{1,00,000}$]
2. What is the probability of getting exactly 3 heads in five throws of a single coin? [Ans. $\frac{5}{16}$]
3. If three coins are tossed simultaneously, what is the probability that they will fall alike? [Ans. $\frac{1}{4}$]
4. Eight coins are tossed simultaneously. What is the probability that they will fall 6 heads and 2 tails up? [Ans. $\frac{28}{256}$]
5. Find the probability of having at least one head in 5 throws with a coin. [Ans. $\frac{31}{32}$]
6. A coin is tossed six times. What is the probability of obtaining four or more heads? [Ans. 0.3437]

o Mathematical Expectation

If a person is to receive a particular amount of money on the happening (or success) of an event, then the multiplication of probability of that event (P) with the amount to be received with happening of the event is known as mathematical expectation. Symbolically,

$$M.E. = P \times M$$

Where, $M.E.$ = Mathematical expectation

P = Probability of happening (or success) of the event

M = Money to be received on the happening of the event.

For example, if the probability of happening of an event (P) is $\frac{1}{6}$ and the money to be received on the happening of such event (M) is Rs. 600, then

$$M.E. = \frac{1}{6} \times 600 = \text{Rs. } 100$$

The following examples illustrate the applications of mathematical expectation:

Example 94. A bag contains 2 white balls and 3 black balls. Four persons A, B, C, D in the order named each draws one ball and does not replace it. The first to draw a white ball receives Rs. 50. Determine their expectations.

Solution: First of all, the ball will be drawn by A. Hence, the probability of drawing white ball by him $P(A) = \frac{2}{2+3} = \frac{2}{5}$

A's expectation $= \frac{2}{5} \times 50 = \text{Rs. } 20$

If A draws a black ball, then the probability of drawing white ball by B, *i.e.*,

$$P(\bar{A}) \times P(B/\bar{A}) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$$

B's expectation $= \frac{3}{10} \times 50 = \text{Rs. } 15$

If A and B both draw black balls, then the probability of drawing white ball by C

$$= P(\bar{A}) \times P(\bar{B}/\bar{A}) \times P(C/\bar{A}\,\bar{B})$$

$$= \frac{3}{5} \times \frac{2}{4} \times \frac{2}{3} = \frac{1}{5}$$

C's expectation $= \frac{1}{5} \times 50 = \text{Rs. } 10$

If A, B and C all draw black balls, then the probability of drawing white ball by D

$$= P(\bar{A}) \times P(\bar{B}/\bar{A}) \times P(\bar{C}/\overline{AB}) \times P(D/\overline{ABC})$$

$$= \frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} \times \frac{2}{2} = \frac{1}{10}$$

D's expectation $= \frac{1}{10} \times 50 = \text{Rs. } 5$

Example 95. A and B play for a prize of Rs. 1000. A is to throw a die first and is to win if he throws 6. If he fails B is to throw and is to win if throws 6 or 5. If he fails A is to throw again and to win if he throws 6, 5 or 4 and so on. Find their respective expectations.

Solution: Probability of A's winning in the 1st throw (*i.e.*, he throws 6) $= \frac{1}{6}$

Probability of B's winning in the 2nd throw (*i.e.*, he throws 6 or 5) $= \frac{5}{6} \times \frac{2}{6} = \frac{5}{18}$

Probability of A's winning in the 3rd throw (6 or 5 or 4) $= \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} = \frac{5}{18}$

Probability of B's winning in the 4th throw (6 or 5 or 4 or 3) $= \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{4}{6} = \frac{5}{27}$

Probability of A's winning in the 5th throw (6 or 5 or 4 or 3 or 2)

$$= \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times \frac{5}{6} = \frac{25}{324}$$

Probability of B's winning in the 6th throw (6 or 5 or 4 or 3 or 2 or 1)

$$= \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times \frac{1}{6} \times \frac{6}{6} = \frac{5}{324}$$

A's total chances of success $= \frac{1}{6} + \frac{5}{18} + \frac{25}{324} = \frac{169}{324}$

B's total chances of success $= \frac{5}{18} + \frac{5}{27} + \frac{5}{324} = \frac{155}{324}$

For a prize of Rs. 1,000

A's expectation $= p \times m = \frac{169}{324} \times 1{,}000 =$ Rs. 521.6

B's expectation $= p \times m = \frac{155}{324} \times 1{,}000 =$ Rs. 478.4

Example 96. A and B play for a prize of Rs. 99. The prize is to be won by a player who first throws 6 with one die. A first throws and if he fails B throws and if he fails A again throws and so on. Find their respective expectations.

Solution: The probability of throwing 6 with a single die $= \frac{1}{6}$

The probability of not throwing 6 with single die $= 1 - \frac{1}{6} = \frac{5}{6}$

If A is to win, he should throw 6 in the 1st, 3rd or 5th...throws

If B is to win, he should throw 6 in the 2nd, 4th, 6th...throws

A's chance of success is given by

$$= \frac{1}{6} + \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right) + \ldots \infty$$

$$= \frac{1}{6} \cdot \left[1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \ldots \infty\right]$$ [Infinite GP series: $S = 1 + a + a^2 + \ldots \infty$]

$$= \frac{1}{6} \cdot \left[\frac{1}{1 - \left(\frac{5}{6}\right)^2}\right] = \frac{1}{6} \times \frac{36}{11} = \frac{6}{11} \qquad \left(\therefore S_\infty = \frac{1}{1-a} = \frac{\text{First Term}}{1 - \text{Common Ratio}}\right)$$

A's expectation $= p \times m = \frac{6}{11} \times 99 =$ Rs. 54

B's chance of success is given by

$$= \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^5 \cdot \left(\frac{1}{6}\right) + \ldots \infty$$

$$= \frac{5}{6} \times \frac{1}{6} \cdot \left[1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \ldots \infty\right]$$

$$= \frac{5}{6} \times \frac{1}{6} \cdot \left[\frac{1}{1 - \left(\frac{5}{6}\right)^2}\right] = \frac{5}{6} \times \frac{1}{6} \times \frac{36}{11} = \frac{5}{11}$$

B's expectation $=$ Rs. $99 \times \frac{5}{11} =$ Rs. 45.

Example 97. A bag contains 6 black and 9 white balls. A person draws out 2 balls. If on every black ball he gets Rs. 20 and on every white ball Rs. 10, find out his expectation.

Solution: There may be the following three options for drawing 2 balls:

(*i*) Both are white, (*ii*) Both are black, (*iii*) One is white and other is black.

(*i*) Both balls are white

$$P(2W) = p = \frac{{}^9C_2}{{}^{15}C_2} = \frac{12}{35}$$

Expectation $= p \times m = \frac{12}{35} \times 10 \times 2 =$ Rs. 6.86

(*ii*) Both balls are black

$$P(2B) = p = \frac{{}^6C_2}{{}^{15}C_2} = \frac{1}{7}$$

Expectation $= p \times m = \frac{1}{7} \times 20 \times 2 =$ Rs. 5.71

(iii) **One ball is white and the other is black**

$$P(1W\,1B) = p = \frac{{}^{6}C_1 \times {}^{9}C_1}{{}^{15}C_2} = \frac{18}{35}$$

$$\text{Expectation} = p \times m = \frac{18}{35} \times (20+10) = \text{Rs. } 15.43$$

$$\text{Total Expectation} = 6.86 + 5.71 + 15.43 = \text{Rs. } 28$$

Example 98. If it rains, a taxi driver can earn Rs. 1000 per day. If it is fair, he can lose Rs.100 per day. If the probability of rain is 0.4, what is his expectation?

Solution: The distribution of earnings (X) is given as:

X	$X_1 = 1000$	$X_2 = -100$
P	$P_1 = 0.4$	$P_2 = 1 - 0.4 = 0.6$

$$\therefore\; E(X) = P_1 X_1 + P_2 X_2$$
$$= 0.4 \times 1000 + 0.6 \times (-100) = \text{Rs. } 340$$

Example 99. A petrol pump dealer sells an average petrol of Rs. 80,000 on a rainy day and an average of Rs. 95,000 at a clear day. The probability of clear weather is 76% on Tuesday. What will be the expected sale?

Solution: The distribution of earnings (X) is given as:

X	$X_1 = 80{,}000$	$X_2 = 95{,}000$
P	$1 - 0.76 = 0.24$	0.76

$$E(X) = 80{,}000 \times 0.24 + 95{,}000 \times 0.76$$
$$= \text{Rs. } 91{,}400$$

Example 100. A player tosses 3 fair coins. He wins Rs. 12 if 3 heads appear, Rs. 8 if 2 heads appear and Rs. 3 if 1 head appears. On the otherhand, he loses Rs. 25 if 3 tails appear. Find the expected gain of the player.

Solution: If p denotes the probability of getting a head and X denotes the corresponding amount of winning, then the distribution of X is given by:

Heads:	0H	1H	2H	3H
Favourable Events	TTT	HTT, THT, TTH	HHT, HTH, THH	HHH
P	$\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{8}$	$\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$	$\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$	$\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{8}$
X Winning amount	–25	3	8	12

The expected gain of the player is given by:

$$E(X) = \frac{1}{8}(-25) + \frac{3}{8}(3) + \frac{3}{8}(8) + \frac{1}{8}(12)$$
$$= \frac{-25+9+24+12}{8} = \frac{20}{8} = \frac{5}{2} = \text{Rs. } 2.50.$$

Example 101. A player tosses two fair coins. He wins Rs. 5 if 2 heads appear, Rs. 2 if one head appear and Rs. 1 if no head appear. Find his expected gain of the player.

Solution: If p denotes the probability of getting a head and X denotes the corresponding amount of winning, then the probability distribution of X is given by:

Heads:	0H	1H	2H
Favourable Events	TT	HT, TH	HH
P	$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$	$\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$	$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$
X	1	2	5

The expected gain of the player is given by:

$$E(X) = P_1X_1 + P_2X_2 + P_3X_3$$
$$= \frac{1}{4}\times 1 + \frac{1}{2}\times 2 + \frac{1}{4}\times 5 = \text{Rs. } 2.50$$

Example 102. A survey conducted over the last 25 years indicated that in 10 years, the winter was mild, in 8 years it was cold and in the remaining 7 it was very cold. A company sells 1000 woolen coats in a mild year, 1300 in a cold year and 2000 in a very cold year. If a woolen coat costs Rs. 173 and is sold for Rs. 248, find the yearly expected profit of the company.

Solution:

State of Nature	Prob. P(X)	Sale of woollen coat	Profit (X)
Mild winter	$\frac{10}{25}=0.4$	1000	$1000 \times (248 - 173)$
Cold winter	$\frac{8}{25}=0.32$	1300	$1300 \times (248 - 173)$
Very cold winter	$\frac{7}{25}=0.28$	2000	$2000 \times (248 - 173)$

$\therefore$ Expected profit is given by

$$E(X) = 1000 \times 75 \times 0.4 + 1300 \times 75 \times 0.32 + 2000 \times 75 \times 0.28$$
$$= 30{,}000 + 31{,}200 + 42{,}000 = \text{Rs. } 1{,}03{,}200$$

EXERCISE 7.9

1. A bag contains 3 red and 4 green balls. Four persons A, B, C and D in the order named draws one ball and does not replace it. The first to draw a green ball receives Rs. 56. Determine their expectations. **[Ans. Rs. 32, Rs. 16, Rs. 6.40, Rs. 1.60]**
2. A and B play for a prize of Rs. 99. The prize is to won by a player who first throws '3' with one die. If A throws first and if he fails, B throws and if he fails A again throws and so on. Find their respective expectations. **[Ans. Rs. 54, Rs. 45]**

3. A and B play for a prize of Rs. 500. A is to throw a die first and is to win if he throws 6. If he fails B is to throw and is to win if he throws 6 or 5. If he fails A is to throw again and to win if he throws 6, 5 or 4 and so on. Find their respective expectations. **[Ans. Rs. 260.80, Rs. 239.69]**
4. A box contains 4 white and 6 black balls. A person draws 2 balls and is given Rs. 14 for every white ball and Rs. 7 for every black ball. What is his expectation? **[Ans. Rs. 19.50]**
5. If it rains, a dealer of an umbrella can earns Rs. 300 per day. If it does not rain, he bears a loss of Rs. 80 per day. What is his expectation if the probability of rainy days is 0.57? **[Ans. Rs. 136.60]**
6. A person receives Rs. 400 for a head and losses Rs. 300 for a tail when a coin is tossed. Find his expectation. **[Ans. Rs. 50]**
7. A player tosses 3 fair coins. He wins Rs. 10 if 3 heads appear, Rs. 6 if 2 heads appear, Rs.2 if 1 head appears. On the other hand, he loses Rs. 25 if 3 tails occurs. Find the expected gain of the player. **[Ans. Rs. 1.13]**
8. A throws a coin 3 times. If he gets a head all the three times he is to get a prize of Rs. 120. The entry fee for the game is Rs. 12. What is the mathematical expectation of A?
[Ans. $E(A) = 120 \times \frac{1}{8} + (-12) \times \frac{7}{8} = 15 - 10.5 =$ **Rs. 4.5]**

---

(3) Bayes' Theorem

Bayes' Theorem is named after the British Mathematician Thomas Bayes and it was published in the year 1763. With the help of Bayes' Theorem, prior probability are revised in the light of some sample information and posterior probabilities are obtained. This theorem is also called **Theorem of Inverse Probability.** Suppose in a factory, two machines A_1 and A_2 are manufacturing goods. Further suppose that machine A_1 and A_2 manufacture respectively 70% and 30% of the total with 5% and 3% of total defective bolts. Suppose an item is selected from the total production and found to be defective. And if we want to find out the probability that it was manufactured by machine A_1 or machine A_2, then this can be found by using Bayes' Theorem. Take another example, Suppose an urn contains 6 black and 4 white balls. Another urn contains 4 black and 6 white balls. A ball is drawn from one of the urn and found to be black. And if we want to find out the probability that it came from 1st urn or 2nd urn. This can be found by using Bayes' Theorem.

▶ Statement of Bayes' Theorem

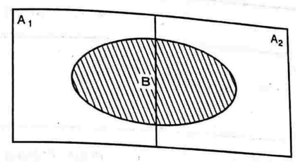
If A_1 and A_2 are mutually exclusive and exhaustive events and B be an event which can occur in combination with A_1 and A_2, then the conditional probability for event A_1 and A_2 given the event B is given by:

$$P(A_1/B) = \frac{P(A_1).P(B/A_1)}{P(A_1).P(B/A_1) + P(A_2).P(B/A_2)}$$

Similarly,

$$P(A_2/B) = \frac{P(A_2).P(B/A_2)}{P(A_1).P(B/A_1) + P(A_2).P(B/A_2)}$$

Proof of the Theorem: Consider the graph given below:



Since, A_1 and A_2 are mutually exclusive events and since the event B occurs with only one of them, so that

$$B = BA_1 + BA_2 \quad \text{or} \quad B = A_1B + A_2B$$

By the addition theorem of probability, we have

$$P(B) = P(A_1B) + P(A_2B) \quad ...(i)$$

Now, by multiplication theorem, we have

$$P(A_1B) = P(A_1).P(B/A_1) \quad ...(ii)$$

$$P(A_2B) = P(A_2).P(B/A_2) \quad ...(iii)$$

Substituting the values of $P(A_1 B)$ and $P(A_2 B)$ in (*i*), we get

$$P(B) = P(A_1).P(B/A_1) + P(A_2).P(B/A_2) \quad ...(iv)$$

Hence,

$$P(B) = \sum_{i=1}^{2} P(A_i).P(B/A_i) \quad ...(v)$$

Again by the theorem of conditional probability, we have

$$P(A_1/B) = \frac{P(A_1B)}{P(B)} \quad ...(vi)$$

Substituting the values of $P(A_1 B)$ and $P(B)$ from (*ii*) and (*iv*) in equation (*vi*), we get

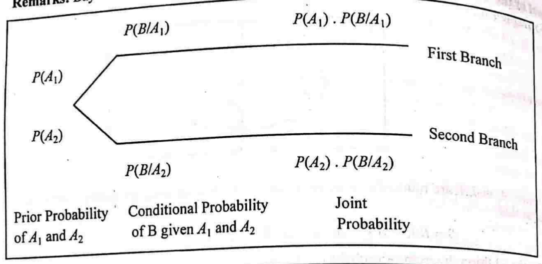
$$P(A_1/B) = \frac{P(A_1).P(B/A_1)}{P(A_1).P(B/A_1) + P(A_2).P(B/A_2)}$$

Similarly,

$$P(A_2/B) = \frac{P(A_2).P(B/A_2)}{P(A_1).P(B/A_1) + P(A_2).P(B/A_2)}$$

The probabilities $P(A_1)$ and $P(A_2)$ are called prior probabilities and probabilities $P(A_1/B)$ and $P(A_2/B)$ are called posterior probabilities.

Remarks: Bayes' Theorem can be expressed by means of the following figure:



Now,

$$P(A_1/B) = \frac{\text{Joint probability of the first branch}}{\text{Sum of the joint probabilities of the two branches}}$$

$$= \frac{P(A_1).P(B/A_1)}{P(A_1).P(B/A_1)+P(A_2).P(B/A_2)}$$

Similarly,

$$P(A_2/B) = \frac{\text{Joint probability of the 2nd branch}}{\text{Sum of the joint probabilities of the two branches}}$$

$$= \frac{P(A_2).P(B/A_2)}{P(A_1).P(B/A_1)+P(A_2).P(B/A_2)}$$

Generalisation

Bayes' Theorem can be extended to three or more events. If A_1, A_2 and A_3 are three mutually exclusive events and B is an event which can occur in combination with A_1, A_2 and A_3, then

$$P(A_1/B) = \frac{P(A_1).P(B/A_1)}{P(A_1).P(B/A_1)+P(A_2).P(B/A_2)+P(A_3).P(B/A_3)}$$

$$P(A_2/B) = \frac{P(A_2).P(B/A_2)}{P(A_1).P(B/A_1)+P(A_2).P(B/A_2)+P(A_3).P(B/A_3)}$$

$$P(A_3/B) = \frac{P(A_3).P(B/A_3)}{P(A_1).P(B/A_1)+P(A_2).P(B/A_2)+P(A_3).P(B/A_3)}$$

The following examples will illustrate the applications of Bayes' Theorem:

Example 103. In a bolt factory machine A, B and C manufacture respectively 25%, 35% and 40% of the total. Of their output 5, 4, 2 per cent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the probability that it was manufactured by machine C?

Solution: Let A, B and C be the events of drawing a bolt produced by machine A, B and C respectively and let D be the event that the bolt is defective.

We are given the information:

$P(A) = 25\% = \frac{25}{100} = 0.25$

$P(B) = 35\% = \frac{35}{100} = 0.35$

$P(C) = 40\% = \frac{40}{100} = 0.40$

The conditional probabilities are:

$P(D/A) = 5\% = \frac{5}{100} = 0.05$

$P(D/B) = 4\% = \frac{4}{100} = 0.04$

$P(D/C) = 2\% = \frac{2}{100} = 0.02$

Putting the given information in the table given below:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Col. (2) × (3)
A	P(A) = 0.25	P(D/A) = 0.05	0.25 × 0.05
B	P(B) = 0.35	P(D/B) = 0.04	0.35 × 0.04
C	P(C) = 0.40	P(D/C) = 0.02	0.40 × 0.02

We have to calculate P(C/D), *i.e.*, the probability that the defective item was produced by machine C.

$$P(C/D) = \frac{\text{Joint Probability of the machine C}}{\text{Sum of Joint Probability of three machines}}$$

$$= \frac{0.40 \times 0.02}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02}$$

$$= \frac{0.008}{0.0125 + 0.014 + 0.008} = \frac{0.008}{0.0345}$$

$$= 0.2318 \text{ or } 23.18\%.$$

Example 104. A manufacturing firm produces steel pipes in three plants with daily production volumes of 500, 1000 and 2000 units respectively. According to past experience, it is known that the fraction of defective output produced by three plants are respectively 0.005, 0.008, 0.010. If a pipe is selected from a day's total production and found to be defective, find the probability that it came from the first plant.

Solution: Let E_1, E_2 and E_3 be the events of selecting steel pipes by plants I, II and III and let D be the event that the pipe is defective.

We are given the information:

$$P(E_1)=\frac{500}{500+1000+2000}=\frac{1}{7}$$

$$P(E_2)=\frac{1000}{500+1000+2000}=\frac{2}{7}$$

$$P(E_3)=\frac{2000}{500+1000+2000}=\frac{4}{7}$$

Conditional probabilities are:

$$P(D/E_1)=0.005$$

$$P(D/E_2)=0.008$$

$$P(D/E_3)=0.010$$

Putting the given information in the table as shown below:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Col. (2) × (3)
E_1	$P(E_1)=1/7$	$P(D/E_1)=0.005$	$1/7\times0.005$
E_2	$P(E_2)=2/7$	$P(D/E_2)=0.008$	$2/7\times0.008$
E_3	$P(E_3)=4/7$	$P(D/E_3)=0.010$	$4/7\times0.010$

We have to calculate $P(E_1/D)$, *i.e.*, the probability that the defective pipe was produced by Ist plant.

$$P(E_1/D)=\frac{\text{Joint Probability of the Ist Plant}}{\text{Sum of Joint Probability of three plants}}$$

$$=\frac{\frac{1}{7}\times0.005}{\frac{1}{7}\times0.005+\frac{2}{7}\times0.008+\frac{4}{7}\times0.010}$$

$$=\frac{0.005}{0.005+0.016+0.040}=\frac{0.005}{0.061}=\frac{5}{61}$$

Example 105. A company has two plants to manufacture scooters. Plant I manufactures 70% of the scooters and Plant II manufactures 30%. At Plant I, 80% scooters are rated standard quality and at Plant II, 90% of the scooters are rated standard quality. A scooter is picked up at random and is found to be standard quality. What is the chance that it comes from Plant I?

Solution: Let A_1 and A_2 be the events that the selected scooter is manufactured by Plant I and Plant II and let B be the event that the scooter is of standard quality.

We are given the information:

$$P(A_1)=70\%=\frac{70}{100}$$

$$P(A_2)=30\%=\frac{30}{100}$$

Conditional probabilities are:

$$P(B/A_1)=80\%=\frac{80}{100}$$

$$P(B/A_2)=90\%=\frac{90}{100}$$

Putting the given information in the table given below:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Col. (2) × (3)
A_1	$P(A_1)=\frac{70}{100}$	$P(B/A_1)=\frac{80}{100}$	$\frac{70}{100}\times\frac{80}{100}$
A_2	$P(A_2)=\frac{30}{100}$	$P(B/A_2)=\frac{90}{100}$	$\frac{30}{100}\times\frac{90}{100}$

We have to find $P(A_1/B)$, *i.e.*, the probability that the standard quality scooter was produced by Plant I.

$$P(A_1/B)=\frac{\text{Joint Probability of the Ist Plant}}{\text{Sum of Joint Probability of Both Plants}}$$

$$=\frac{\frac{70}{100}\times\frac{80}{100}}{\frac{70}{100}\times\frac{80}{100}+\frac{30}{100}\times\frac{90}{100}}=\frac{70\times80}{70\times80+30\times90}=\frac{56}{83}$$

Example 106. An insurance company insured 2000 scooter drivers, 4000 car drivers and 6000 truck drivers. The probability of accident is 0.01, 0.03 and 0.15 respectively. One of the insured person meets an accident. What is the probability that he is a scooter driver?

Solution: Let A_1, A_2 and A_3 be the events that the insured person is a scooter driver, car driver and truck driver respectively and let B be the event that the insured person meets an accident.

We are given the information:

$$P(A_1)=\frac{2000}{12000}=\frac{1}{6}$$

$$P(A_2)=\frac{4000}{12000}=\frac{1}{3}$$

$$P(A_3)=\frac{6000}{12000}=\frac{1}{2}$$

Conditional Probabilities are:

$$P(B/A_1)=0.01=\frac{1}{100}$$

$$P(B/A_2)=0.03=\frac{3}{100}$$

$$P(B/A_3)=0.15=\frac{15}{100}$$

Putting the information in the table given below:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Col. (2) × (3)
A_1	$P(A_1)=\frac{1}{6}$	$P(B/A_1)=\frac{1}{100}$	$\frac{1}{6}\times\frac{1}{100}$
A_2	$P(A_2)=\frac{1}{3}$	$P(B/A_2)=\frac{3}{100}$	$\frac{1}{3}\times\frac{3}{100}$
A_3	$P(A_3)=\frac{1}{2}$	$P(B/A_3)=\frac{15}{100}$	$\frac{1}{2}\times\frac{15}{100}$

We have to find $P(A_1/B)$, *i.e.*, the probability of a scooter driver given that he meets an accident.

By Bayes' theorem,

$$P(A_1/B)=\frac{\frac{1}{6}\times\frac{1}{100}}{\frac{1}{6}\times\frac{1}{100}+\frac{1}{3}\times\frac{3}{100}+\frac{1}{2}\times\frac{15}{100}}$$

$$=\frac{\frac{1}{6}}{\frac{1}{6}\times 1+\frac{1}{3}\times 3+\frac{1}{2}\times 15}=\frac{0.166}{0.166+1+7.5}=\frac{0.166}{8.666}=0.0191$$

Example 107. There are three machines A, B and C in a factory. Their daily outputs are in the ratio of 2:3:1. Past experience shows that 2%, 4% and 5% of the item produced by A, B and C respectively are defective. If an item selected at random is found to be defective, find the probability that it was produced by A or B.

Solution: Let A, B and C denote events that the item is manufactured by A, B and C machines respectively and let D be the event of selecting a defective item.

We are given:

$$P(A)=\frac{2}{2+3+1}=\frac{2}{6} \qquad P(D/A)=2\%=\frac{2}{100}$$

$$P(B)=\frac{3}{2+3+1}=\frac{3}{6} \qquad P(D/B)=4\%=\frac{4}{100}$$

$$P(C)=\frac{1}{2+3+1}=\frac{1}{6} \qquad P(D/C)=5\%=\frac{5}{100}$$

Putting the given information in the form of the table as follows:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Col. (2) × (3)
A	$P(A)=\frac{2}{6}$	$P(D/A)=\frac{2}{100}$	$\frac{2}{6}\times\frac{2}{100}$
B	$P(B)=\frac{3}{6}$	$P(D/B)=\frac{4}{100}$	$\frac{3}{6}\times\frac{4}{100}$
C	$P(C)=\frac{1}{6}$	$P(D/C)=\frac{5}{100}$	$\frac{1}{6}\times\frac{5}{100}$

We have to find P(A/D) + P(B/D), *i.e.*, defective item is produced by machine A or B

By Bayes' theorem,

(*i*)
$$P(A/D)=\frac{\frac{2}{6}\times\frac{2}{100}}{\frac{2}{6}\times\frac{2}{100}+\frac{3}{6}\times\frac{4}{100}+\frac{1}{6}\times\frac{5}{100}}=\frac{\frac{4}{600}}{\frac{4}{600}+\frac{12}{600}+\frac{5}{600}}=\frac{4}{21}$$

(*ii*)
$$P(B/D)=\frac{\frac{3}{6}\times\frac{4}{100}}{\frac{2}{6}\times\frac{2}{100}+\frac{3}{6}\times\frac{4}{100}+\frac{1}{6}\times\frac{5}{100}}=\frac{\frac{12}{600}}{\frac{4}{600}+\frac{12}{600}+\frac{5}{600}}=\frac{12}{21}$$

Hence, the required probability $=\frac{4}{21}+\frac{12}{21}=\frac{16}{21}$

Example 108. A company produces certain type of sophisticated item by three machines. The respective daily production figures are: Machine A 300 units, Machine B 450 units and Machine C 250 units. Past experience shows that the percentage of defective in the three machines are 0.1, 0.2 and 0.7 respectively for the machines A, B and C. An item is drawn at random from a day's production and is found to be defective. What is the probability that it is not produced by machine C?

Solution: Let A, B, C denote the events that the item is manufactured by machines A, B and C respectively and let D denote the event that the defective item is manufactured.

We are given:

$$P(A)=\frac{300}{300+450+250}=\frac{300}{1000}=0.30$$

$$P(B)=\frac{450}{1000}=0.45$$

$$P(C)=\frac{250}{1000}=0.25$$

The conditional probabilities are:

$$P(D/A)=0.1\%=\frac{1}{1000}=0.001$$

$$P(D/B)=0.2\%=\frac{2}{1000}=0.002$$

$$P(D/C)=0.7\%=\frac{7}{1000}=0.007$$

Putting the information in the table given below:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Col. (2) × (3)
A	P(A) = 0.30	P(D/A) = 0.001	0.30 × 0.001
B	P(B) = 0.45	P(D/B) = 0.002	0.45 × 0.002
C	P(C) = 0.25	P(D/C) = 0.007	0.25 × 0.007

If the defective item is not produced by machine C, it means that either it is produced by machine A or by machines B.

$\Rightarrow$ $P(\overline{C}/D)=$ P (either A or B) = P(A/D) + P(B/D)

By Bayes Rule, we have

$$P(A/D)=\frac{0.30\times0.001}{0.30\times0.001+0.45\times0.002+0.25\times0.007}=\frac{30}{295}$$

$$P(B/D)=\frac{0.45\times0.002}{0.30\times0.001+0.45\times0.002+0.25\times0.007}=\frac{90}{295}$$

Thus, $P(\overline{C}/D)=P(A/D)+P(B/D)$

$$=\frac{30}{295}+\frac{90}{295}=\frac{120}{295}=\frac{24}{59}$$

Example 109. There are two urns. Urn I contains 1 white and 6 red balls. Urn II has 4 white and 3 red balls. One of the urn is selected at random and a ball is drawn from it and found to be white. What is the probability that it is drawn from the 1st urn?

Solution: Let A_1 and A_2 stand for the events Urn I is chosen and Urn II is chosen respectively and let W stand for the event that white ball is chosen.

Thus, we are given:

$P(A_1)=\frac{1}{2}$ $\quad P(W/A_1)=\frac{1}{7}$

$P(A_2)=\frac{1}{2}$ $\quad P(W/A_2)=\frac{4}{7}$

Putting the information in the table given below:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Col. (2)×(3)
A_1	$P(A_1)=\frac{1}{2}$	$P(W/A_1)=\frac{1}{7}$	$\frac{1}{2}\times\frac{1}{7}$
A_2	$P(A_2)=\frac{1}{2}$	$P(W/A_2)=\frac{4}{7}$	$\frac{1}{2}\times\frac{4}{7}$

We have to find $P(A_1/W)$, *i.e.*, the probability that the white ball comes from urn I

Using Bayes' Theorem,

$$P(A_1/W)=\frac{\text{Joint Probability of the Ist urn}}{\text{Sum of Joint Probability of two urns}}$$

$$=\frac{\frac{1}{2}\times\frac{1}{7}}{\frac{1}{2}\times\frac{1}{7}+\frac{1}{2}\times\frac{4}{7}}=\frac{\frac{1}{14}}{\frac{5}{14}}=\frac{1}{5}$$

Example 110. A, B and C are three candidates for the post of a Director in a company. Their respective chances of selection are in the ratio of 4 : 5 : 3. The probability that A, if selected will introduce the internet trading in the company is 0.30. Similarly, the probability of B and C are 0.50 and 0.60 respectively. Find the probability that the company will introduce internet trading. Also find the probability that Director B introduced the internet trading in the company.

Solution: Let A_1, A_2 and A_3 denote the events that the persons A, B and C respectively are selected as Director of the company and let E be the event of introducing internet trading in the company. Then we are given:

$P(A_1)=\frac{4}{4+5+3}=\frac{4}{12}$ $\quad P(A_2)=\frac{5}{4+5+3}=\frac{5}{12}$ $\quad P(A_3)=\frac{3}{4+5+3}=\frac{3}{12}$

$P(E/A_1)=0.30$ $\quad P(E/A_2)=0.50$ $\quad P(E/A_3)=0.60$

Putting the information in the table given below:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Cols. (2)×(3)
A_1	$P(A_1)=\frac{4}{12}$	$P(E/A_1)=0.30$	$\frac{4}{12}\times0.30$
A_2	$P(A_2)=\frac{5}{12}$	$P(E/A_2)=0.50$	$\frac{5}{12}\times0.50$
A_3	$P(A_3)=\frac{3}{12}$	$P(E/A_3)=0.60$	$\frac{3}{12}\times0.60$

(*i*) Now, *P*(Internet trading is introduced in the company)

$$P(E)=P(A_1E \text{ or } A_2E \text{ or } A_3E)=P(A_1E)+P(A_2E)+P(A_3E)$$

$$=P(A_1).P(E/A_1)+P(A_2).(E/A_2)+P(A_3).P(E/A_3)$$

$$=\frac{4}{12}\times0.30+\frac{5}{12}\times0.50+\frac{3}{12}\times0.60=\frac{55}{120}=\frac{11}{24}$$

(*ii*) We have to find $P(A_2/E)$, *i.e.*, internet trading is introducing by Director B

By Bayes' Theorem, we have

P(Director B introduces internet trading)

$$P(A_2/E)=\frac{\text{Joint probability of the 2nd}}{\text{Sum of the joint probability}}$$

$$=\frac{\frac{5}{12}\times0.50}{\frac{4}{12}\times0.30+\frac{5}{12}\times0.50+\frac{3}{12}\times0.60}=\frac{\frac{5}{12}\times\frac{1}{2}}{\frac{11}{24}}=\frac{5}{11}$$

Example 111. In a railway reservation office, two clerks are engaged in checking reservation forms. On an average, the first clerk (A_1) checks 55% of the forms while the second (A_2) does the remaining. A_1 has an error rate of 0.03 and A_2 as much as 0.02. A reservation form is selected at random from the total number of forms checked during a day and is found to have an error. Find the probability that it was checked by A_1 and A_2 respectively.

Solution: Let B_1 and B_2 be the event that the forms are checked by clerk (A_1) and clerk (A_2). Let A be the event that a form selected at random has an error.

Now, $P(B_1) = 0.55$, $P(B_2) = 0.45$

$P(A/B_1) = 0.03$, $P(A/B_2) = 0.02$

Putting the information in the table given below:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Cols. (2) × (3)
B_1	$P(B_1) = 0.55$	$P(A/B_1) = 0.03$	0.55×0.03
B_2	$P(B_2) = 0.45$	$P(A/B_2) = 0.02$	0.45×0.02

We have to find $P(B_1/A)$ and $P(B_2/A)$

Using Bayes' Theorem,

(i) $$P(B_1/A) = \frac{\text{Joint Probability of the 1st clerk}}{\text{Sum of Joint Probability of the two}}$$

$$= \frac{0.55 \times 0.03}{0.55 \times 0.03 + 0.45 \times 0.02} = \frac{0.0165}{0.0165 + 0.009} = 0.647$$

(ii) $$P(B_2/A) = \frac{\text{Joint Probability of the 2nd clerk}}{\text{Sum of Joint Probability of the two}}$$

$$= \frac{0.45 \times 0.02}{0.55 \times 0.03 + 0.45 \times 0.02} = \frac{0.009}{0.0165 + 0.009}$$

$= 0.3529$

EXERCISE 7.10

1. A factory has two machines, machine I produces 30% of the items of output and machine II produces 70% of the items. Further, 5% of the items produced by the machine I were defective and only 1% produced by machine II were defective. If a defective item is drawn at random, what is the probability that it was produced by machine I? **[Ans. 15/22]**
2. There are 4 boys and 2 girls in Room No. I and 5 boys and 3 girls in Room No. II. A girl from one of two rooms laughed loudly. What is the probability that the girl who laughed loudly was from Room No. II? **[Ans. 9/17]**
3. A purse contains three one rupee coins and four 50 paise coins. Another purse contains four one-rupees coins and five 50 paise coins. A one-rupee coin has been taken out from one of the purses. Find out the probability that it is from the first purse. **[Ans. 27/55]**
4. There are two identical boxes containing respectively 4 white and 3 red balls, 3 white and 7 red balls. A box is chosen at random and a ball is drawn from it. If the ball drawn is white, what is the probability that it is from first box? **[Ans. 40/61]**

5. A manufacturing firm produces sheet pipes in three plants with daily production volume of 250, 350 and 400 units respectively. According to past experiences, it is known that fraction of defective outputs produced by plants are respectively 0.05, 0.04 and 0.02. If a pipe is selected from a day's total production and found to be defective, find out the probability that it came from 1st machine. **[Ans. 25/69]**
6. There are three identical boxes containing respectively 1 white and 3 red balls, 2 white and 1 red balls, 4 white and 3 red balls. One box is chosen at random and two balls are drawn: (*i*) find the probability that the balls are white and red, (*ii*) if the balls are white and red, what is the probability that they are from the second box? **[Ans. (*i*) 73/126; (*ii*) 28/73]**
7. You note that your officer is happy in 60% cases of your calls. You have also noticed that if he is happy, he accedes to your requests with a probability of 0.4, whereas if he is not happy, he accedes to your requests with a probability of 0.1. You call on him one day and he accedes to your request. What is the probability of his being happy?

   [**Hint:** $\frac{0.6 \times 0.4}{0.6 \times 0.4 + 0.4 \times 0.1}$] **[Ans. 6/7]**
8. Three persons A, B and C are being considered for the appointment as Vice-Chancellor of a University whose chances of being selected for the post are in the proportion 4 : 2 : 3 respectively. The probability that A, if selected will introduce democratisation in the University structure is 0.3, the corresponding probabilities for B and C doing the same are respectively 0.5 and 0.8. What is the probability that democratisation would be introduced in the University. Also find the probability that Vice Chancellor B introduced democratisation in the University. **[Ans. $\frac{23}{45}, \frac{5}{23}$]**
9. In a university, 30 per cent of the students doing a course in Statistics use the book authored by A_1, 45 per cent use the book authored by A_2, and 25 per cent use the book authored by A_3. The proportion of students who learnt about each of these books through their teachers are: $A_1 = 0.50$, $A_2 = 0.30$, and $A_3 = 0.20$. One of the students selected at random revealed that he learnt about the book he is using through the teachers. Find the probabilities that the book used is authored by A_1, A_2, and A_3, respectively. **[Ans. 0.45, 0.40, 0.15]**

MISCELLANEOUS SOLVED EXAMPLES

Example 112. A bag contains 7 white, 5 red and 8 black balls. Two balls are drawn at random. Find the probability that they will be white.

Solution: Total number of balls in the bag = 7 + 5 + 8 = 20.

2 balls can be drawn from 20 balls in $^{20}C_2$ ways.

2 white balls can be drawn from 7 white balls in $^{7}C_2$ ways.

$\therefore$ Required Probability $= \frac{^{7}C_2}{^{20}C_2} = \frac{21}{190}$

Example 113. A bag contains 8 white and 4 red balls. Five balls are drawn at random. What is the probability that 2 of them are red and 3 white?

Solution: Total number of balls in the bag = 8 + 4 = 12

Number of balls drawn = 5

5 balls can be drawn from 12 balls in ${}^{12}C_5$ ways.

2 red balls can be drawn from 4 red balls in ${}^{4}C_2$ ways.

3 white balls can be drawn from 8 white balls in ${}^{8}C_3$ ways.

$\therefore$ The number of favourable cases to 2 red and 3 white balls = ${}^{4}C_2 \times {}^{8}C_3$

$\therefore$ Required Probability $= \dfrac{{}^{4}C_2 \times {}^{8}C_3}{{}^{12}C_5} = \dfrac{14}{33}$

Example 114. A committee of 5 members is to be formed out of a group of 8 boys and 7 girls. Find the probability that in a committee— (*i*) there will be 3 boys and 2 girls and (*ii*) at least one girl.

Solution: (*i*) Probability of 3 boys and 2 girls is $= \dfrac{{}^{8}C_3 \times {}^{7}C_2}{{}^{15}C_5} = \dfrac{56}{143}$

(*ii*) Probability of at least one girl is:

Probability of no girl $= \dfrac{{}^{8}C_5}{{}^{15}C_5}$

$\therefore$ Probability of at least one girl $= 1 - \dfrac{{}^{8}C_5}{{}^{15}C_5} = 1 - \dfrac{8}{429} = \dfrac{421}{429}$.

Example 115. Five men in a company of 20 are graduates. If 3 men are picked out at random, what is the probability that (*i*) all are graduates and (*ii*) at least one being graduate?

Solution: (*i*) Probability of 3 graduates $= \dfrac{{}^{5}C_3}{{}^{20}C_3} = \dfrac{1}{114}$.

(*ii*) Probability of at least one graduate:

Probability of zero graduate $= \dfrac{{}^{15}C_3}{{}^{20}C_3} = \dfrac{91}{228}$.

$\therefore$ Probability of at least one graduate $= 1 - \dfrac{{}^{15}C_3}{{}^{20}C_3} = 1 - \dfrac{91}{228} = \dfrac{137}{228}$.

Example 116. If A and B be events in a sample space such that $P(A) = 0.3$, $P(\overline{B}) = 0.4$ and $P(A \cup B) = 0.8$, find

(*i*) $P(A \cap B)$ and

(*ii*) $P(\overline{A} \cap \overline{B})$

Solution: (*i*) We know that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$\Rightarrow \quad P(A \cup B) = P(A) + \{(1 - P(\overline{B}))\} - P(A \cap B) \qquad [\because P(B) = 1 - P(B^C)]$

$\Rightarrow \quad 0.8 = 0.3 + (1 - 0.4) - P(A \cap B)$

$\Rightarrow \quad 0.8 = 0.3 + 0.6 - P(A \cap B)$

$\therefore \quad P(A \cap B) = 0.9 - 0.8$

$= 0.1$

(*ii*) Now, $P(\overline{A} \cap \overline{B}) = P(A \cup B)^C = 1 - P(A \cup B)$ [By De Morgan's Law]

$= 1 - 0.8 = 0.2$

Example 117. One bag contains 4 white and 2 black balls. Another bag contains 3 white and 5 black balls. If one ball is drawn from each bag, find the probability that (*i*) both are white (*ii*) both are black and (*iii*) one is white and one is black.

Solution: (*i*) Probability of drawing a white ball from the 1st bag $= \dfrac{4}{6}$

Probability of drawing a white ball from the 2nd bag $= \dfrac{3}{8}$

Since, the events are independent, the probability that the balls are white

$= \dfrac{4}{6} \times \dfrac{3}{8} = \dfrac{1}{4}$

(*ii*) Probability of drawing a black ball from the 1st bag $= \dfrac{2}{6}$

Probability of drawing a black ball from the second bag $= \dfrac{5}{8}$

Since, the events are independent, the probability that both the balls are white

$= \dfrac{2}{6} \times \dfrac{5}{8} = \dfrac{5}{24}$

(*iii*) There are two possibilities, *viz*. (*a*) either 1st is white and the 2nd is black or (*b*) 1st is black and 2nd is white.

The probability that one is white and one is black.

$= P(WB) + P(BW)$

$= \left(\dfrac{4}{6}\right) \times \left(\dfrac{5}{8}\right) + \left(\dfrac{2}{6}\right)\left(\dfrac{3}{8}\right) = \dfrac{20}{48} + \dfrac{6}{48} = \dfrac{26}{48} = \dfrac{13}{24}$.

Example 118. A card is drawn from a pack of playing cards and then another card is drawn without the first being replaced. What is the probability of drawing (*i*) two aces (*ii*) two spades?

Solution: (*i*) The probability that the first card is an ace $= \dfrac{4}{52}$

When an ace has been drawn, there are three aces in 51 cards left.

$\therefore$ The probability that the second card should also be an ace $= \frac{3}{51}$

Hence, the probability that both are aces is $= \frac{4}{52} \times \frac{3}{51} = \frac{1}{221}$

(*ii*) The probability that the first card is a spade $= \frac{13}{52}$.

When a spade card has been drawn, there are 12 cards of spade in 51 cards left.

$\therefore$ The probability that the second card should also be a spade $= \frac{12}{51}$.

Hence, the probability that both are cards of spade is $= \frac{13}{52} \times \frac{12}{51} = \frac{1}{17}$.

Example 119. Three dice are rolled simultaneously. Find the probability of getting a total of (*i*) not more than 5 (*ii*) at least 15 and (*iii*) exactly 8.

Solution: If three dice are thrown together, the total number of exhaustive cases are:

$$= 6 \times 6 \times 6 = 216$$

Die I	*Die II*	*Die III*
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6

(*i*) The favourable cases showing a total of not more than 5 (*i.e.*, 3, 4, 5)

Number of cases favourable to a total of 3 are: (1+1+1) = 1

Number of cases favourable to a total of 4 are:

(1+2+1), (1+1+2), (2+1+1) = 3

Number of cases favourable to a total of 5 are:

(1+2+2), (2+1+2), (2+2+1), (1+3+1), (1+1+3), (3+1+1) = 6

Total number of cases favourable showing not more than 5 (*i.e.*, 3, 4, 5)

= 1+3+6 = 10.

$\therefore$ Probability of getting a total of not more than 5 $= \frac{10}{216} = \frac{5}{108}$

(*ii*) The favourable cases showing a total of at least 15 (*i.e.*, 15, 16, 17, 18)

Number of cases favourable to 15 are:

(3+6+6), (6+3+6), (6+6+3), (4+5+6), (4+6+5),

(5+4+6), (5+6+4), (6+4+5), (6+5+4), (5+5+5) = 10

Number of cases favourable to 16 are:

(4+6+6), (6+4+6), (6+6+4), (5+6+5), (5+5+6), (6+5+5) = 6

No. of cases favourable to 17 are: (5+6+6), (6+5+6), (6+6+5) = 3

No. of cases favourable to 18 are: (6+6+6) = 1

Total Number of cases favourable to at least 15, *i.e.*, 15,16,17 and 18

= 10+6+3+1=20

Probability of getting a total of at least 15 $= \frac{20}{216} = \frac{5}{54}$

(*iii*) The favourable cases showing a total of exactly 8 are:

(1+5+2), (1+2+5), (5+1+2), (5+2+1), (2+1+5), (2+5+1), (1+6+1), (1+1+6), (6+1+1), (2+2+4), (2+4+2), (4+2+2), (2+3+3), (3+2+3), (3+3+2), (3+1+4), (3+4+1), (1+3+4), (1+4+3), (4+3+1), (4+1+3) = 21

$\therefore$ Probability of getting a total of exactly 8 $= \frac{21}{216} = \frac{7}{72}$

Important Note:

The probability of getting sum from 3 to 18 can be easily ascertained by remembering the following probability distribution in case of three dice:

X:	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
P(X):	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$

Example 120. A can hit a target 4 times in 5 shots. B 3 times in 4 shots and C twice in 3 shots. They fire a volley. What is the probability that at least two shots hit?

Solution: There are four possibilities:

(*i*) A and B hit and C does not hit.

(*ii*) A and C hit and B does not hit.

(*iii*) B and C hit and A does not hit.

(*iv*) A, B and C hit the target.

(*i*) Probability in the 1st case $= \frac{4}{5} \times \frac{3}{4} \times \left(1 - \frac{2}{3}\right) = \frac{12}{60}$

(*ii*) Probability in the 2nd case $= \frac{4}{5} \times \frac{2}{3} \times \left(1 - \frac{3}{4}\right) = \frac{8}{60}$

(*iii*) Probability in the 3rd case $= \frac{3}{4} \times \frac{2}{3} \times \left(1 - \frac{4}{5}\right) = \frac{6}{60}$

(*iv*) Probability in the 4th case $= \frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} = \frac{24}{60}$

Since, these are mutually exclusive cases,

$\therefore$ Required Probability $= \frac{12}{60} + \frac{8}{60} + \frac{6}{60} + \frac{24}{60} = \frac{50}{60} = \frac{5}{6}$.

Example 121. A, B and C, in order, toss a coin. The first one to throw a head wins. If A starts first, find their respective chances of winning.

Solution: The chance of throwing a head with a single coin $= \frac{1}{2}$

The chance of not throwing a head with a single coin $= 1 - \frac{1}{2} = \frac{1}{2}$

If A is to win, he should throw a head in the 1st or 4th or 7th, ... throws.
If B is to win, he should throw a head in the 2nd or 5th or 8th, ... throws.
If C is to win, he should throw a head in the 3rd or 6th or 9th, ... throws.
The chances that a head is thrown in the Ist, 2nd, 3rd, 4th, 5th, 6th,........ throws are

$$\frac{1}{2}, \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3, \left(\frac{1}{2}\right)^5, \left(\frac{1}{2}\right)^6, ...$$

$\therefore$ A's chance $= \frac{1}{2} + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^7 + ... \infty$

$$\left[S_\infty = 1 + a + a^2 + ...\infty = \frac{1}{1-a} = \frac{\text{First Term}}{1 - \text{Common Ratio}}\right]$$

$$= \frac{1}{2} \cdot \left[\frac{1}{1 - \left(\frac{1}{2}\right)^3}\right] = \frac{1}{2} \times \frac{8}{7} = \frac{4}{7}$$

$$\text{B's chance} = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^8 + ...\infty = \left(\frac{1}{2}\right)^2 \cdot \left[\frac{1}{1 - \left(\frac{1}{2}\right)^3}\right] = \frac{1}{4} \times \frac{8}{7} = \frac{2}{7}$$

$$\text{C's chance} = 1 - P(A) - P(B) = 1 - \frac{4}{7} - \frac{2}{7} = \frac{1}{7}.$$

Example 122. An urn contains 5 red, 3 white and 4 black balls. Three balls are drawn at random one after the other. Find the following probabilities:

(*i*) E_1, all the three balls are red
(*ii*) E_2, one is red and two are white
(*iii*) E_3, at least one is white.

Solution: Total number of balls in the urn $= 5 + 3 + 4 = 12$

Number of ways of drawing 3 balls out of 12 $= {}^{12}C_3 = \frac{12 \times 11 \times 10}{3 \times 2 \times 1} = 220$

(*i*) Number of ways of drawing 3 red balls out of 5 $= {}^5C_3 = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$

$\therefore$ Required probability $= \frac{10}{220} = \frac{1}{22}$

(*ii*) Number of ways of drawing 1 red ball from 5 red balls and 2 white balls from 3 white balls $= {}^5C_1 \times {}^3C_2 = \frac{5 \times 3 \times 2}{2 \times 1} = 15$

$\therefore$ Required probability $= \frac{15}{220}$

(*iii*) Number of ways that no ball is white, *i.e.*, number of ways of drawing 3 non-white balls from 5 + 4 (= 9 balls) $= {}^9C_3 = \frac{9 \times 8 \times 7}{3 \times 2 \times 1} = 84$

$\therefore$ P (zero white ball) $= \frac{84}{220}$

P(at least one white) $= 1 -$ P (zero white ball) $= 1 - \frac{84}{220} = \frac{136}{220}$

Example 123. In each of a set of games, it is 2 to 1 in favour of the winner of the previous game, what is the chance that the player who wins the first game shall win at least three of the next four games.

Solution: The player who has won the first game has the following chance of winning or losing the other games:

(*i*) Of losing the first game and winning the remaining three

$$\frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{81}$$

(*ii*) Of losing the second game and winning the remaining three

$$\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{81}$$

(*iii*) Of losing the third game and winning the remaining three

$$\frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{4}{81}$$

(*iv*) Of losing the fourth game and winning the remaining three

$$\frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} = \frac{8}{81}$$

(*v*) Of winning all the four games

$$\frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{16}{81}$$

Since, these events are mutually exclusive, the required probability will be:

$$\frac{4}{81} + \frac{4}{81} + \frac{4}{81} + \frac{8}{81} + \frac{16}{81} = \frac{36}{81} = \frac{4}{9}$$

Example 124. The odds against a certain event are 5 to 2 and odds in favour of another event are 6 to 5. Find the probability that at least one of these events will happen.

Solution: Odds against first event :: 5 : 2

$\therefore$ The probability that the first event will not happen $= \frac{5}{5+2} = \frac{5}{7}$

Odds in favour of 2nd event :: 6 : 5

$\therefore$ The probability that the second event will not happen $= \frac{5}{6+5} = \frac{5}{11}$.

Since, the events are independent, the joint probability of not happening both the events

$$= \frac{5}{7} \times \frac{5}{11} = \frac{25}{77}.$$

Hence, the probability that at least one will happen $= 1 - \frac{25}{77} = \frac{52}{77}$.

Example 125. There are four hotels in a certain city. If 3 men check into hotels in a day, find the probability that each check into (*i*) a different hotel and (*ii*) the same hotel.

Solution: (*i*) Each check into a different hotel

Required Probability $= \frac{4}{4} \times \frac{3}{4} \times \frac{2}{4} = \frac{24}{64} = \frac{3}{8}$

(*ii*) Each check into the same hotel

Required Probability $= \frac{4}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{4}{64} = \frac{1}{16}$.

Example 126. Three newspapers A, B and C are published in a certain city. It is estimated from a survey of adult population, 20% read A, 16% read B, 14% read C, 8% read both A and B, 5% read both A and C, 4% read both B and C, 2% read all the three. What is the probability that a normally chosen person read at least one of the papers?

Solution: Given, $P(A) = 20\% = \frac{20}{100}$, $P(B) = 16\% = \frac{16}{100}$, $P(C) = 14\% = \frac{14}{100}$

$P(AB) = 8\% = \frac{8}{100}$, $P(AC) = 5\% = \frac{5}{100}$, $P(BC) = 4\% = \frac{4}{100}$; $P(ABC) = \frac{2}{100}$

$\therefore$ P(A+B+C) = Probability that the person read at least one of the papers.

= Probability that the person reads A, B or C

$= P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$

$$= \frac{20}{100} + \frac{16}{100} + \frac{14}{100} - \frac{8}{100} - \frac{5}{100} - \frac{4}{100} + \frac{2}{100} = \frac{35}{100} = 35\%$$

Example 127. There are 10 boys and 20 girls in a class; in which half boys and half girls have blue eyes. One representative is selected at random from the class. What is the probability that he is boy or his eyes are blue?

Solution: The given information in tabular form

	Boys	Girls	Total
Blue eyes	5	10	15
Not blue eyes	5	10	15
Total	10	20	30

Let A = Boy, $\therefore$ $P(A) = \frac{10}{30}$ and B = Blue eyes, $\therefore$ $P(B) = \frac{15}{30}$

$A \cap B$ = A boy with blue eyes

$\therefore$ $P(A \cap B) = \frac{5}{30}$

Hence, required probability = $P(A$ or $B)$

$$= P(A) + P(B) - P(A \cap B) = \frac{10}{30} + \frac{15}{30} - \frac{5}{30} = \frac{20}{30} = \frac{2}{3}$$

Example 128. In a box, there are 5 red, 3 blue and 2 white balls. Three balls are chosen randomly with replacement. Find the probability that:

(*i*) all three balls are red (*ii*) no ball is red

(*iii*) at least one ball is red (*iv*) balls are either red or blue.

Solution: (*i*) (3R)= P(R) . P(R). P(R) $= \frac{5}{10} \times \frac{5}{10} \times \frac{5}{10} = \frac{125}{1000} = \frac{1}{8}$

(*ii*) No ball is red, *i.e.*, all the three balls are non-red. Following are the possibilities of 3 non-red balls.

Blue	White	Probability
3	0	$= \frac{3}{10} \times \frac{3}{10} \times \frac{3}{10} = \frac{27}{1000}$
2	1	$= \frac{3}{10} \times \frac{3}{10} \times \frac{2}{10} + \frac{3}{10} \times \frac{2}{10} \times \frac{3}{10} + \frac{2}{10} \times \frac{3}{10} \times \frac{3}{10} = \frac{54}{1000}$
1	2	$= \frac{3}{10} \times \frac{2}{10} \times \frac{2}{10} + \frac{2}{10} \times \frac{3}{10} \times \frac{2}{10} + \frac{2}{10} \times \frac{2}{10} \times \frac{3}{10} = \frac{36}{1000}$

Required Probability $= \frac{27}{1000} + \frac{54}{1000} + \frac{36}{1000} = \frac{117}{1000}$

(*iii*) P (at least one ball is red) = 1 – P [no ball is red]

$= 1 - P$ [all 3 balls are non-red] $= 1 - \frac{117}{1000} = \frac{883}{1000}$

(*iv*) P (Either 3R or 3B) = P(3R) + P(3B)

= P(R).P(R). P(R) + P(B). P(B).P(B)

$$= \frac{5}{10} \times \frac{5}{10} \times \frac{5}{10} + \frac{3}{10} \times \frac{3}{10} \times \frac{3}{10} = \frac{125}{1000} + \frac{27}{1000} = \frac{152}{1000}$$

Example 129. Two factories manufacture the same machine parts. Each part is classified as having either 0, 1, 2 or 3 manufacturing defects. The joint probability for this is given below:

Manufacturer	Number of Defects			
	0	1	2	3
X	0.1250	0.0625	0.1875	0.1250
Y	0.0625	0.0625	0.1250	0.2500

(*i*) A part is observed to have no defect. What is the probability that it was produced by X manufacturer?

(*ii*) A part is known to have been produced by manufacturer X. What is the probability that the part has no defects?

(*iii*) A part is known to have two or more defects. What is the probability that it was manufactured by X?

(*iv*) A part is known to have one or more defects. What is the probability that it was manufactured by Y?

Solution: Let A, B, C and D denote part of having 0, 1, 2 and 3 defects.

(*i*) $P(X/A) = \dfrac{\text{Joint probability of zero defects by X}}{\text{Total joint probability of zero defects by X and Y}}$

$= \dfrac{0.1250}{0.1250+0.0625} = \dfrac{0.1250}{0.1875} = 0.667$

(*ii*) $P(A/X) = \dfrac{\text{Joint probability of zero defects by X}}{\text{Total joint probability of zero defects by X}} = \dfrac{0.1250}{0.5000} = 0.25$

(*iii*) $P(X/C \text{ or } D) = \dfrac{\text{Joint probability of 2 or 3 defects by X}}{\text{Joint probability of 2 or 3 defects by X and Y both}}$

$= \dfrac{0.3125}{0.6875} = 0.455$

(*iv*) $P(Y/B \text{ or } C \text{ or } D) = \dfrac{\text{Joint probability of 1 or 2 or 3 defects by Y}}{\text{Total joint probability of 1 or 2 or 3 defects both by X and Y}}$

$= \dfrac{0.4375}{0.8125} = 0.538$

Example 130. 4 letters to each of which corresponds an envelope are placed in the envelopes at random. What is the probability that (*i*) no letter is placed in the right envelope, (*ii*) all letters are placed in the right envelope, and (*iii*) all letters are not placed in right envelopes.

Solution: Total number of letters = 4

Total number of envelopes = 4

Total number of ways in which 4 letters can be put into different envelopes one in each = $4 \times 3 \times 2 \times 1 = 24$

(*i*) Let us mark the letters by A, B, C, D and envelopes by I, II, III, IV

Because A is not to go in I.

$\therefore$ If A goes in II the others will go in wrong envelopes as under

I	II	III	IV
B	A	D	C
C	A	D	B
D	A	B	C

If A goes in III the others will go in wrong envelopes as under

I	II	III	IV
D	C	A	B
B	D	A	C
C	D	A	B

Similarly, if A goes in IV the others will go wrong as under

I	II	III	IV
D	C	B	A
B	C	D	A
C	D	B	A

Thus, in all there are nine cases in which all the envelopes have wrong letters.

$\therefore$ Probability that no letter is placed in right envelope $= \dfrac{9}{24} = \dfrac{3}{8}$.

(*ii*) Since, there is only way in which correct letter can go in correct envelope.

$\therefore$ Probability that all the four letters go in right envelope $= \dfrac{1}{24}$

(*iii*) Continuing from part (*ii*), Prob. that all letters are not put in right envelops

$= 1 - \dfrac{1}{24} = \dfrac{23}{24}$

Example 131. The probabilities of X, Y and Z becoming managers are $\frac{4}{9}, \frac{2}{9}$ and $\frac{1}{3}$ respectively. The probabilities that the Bonus scheme will be introduced if *X*, *Y* and *Z* become managers are $\frac{3}{10}, \frac{1}{2}$ and $\frac{4}{5}$ respectively.

(*i*) What is the probability that the bonus scheme will be introduced?

(*ii*) If the bonus scheme has been introduced, what is the probability that the manager appointed was X?

Solution: Given, $P(X) = \dfrac{4}{9}$, $P(Y) = \dfrac{2}{9}$, $P(Z) = \dfrac{1}{3}$

$P(B/X) = \dfrac{3}{10}$, $P(B/Y) = \dfrac{1}{2}$, $P(B/Z) = \dfrac{4}{5}$

Putting the given information in the form of a table as follows:

Events (1)	Prior Probability (2)	Conditional Probability (3)	Joint Probability Cols. (2) × (3)
X	$P(X)=\frac{4}{9}$	$P(B/X)=\frac{3}{10}$	$\frac{4}{9}\times\frac{3}{10}$
Y	$P(Y)=\frac{2}{9}$	$P(B/Y)=\frac{1}{2}$	$\frac{2}{9}\times\frac{1}{2}$
Z	$P(Z)=\frac{1}{3}$	$P(B/Z)=\frac{4}{5}$	$\frac{1}{3}\times\frac{4}{5}$

(*i*)

$$P(B) = P(XB) + P(YB) + P(ZB)$$
$$= P(X).P(B/X) + P(Y).P(B/Y) + P(Z).P(B/Z)$$
$$= \frac{4}{9}\times\frac{3}{10}+\frac{2}{9}\times\frac{1}{2}+\frac{1}{3}\times\frac{4}{5}=\frac{12+10+24}{90}=\frac{46}{90} \text{ or } \frac{23}{45}$$

(*ii*) Using Bayes theorem, the required probability is:

$$P(X/B) = \frac{\text{Joint probability of the X}}{\text{Sum of the joint probabilities}} = \frac{\frac{4}{9}\times\frac{3}{10}}{\frac{4}{9}\times\frac{3}{10}+\frac{2}{9}\times\frac{1}{2}+\frac{1}{3}\times\frac{4}{5}}=\frac{6}{23}$$

Example 132. The probability that a person stopping at a petrol pump will ask to have his tyres checked is 0.12, the probability that he will ask to have his oil checked is 0.29 and the probability that he will ask to have both of them checked is 0.07. (*i*) What is the probability that a person who has oil checked will also have tyre checked? (*ii*) What is the probability that a person stopping at the petrol pump will have either tyres or oil checked? (*iii*) What is the probability that a person stopping at the petrol pump will have neither his tyres nor oil checked?

Solution: Let A denote the event that a person stopping at a petrol pump will have tyres checked and B denote the event that he will get his oil checked. Then we are given:

$$P(A) = 0.12,\ P(B) = 0.29,\ P(AB) = 0.07$$

(*i*) The probability that a person who has oil checked will also have tyre checked is given by:

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{0.07}{0.29} = 0.24$$

(*ii*) The probability that a person stopping at the pump will have either tyres checked or oil checked is given by:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$
$$= 0.12 + 0.29 - 0.07 = 0.41 - 0.07 = 0.34$$

(*iii*) The probability that a person stopping at the pump will have neither his tyres nor oil checked is given by:

$$P(\bar{A} \cap \bar{B}) = P(A \cup B)^c = 1 - [P(A+B)]$$
$$= 1 - [P(A) + P(B) - P(AB)] = 1 - [0.34] = 0.66$$

Example 133. A problem of Statistics is given to two students A and B whose chances of solving it independently are $\frac{1}{2}$ and $\frac{1}{3}$. What is the probability that: (*i*) the problem is solved, (*ii*) only one of them solve the problem, (*iii*) only A will solve the problem.

Solution: Given, $P(A)=\frac{1}{2}$, $P(B)=\frac{1}{3}$, $P(\bar{A})=\frac{1}{2}$, $P(\bar{B})=\frac{2}{3}$

(*i*) P(the problem will be solved) $= 1 - P(\bar{A}).P(\bar{B}) = 1-\frac{1}{2}\times\frac{2}{3}=1-\frac{1}{3}=\frac{2}{3}$

(*ii*) P(only one of them solves the problem) $= P(A\bar{B}) + P(B\bar{A})$

$$= P(A).P(\bar{B}) + P(B).P(\bar{A})$$
$$= \frac{1}{2}\times\frac{2}{3}+\frac{1}{3}\times\frac{1}{2}=\frac{2}{6}+\frac{1}{6}=\frac{3}{6}$$

(*iii*) P(only A will solve the problem) $= P(A\bar{B}) = P(A).P(\bar{B}) = \frac{1}{2}\times\frac{2}{3}=\frac{1}{3}$

IMPORTANT FORMULAE

1. Definition of Probability

(*i*) Classical Definition

$$P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of equally likely cases}} = \frac{m}{n}$$

(*ii*) Statistical Definition

$$P(A) = \lim_{n\to\infty}\frac{m}{n}$$

2. Theorems of Probability

(*i*) Addition Theorem

(*a*) When A and B are mutually exclusive events, then:

$$P(A + B) = P(A) + P(B)$$

(*b*) When A and B are not mutually exclusive events, then:

$$P(A + B) = P(A) + P(B) - P(AB)$$

(*ii*) Multiplication Theorem

(*a*) When A and B are independent events, then

$$P(AB) = P(A) \times P(B)$$

(*b*) When A and B are dependent events, then

$$P(AB) = P(A) . P(B/A)$$
$$= P(B) . P(A/B)$$

(iii) Bayes' Theorem

If an event B can only occur in combination with one of the two mutually exclusive events A_1 and A_2 and if B actually happens, then the probability that it was preceded by the particular event A_i is given by:

$$P(A_i / B) = \frac{P(A_i).P(B / A_i)}{\sum_{i=1}^{2} P(A_i).P(B / A_i)} \quad \text{where, } i = 1, 2.$$

QUESTIONS

1. Define probability and explain the importance of this concept in Statistics.
2. Explain various approaches to probability.
3. Explain the following:
   (i) Mutually Exclusive and Equally-likely Events.
   (ii) Simple and Compound Events.
   (iii) Independent and Dependent Events.
   (iv) Exhaustive and Complementary Events.
4. State and prove addition theorem of probability.
5. State and prove multiplication theorem of probability.
6. Explain the concept of conditional probability.
7. Explain with examples to the rules of addition and multiplication in the theory of probability.
8. State and prove Bayes' theorem.
9. Explain:
   (i) Sample space (ii) Probability of an event
   (iii) Equally likely cases (iv) Additive and Multiplicative rules
   (v) Bayes' theorem.
10. Explain with suitable examples the mathematical and statistical definitions of probability. Discuss the importance of probability in decision making.
11. State the addition and multiplication theorems of probability, with two different examples illustrating the application of these theorems.
12. State and prove multiplication theorem for two independent events. What will be the form of the theorem if the two events are not independent of one another.
13. State and prove addition theorem for two mutually exclusive events. What would be the form of the theorem if the events are not mutually exclusive?
14. Explain the following:
   (i) Dependent events (ii) Independent events
   (iii) Complementary events.

Probability Distributions—Binomial and Poission

8

◘ INTRODUCTION

In statistics, we study different types of distributions. They are broadly classified into two headings:

(1) Observed Frequency Distribution

(2) Theoretical or Probability Distribution

◘ (1) OBSERVED FREQUENCY DISTRIBUTION

Observed frequency distribution refers to those frequency distributions which are obtained by actual observations or experiments. For example, the observed frequency distribution of the marks obtained by 70 students of a class is as follows:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of Students:	5	15	20	25	5

The observed frequency distribution are generally analysed by using various statistical devices like averages, dispersion, skewness, etc.

◘ (2) THEORETICAL OR PROBABILITY DISTRIBUTION

Theoretical frequency distribution refers to those distributions which are not obtained by actual observations or experiments but are mathematically deduced under certain assumptions. Theoretical frequency distributions are also called **Probability Distributions or Expected Frequency Distribution.** For example, if four coins are tossed 160 times and the probability of getting a head is considered a success, then on the basis of theory of probability the expected frequency distribution will be as follows:

No. of Success (X)	Probability (p)	Expected Frequency
0	1/16	$160 \times 1/16 = 10$
1	4/16	$160 \times 4/16 = 40$
2	6/16	$160 \times 6/16 = 60$
3	4/16	$160 \times 4/16 = 40$
4	1/16	$160 \times 1/16 = 10$
	$\Sigma p = 1$	160

Thus, the theoretical frequency distributions are not based on actual observations but are mathematically deducted under certain assumptions.

Uses of Theoretical Frequency Distribution

Theoretical distribution play an important role in statistical theory. The main uses of theoretical distribution are as follows:

(1) Theoretical distributions are useful in analysing the nature of given distribution under certain assumptions.

(2) The expected frequencies obtained from the theoretical frequency distribution are useful for making logical decisions.

(3) Theoretical frequency distributions helps in comparing actual and expected frequencies and then to determine whether the difference between the two is significant or is due to fluctuations of sampling.

(4) Theoretical distribution helps in making predictions, projection and forecasting.

(5) Theoretical distributions are useful in solving many business and other problems. Poisson distribution is useful in making important decisions regarding quality control. Normal distribution helps in determining the stock of ready market garments of different sizes.

(6) In such cases where the actual experiments are not possible or in case of high cost involved in the collection of actual observation, theoretical frequency distribution can be substituted in place of observed frequency distributions.

From the foregoing discussion, it is thus clear that the study of theoretical frequency distributions is very useful. To quote **Meril and Fox, "Theoretical distributions play important role in statistical Theory".**

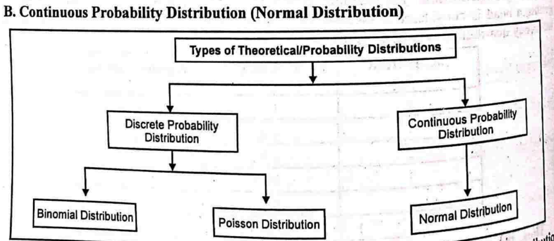
Types of Theoretical or Probability Distributions

The main type of theoretical distributions are:

A. Discrete Probability Distributions

- **(i) Binomial Distribution**
- **(ii) Poisson Distribution**

B. Continuous Probability Distribution (Normal Distribution)



In this chapter, we shall study only Binomial and Poisson Distribution. The normal distribution will be discussed in the next chapter.

(1) BINOMIAL DISTRIBUTION

Binomial distribution is a discrete probability distribution. This distribution was discovered by a Swiss Mathematician James Bernoulli. It is used in such situations where an experiment results in two possibilities - success and failure. Binomial distribution is a discrete probability distribution which expresses the probability of one set of two alternatives—success (p) and failure (q).

Definition of Binomial Distribution

Binomial distribution is defined and given by the following probability function:

$$P(X=x) = {}^nC_x \, q^{n-x} \cdot p^x$$

Where, p = probability of success, q = probability of failure = $1-p$, n = number of trials, $P(X=x)$ = probability of x successes in n trials.

By substituting the different values of X in the above probability function of the Binomial distribution, we can obtain the probability of 0, 1, 2, ..., n successes as follows:

Number of Success (X)	Probability of Success P(X = x)
0	${}^nC_0 \, q^{n-0} \cdot p^0 = q^n$
1	${}^nC_1 \, q^{n-1} \cdot p^1 = n \, q^{n-1} \cdot p^1$
2	${}^nC_2 \, q^{n-2} \cdot p^2 = \frac{n(n-1)}{2 \times 1} q^{n-2} \cdot p^2$
x ⋮ n	${}^nC_x \, q^{n-x} \cdot p^x$ ⋮ ${}^nC_n \, q^{n-n} \cdot p^n = p^n$

Conditions or Assumptions to Apply Binomial Distribution

Binomial distribution can be used only under the following conditions:

(1) Finite Number of Trials: Under binomial distribution, an experiment is performed under identical conditions for a finite and fixed number of trials, *i.e.*, number of trials is finite.

(2) Mutually Exclusive Outcomes: Each trial must result in two mutually exclusive outcomes—success or failure. For example, if a coin is tossed, then either the head (H) may turn up or the tail (T) may turn up.

(3) The probability of success in each trial is constant: In each trial, the probability of success, denoted by p remains constant. In other words, the probability of success in different trials does not change. For example, in tossing a coin, the probability of getting a head in each toss remains the same, *i.e.*, $p = P(H) = \frac{1}{2}$.

(4) Trials are independent: In binomial distribution, statistical independent among trials is assumed, *i.e.*, the outcome of any trial does not affect the outcomes of the subsequent trials.

Properties/Characteristics of Binomial Distribution

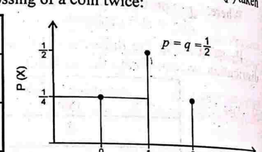
The following are the important properties or characteristics of binomial distribution:

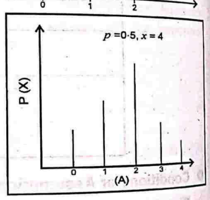
(1) Theoretical Frequency Distribution: The binomial distribution is a theoretical frequency distribution which is based on Binomial Theorem of algebra. With the help of this distribution, we can obtain the theoretical frequencies by multiplying the probability of success by the total number (N).

(2) Discrete Probability Distribution: The binomial distribution is a discrete probability distribution in which the number of successes 0, 1, 2, 3,......*n* are given in whole numbers and not in fractions.

(3) Line Graph: The binomial distribution can be presented graphically by means of a line graph. The number of successes (X) is taken on the X-axis and the probability of successes (*p*) taken on the Y-axis. The following line graph is based on tossing of a coin twice:

Number of Heads (X)	Probability P(X = *x*)
0	${}^2C_0\left(\frac{1}{2}\right)^2=\frac{1}{4}$
1	${}^2C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^1=\frac{1}{2}$
2	${}^2C_2\left(\frac{1}{2}\right)^2=\frac{1}{4}$

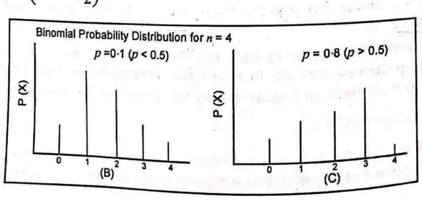




(4) Shape of Binomial Distribution: The shape of binomial distribution depends on the values of *p* and *q*.

(*i*) If $p = q = \frac{1}{2}$, then the binomial distribution is symmetrical (see figure A).

(*ii*) When $p \neq q \neq \frac{1}{2}$, the binomial distribution is skewed, *i.e.*, asymmetrical. It is positively skewed when $p < q\left(i.e., p<\frac{1}{2}\right)$ and negatively skewed when $p > q\left(i.e., p>\frac{1}{2}\right)$. See figures (B) and (C) given below:



(5) Main Parameters: The binomial distribution has two parameters *n* and *p*. The entire distribution can be known from these two parameters.

(6) Constants of Binomial Distribution: The constants of Binomial distribution are obtained by using the formula.

Mean $=(\overline{X})=np$

Moment coeff. of skewness $=\sqrt{\beta_1}=\frac{q-p}{\sqrt{npq}}$

Variance $=\sigma^2=npq$

$S.D.=\sigma=\sqrt{npq}$

Moment coeff. of Kurtosis $=\beta_2=3+\frac{1-6pq}{npq}$

[For Proof of $\overline{X}$ and σ^2 see Example 52]

(7) Uses: It has been found useful in those fields where the outcome is classified into success and failure. In other words, it is useful in coin experiment, dice throwing, manufacturing of items by a company, etc.

Applications of Binomial Distribution

The practical applications of binomial distribution are studied under the following headings:

(A) Application of Binomial Distribution Formula

(B) To Find n, p and q from $\overline{X}$ and σ

(C) To Find $\overline{X}$ and σ when n, p and q are given

(D) Fitting of Binomial Distribution.

(A) Application of Binomial Distribution Formula

When we are given the probability of occurrence of an event relating to a problem, *i.e.*, the value of *p* and *q* are given, then we can find the probability of the happening of the event exactly *x* times out of *n* trials by using the formula: $[P(X=x)={}^nC_x\,q^{n-x}\,p^x]$.

Example 1. A fair coin is tossed thrice. Find the probability of getting:

(*i*) exactly 2 Heads

(*ii*) at least 2 Heads

(*iii*) at the most 2 Heads

Solution: Let p = probability of getting head when a coin is tossed $=\frac{1}{2}$

q = the probability of tail $=\frac{1}{2}$

and $n = 3$, $P(X=x) = {}^nC_x\, q^{n-x} . p^x$

(*i*) $P(2H) = {}^3C_2\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2=3\times\frac{1}{2}\times\frac{1}{4}=\frac{3}{8}$

(*ii*) P (at least 2 Heads) $= P(2H) + P(3H)$

$$= {}^3C_2\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2 + {}^3C_3\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3$$

$$= 3\times\frac{1}{8} + 1\times\frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

(*iii*) P (at most 2 Heads) $= P(0H) + P(1H) + P(2H)$

$= 1 - P(3H)$

$$= 1 - {}^3C_3\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3$$

$$= 1 - 1\times\frac{1}{8} = 1 - \frac{1}{8} = \frac{7}{8}$$

Example 2. Four coins are tossed simultaneously. What is the probability of getting (*i*) No head (*ii*) No tail and (*iii*) Two heads only?

Solution: Let p = probability of getting head when a coin is thrown $= \frac{1}{2}$

$\therefore$ q = the probability of tail $= 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$

and $n = 4$ $P(X = x) = {}^nC_x\, q^{n-x}\,.\,p^x$

(*i*) $$P(0H) = {}^4C_0\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = 1\times\frac{1}{16} = \frac{1}{16}$$

(*ii*) $$P(0T) = P(4H) = {}^4C_4\left(\frac{1}{2}\right)^0.\left(\frac{1}{2}\right)^4 = 1\times\frac{1}{16} = \frac{1}{16}$$

(*iii*) $$P(2H) = {}^4C_2\,.\left(\frac{1}{2}\right)^2.\left(\frac{1}{2}\right)^2 = 6\times\frac{1}{16} = \frac{6}{16} = \frac{3}{8}$$

Example 3. Eight coins are thrown simultaneously. Find the probability of getting at least 6 heads.

Solution: Let p = probability of getting a head. q = probability of getting a tail.

Here, $p = \frac{1}{2}, q = \frac{1}{2}, n = 8,$

$$P(X = x) = {}^nC_x\, q^{n-x}\,.\,p^x$$

P (at least 6 heads) $= P(6H) + P(7H) + P(8H)$

$$= {}^8C_6\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 + {}^8C_7\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7 + {}^8C_8\left(\frac{1}{2}\right)^8$$

$$= 28\times\frac{1}{256} + 8\times\frac{1}{256} + 1\times\frac{1}{256}$$

$$= \frac{37}{256}$$

Example 4. The probability of a bomb hitting a target is $\frac{1}{5}$. Two bombs are enough to destroy a bridge. If six bombs are fired at the bridge, find the probability that the bridge is destroyed.

Solution: Let p = probability of a bomb hitting a target, q = probability of not hitting the target.

Here, $p = \frac{1}{5}$ $\quad\therefore\ q = \frac{4}{5}$ $\quad(\because\ q = 1 - p)$

Also, $n = 6$ $\quad P(X = x) = {}^nC_x\, q^{n-x}\,.\,p^x$

The bridge will be destroyed if two or more of 6 bombs hit it.

$\therefore$ Required probability $= P(2) + P(3) + P(4) + P(5) + P(6)$

$= 1 - [P(0) + P(1)]$

$$= 1 - \left[{}^6C_0\left(\frac{4}{5}\right)^6\left(\frac{1}{5}\right)^0 + {}^6C_1\left(\frac{4}{5}\right)^5\left(\frac{1}{5}\right)^1\right]$$

$$= 1 - \left[1\times\left(\frac{4}{5}\right)^6 + 6\times\frac{(4)^5}{(5)^6}\right] = 1 - \frac{4^6 + 6\times 4^5}{5^6}$$

$$= 1 - \frac{10240}{15625} = \frac{15625 - 10240}{15625} = \frac{5385}{15625} = 0.345$$

Example 5. A die is thrown 5 times. If getting an odd number is a success, what is the probability of getting (*i*) 4 successes (*ii*) at least 4 successes?

Solution: Total No. of cases in a die = 6

Favourable cases (for odd number) = 3

Let p = probability of getting a odd number

Here, $$p = \frac{3}{6} = \frac{1}{2}$$

$\therefore$ $$q = 1 - \frac{1}{2} = \frac{1}{2}$$

Also $n = 5$

$\therefore$ $$P(X = x) = {}^nC_x\, q^{n-x}\,.\,p^x$$

(*i*) $$P(4\text{ Successes}) = {}^5C_4\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = 5\times\frac{1}{32} = \frac{5}{32}$$

(*ii*) P(At least 4 successes) $= P(4) + P(5)$

$$= {}^5C_4\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 + {}^5C_5\left(\frac{1}{2}\right)^5$$

$$= 5\times\frac{1}{32} + 1\times\frac{1}{32} = \frac{5+1}{32} = \frac{6}{32} = \frac{3}{16}$$

Example 6. The incidence of occupational disease in an industry is such that the workers have a 20% chances of suffering from it. What is the probability that out of six workers chosen at random, four or more will suffer from disease?

Solution: Let p = probability of a man suffering from disease.

$$\therefore \quad p=\frac{20}{100}=\frac{1}{5}$$

$$\therefore \quad q=1-\frac{1}{5}=\frac{4}{5}$$

Also $n=6$

$$P(X=x)={}^{n}C_{x}\, q^{n-x}\cdot p^{x}$$

Required probability = P(4) + P(5) + P(6)

$$={}^{6}C_{4}\left(\frac{4}{5}\right)^{2}\left(\frac{1}{5}\right)^{4}+{}^{6}C_{5}\left(\frac{4}{5}\right)^{1}\left(\frac{1}{5}\right)^{5}+{}^{6}C_{6}\left(\frac{4}{5}\right)^{0}\left(\frac{1}{5}\right)^{6}$$

$$=15\times\frac{16}{15625}+6\times\frac{4}{15625}+\frac{1}{15625}$$

$$=\frac{240+24+1}{15625}=\frac{265}{15625}=\frac{53}{3125}=0.01696$$

Example 7. Assuming that half the population is vegetarian, so that the chance of an individual being vegetarian is 1/2 and assuming that 100 investigators each take 10 individuals to see whether they are vegetarian, how many investigators would you expect to report that three or less people were vegetarian?

Solution: Let p = probability of vegetarian $=\frac{1}{2}$

$$\therefore \quad q=1-\frac{1}{2}=\frac{1}{2},\ n=10,\ N=100$$

P (three or less people are vegetarian)

$$=P(X=0)+P(X=1)+P(X=2)+P(X=3)$$

$$={}^{10}C_{0}\left(\frac{1}{2}\right)^{10}\cdot\left(\frac{1}{2}\right)^{0}+{}^{10}C_{1}\left(\frac{1}{2}\right)^{9}\cdot\left(\frac{1}{2}\right)^{1}+{}^{10}C_{2}\left(\frac{1}{2}\right)^{8}\cdot\left(\frac{1}{2}\right)^{2}+{}^{10}C_{3}\left(\frac{1}{2}\right)^{7}\cdot\left(\frac{1}{2}\right)^{3}$$

$$=1\times\frac{1}{1024}+10\times\frac{1}{1024}+45\times\frac{1}{1024}+120\times\frac{1}{1024}=\frac{176}{1024}=\frac{11}{64}$$

No. of investigators who will report 3 or less vegetarian

$$=100\times\frac{11}{64}=17.2 \text{ or } 17 \text{ persons approx.}$$

Example 8. It is observed that 80% of television viewers watch **"Sas Bhi Kabhi Bahu Thee"** programme. What is probability that at least 80% of the viewers in a random sample of five watch this programme?

Solution: If viewing the programme is a success then

$$p=\frac{80}{100}=0.8,$$

and $$q=1-0.8=0.2$$

80% of 5 = 4. Therefore, we are to find the probability that 4 or 5 viewers watch the programme.

Required probability $={}^{5}C_{4}(0.2)^{(5-4)}(0.8)^{4}+{}^{5}C_{5}(0.2)^{0}(0.8)^{5}$

$$=5\times 0.2\times(0.8)^{4}+(0.8)^{5}$$

$$=(0.8)^{4}(1+0.8)$$

$$=0.4096\times 1.8$$

$$=0.73728$$

Example 9. If 8 ships out of 10 ships arrive safely, find the probability that at least one would arrive safely out of 5 ships selected at random.

Solution: p = probability of safe arrival of a ship.

So, $p=\frac{8}{10}=\frac{4}{5}$ and $q=1-p=1-\frac{4}{5}=\frac{1}{5}$

$$n=5$$

$$P(X=x)={}^{n}C_{x}\, q^{n-x}\cdot p^{x}$$

Probability of at least one would arrive safely means that 1 or 2 or 3 or 4 or 5 arrive safely.

P (at least one) = P(1) + P(2) + P(3) + P(4) + P(5)

= 1 − P(0)

$$\text{P (at least one)}=1-{}^{5}C_{0}\left(\frac{1}{5}\right)^{5}\cdot\left(\frac{4}{5}\right)^{0}$$

$$=1-\frac{1}{3125}=\frac{3124}{3125}$$

Example 10. If a die is thrown 6 times and getting 5 or 6 is considered a success, obtain the probability of getting 0, 1, 2, 3, 4, 5 or 6 successes.

Solution: Here, getting 5 or 6 is considered a success,

So, $$p=\frac{2}{6}=\frac{1}{3}$$

Hence, $$q=1-\frac{1}{3}=\frac{2}{3}$$

and $$n=6$$

$$P(X=x)={}^{n}C_{x}\, q^{n-x}\cdot p^{x}$$

No. of Successes	Probability
0	${}^6C_0\left(\frac{2}{3}\right)^6\left(\frac{1}{3}\right)^0 = \frac{64}{729}$
1	${}^6C_1\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^1 = \frac{192}{729}$
2	${}^6C_2\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^2 = \frac{240}{729}$
3	${}^6C_3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^3 = \frac{160}{729}$
4	${}^6C_4\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^4 = \frac{60}{729}$
5	${}^6C_5\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^5 = \frac{12}{729}$
6	${}^6C_6\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^6 = \frac{1}{729}$

Example 11. Out of 1,000 families with 4 children each, what percentage would you expect to have (*i*) at least one boy (*ii*) at the most 2 girls? Assume equal probabilities for boys and girls.

Solution: Let p = probability for a boy $= \frac{1}{2}$

q = probability for a girl $= \frac{1}{2}$ $\quad n = 4, N = 1000$

(*i*) **At least one boy:**

$$\text{P (at least one boy)} = P(1B) + P(2B) + P(3B) + P(4B)$$
$$= 1 - P(0B)$$
$$= 1 - {}^4C_0\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = 1 - \frac{1}{16} = \frac{15}{16}$$

Percentage of families with at least one boy $= \frac{15}{16} \times 100 = 93.75\%$

(*ii*) **At most 2 girls:**

$$\text{P(at most two girls)} = P(0G) + P(1G) + P(2G) = P(4B) + P(3B) + P(2B)$$
$$= {}^4C_4\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 + {}^4C_3\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 + {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2$$
$$= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16}$$

Percentage of such families $= \frac{11}{16} \times 100 = 68.75\%$

Example 12. The probability that an evening student will graduate is 0.8. Determine the probability that out of 5 students (*i*) none (*ii*) one (*iii*) at least one will graduate.

Solution: Let p = probability that a student will graduate

$$\therefore \; p = 0.8, q = 1 - p = 1 - 0.8 = 0.2, n = 5$$
$$P(X = x) = {}^nC_x\, q^{n-x} \cdot p^x$$

(*i*) P (none graduate) = P(OG) $= {}^5C_0\,(0.2)^5\,(0.8)^0 = 0.00032$

(*ii*) P (one graduate) = P(1G) $= {}^5C_1 \cdot (0.2)^4 \cdot (0.8)^1 = 0.0064$

(*iii*) P (at least one graduate) = 1 – P (0 Graduate) = 1 – 0.00032 = 0.99968

Example 13. In a binomial distribution consisting of 6 independent trials, the probability of two and three successes are 0.24576 and 0.08192 respectively. Find the parameter 'p' of the distribution.

Solution: Given, $n = 6$, P(2 successes) = 0.24576, P(3 successes) = 0.08192

According to B.D.,

$$P(X = x) = {}^nC_x\, q^{n-x} \cdot p^x$$
$$P(X = 2) = {}^6C_2\, q^4 \cdot p^2$$

$\because \quad q = 1 - p$

$\therefore \quad P(X=2) = {}^6C_2\,(1-p)^4 \cdot p^2 = 0.24576 \text{ (given)}$ ...(*i*)

$P(X = 3) = {}^6C_3\,(q)^3 \cdot (p)^3$

and $\quad q = 1 - p$

$\therefore \quad P(X=3) = {}^6C_3\,(1-p)^3 \cdot (p)^3 = 0.08192 \text{ (given)}$ ...(*ii*)

Dividing (*i*) by (*ii*)

$$\frac{{}^6C_2\,(1-p)^4\, p^2}{{}^6C_3\,(1-p)^3\, p^3} = \frac{0.24576}{0.08192}$$
$$\frac{15(1-p)^4 \cdot p^2}{20(1-p)^3\, p^3} = 3$$
$$\frac{3}{4}\frac{(1-p)}{p} = 3$$
$$3(1-p) = 12p$$
$$3 - 3p = 12p$$
$$15p = 3$$
$$p = \frac{3}{15} \qquad \therefore\; p = \frac{1}{5}$$

Thus, the value of the parameter 'p' equals $\frac{1}{5}$.

Example 14. A pair of dice is thrown 7 times. If getting a total of 7 is considered a success, find the probability of getting (*i*) no success (*ii*) 6 successes and (*iii*) at least 6 successes.

Solution: Let p = probability of getting a total of $7 = \frac{6}{36} = \frac{1}{6}$

$$\therefore q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}\, n = 7,$$

$$P(X = x) = {}^nC_x\, q^{n-x} \cdot p^x$$

(*i*) $P(0 \text{ Success}) = {}^7C_0 \left(\frac{5}{6}\right)^7 \cdot \left(\frac{1}{6}\right)^0 = \left(\frac{5}{6}\right)^7.$

(*ii*) $P(6 \text{ Successes}) = {}^7C_6 \left(\frac{5}{6}\right)^1 \cdot \left(\frac{1}{6}\right)^6 = 35 \times \left(\frac{1}{6}\right)^7$

(*iii*) P (at least 6 successes) = P(6) + P(7)

$$= {}^7C_6 \left(\frac{5}{6}\right)^1 \cdot \left(\frac{1}{6}\right)^6 + {}^7C_7 \left(\frac{5}{6}\right)^0 \cdot \left(\frac{1}{6}\right)^7$$

$$= 35 . \left(\frac{1}{6}\right)^7 + \left(\frac{1}{6}\right)^7 = 36 . \left(\frac{1}{6}\right)^7.$$

Example 15. The probability of a man hitting a target is $\frac{1}{4}$. (*i*) If he fires 7 times, what is the probability of his hitting the target at least twice? (*ii*) How many times must he fire so that the probability of his hitting target at least once is greater than 2/3?

Solution: (*i*) Probability of hitting the target $p = \frac{1}{4}$

Probability of not hitting the target $q = 1 - \frac{1}{4} = \frac{3}{4}$

Here, $n = 7 \quad P(X = x) = {}^nC_x\, q^{n-x} \cdot p^x$

$$P(X = 0) = {}^7C_0 \left(\frac{3}{4}\right)^7 \left(\frac{1}{4}\right)^0 = \left(\frac{3}{4}\right)^7 = \frac{2187}{16384}$$

$$P(X = 1) = {}^7C_1 \left(\frac{3}{4}\right)^6 \left(\frac{1}{4}\right)^1 = \frac{7 \times (3)^6}{(4)^7} = \frac{7 \times 729}{16384} = \frac{5103}{16384}$$

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$

$$= 1 - \frac{2187}{16384} - \frac{5103}{16384} = \frac{9094}{16384} = 0.555$$

(*ii*) Probability of at least one hit $= P(X \geq 1) = 1 - P(X = 0)$

$$P(X = 0) = {}^nC_0 \left(\frac{3}{4}\right)^n \left(\frac{1}{4}\right)^0 = \left(\frac{3}{4}\right)^n$$

$$P(X \geq 1) = 1 - \left(\frac{3}{4}\right)^n$$

It is required that $P(X \geq 1)$ should be greater than $\frac{2}{3}$.

Hence, $1 - \left(\frac{3}{4}\right)^n > \frac{2}{3}$

$$1 - \frac{2}{3} > \left(\frac{3}{4}\right)^n$$

$$\frac{1}{3} > \left(\frac{3}{4}\right)^n$$

Taking log on both sides,

$$\log\left(\frac{1}{3}\right) > n \log\left(\frac{3}{4}\right)$$

$$\log 1 - \log 3 > n (\log 3 - \log 4)$$

$$0 - 0.4771 > n (0.4771 - 0.6021)$$

$$-0.4771 > n (-0.125)$$

$$0.4771 < 0.125n$$

$$\frac{0.4771}{0.125} < n$$

or $n > 3.82$

$n = 4$ as it cannot be fractional.

EXERCISE 8.1

1. A fair coin is tossed six times. What is the probability of obtaining four or more heads. **[Ans.** 11/32**]**
2. A die is thrown 4 times. Getting a number greater than 2 is a success. Find the probability of (*i*) exactly 1 success (*ii*) less than 3 successes (*iii*) more than 3 successes. **[Ans. (*i*) 0.0988, (*ii*) 0.4074, (*iii*) 0.1975]**
3. One ship out of 10 was sunk on an average in making certain voyage. Find the probability that out of 5 ships, 4 would arrive safely.
   [Hint: $p = \frac{9}{10}$, $q = \frac{1}{10}$**]** **[Ans.** $\frac{32805}{1,00,000}$**]**
4. If hens of a certain breed lay eggs on 5 days a week on an average, find how many days during a session of 100 days a poultry keeps with 5 hens of this breed, will expect to receive at least 4 eggs? **[Ans.** $55.77 \approx 56$ **]**

5. Eight coins are tossed simultaneously. Show that the probability of obtaining 6 heads is $\frac{7}{64}$.

6. Out of 320 families with 5 children each, what percentage would be expected to have (*i*) 2 boys and 3 girls (*ii*) at least one boy? Assume equal probability for boys and girls?
[Ans. (*i*) 31.25% (*ii*) 96.825%]

7. In a binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0.4096 and 0.2048. Find the parameter 'p' of the distribution.
[Ans. $p=0.2$]

8. The chances of catching cold by workers working in ice factory during winter are 20%. What is the probability that of 5 workers, 4 or more will catch cold.
[Ans. 0.0067]

9. An experiment succeeds twice as many times as it fails. Find the chance that in 6 trials, there will be at least 5 successes.
[Ans. $\frac{256}{729}$]

10. A lot of manufactured items contain 20 per cent defectives. A sample of 10 items from this lot is chosen at random. What is the probability the sample contains at most 3 defective items?
[Ans. 0.8791]

11. The probability of failure in Chemistry practical examination is 60%. If 125 batches of 5 students each take the examination, in how many batches 3 or more students would pass?
[Ans. 992/3125, 39.68 $\simeq$ 40 batches]

12. Probability that a bomb dropped from a plane hits the target is 0.4. Two bombs can destroy a bridge. If in all 6 bombs were dropped, find the probability that the bridge will be destroyed.
[Ans. 0.76672]

13. A pair of dice is thrown 7 times. If getting a total of 9 is considered a success, what is the probability of at most 6 successes?
[Ans. $1-\frac{1}{9^7}$]

14. How many tosses of a coin are needed so that the probability of getting at least one head is 0.875 ?
[Ans. $n=3$]

▶ (B) To Find n, p and q from $\overline{X}$ and σ

When we are given the mean ($\overline{X}$) and variance (σ^2) or S.D. (σ) of the binomial distribution, then we can find out n, p and q. The following examples will illustrate the procedure:

Example 16. The mean of a binomial distribution is 20 and standard deviation is 4. Find n, p and q.

Solution: In a B.D., Mean $= np$

$$\text{S.D.} = \sqrt{npq}$$

$$\therefore \quad \overline{X} = np = 20 \quad \text{...(i)}$$

$$\sigma = \sqrt{npq} = 4 \quad \text{...(ii)}$$

Squaring both sides,

$$\Rightarrow \quad \sigma^2 = npq = 16 \quad \text{...(iii)}$$

Dividing (*iii*) by (*i*)

$$\frac{npq}{np} = \frac{16}{20}$$

$$\Rightarrow \quad q = \frac{16}{20} = \frac{4}{5}$$

$$\therefore \quad p = 1 - q = 1 - \frac{4}{5} = \frac{1}{5}$$

Putting the value of p in (*i*)

$$n \times \frac{1}{5} = 20$$

$$\Rightarrow \quad n = 100$$

Hence, $n = 100,\ p = \frac{1}{5},\ q = \frac{4}{5}$

Example 17. Obtain the mean and standard deviation of a binomial distribution for which $P(X=3) = 16\,P(X=7)$ and $n = 10$.

Solution:

$$P(X=3) = {}^{10}C_3\, q^{10-3} p^3 = {}^{10}C_3\, q^7 p^3$$

$$P(X=7) = {}^{10}C_7\, q^{10-7} p^7 = {}^{10}C_7\, q^3 p^7$$

As per the question,

$${}^{10}C_3\, q^7 p^3 = 16\ {}^{10}C_7\, q^3 p^7$$

$$\Rightarrow \quad q^7 p^3 = 16 q^3 p^7 \qquad [\because\ {}^{10}C_3 = {}^{10}C_7]$$

$$\Rightarrow \quad q^4 = 16 p^4 \quad \Rightarrow \quad (q)^4 = (2p)^4 \quad \Rightarrow \quad q = 2p$$

In a binomial distribution

$$p + q = 1 \quad \Rightarrow \quad p + 2p = 1 \quad \Rightarrow \quad p = 1/3$$

$$\therefore \quad q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$$

$$\therefore \quad \text{Mean} = np = \frac{10}{3}$$

$$\text{SD} = \sqrt{npq} = \sqrt{\frac{10}{3} \times \frac{2}{3}} = \frac{\sqrt{20}}{3}$$

Example 18. Is there any fallacy in the statement? The mean of a binomial distribution is 20 and its standard deviation is 7.

Solution: In a B.D., Mean $= np$

$$\text{S.D.} = \sqrt{npq}$$

$$\therefore \quad \overline{X} = np = 20 \quad \text{...(i)}$$

$$\sigma = \sqrt{npq} = 7 \quad \text{...(ii)}$$

Squaring both sides,

$$\Rightarrow \quad \sigma^2 = npq = 49 \quad \text{...(iii)}$$

Dividing (*iii*) by (*i*)

$$\frac{npq}{np}=\frac{49}{20}$$

$\therefore$ $q=\frac{49}{20}=2.45>1$, which is impossible as $p+q=1$

Hence, the statement is wrong.

Example 19. Find the probability of 5 successes in a binomial distribution whose mean and variance are respectively 6 and 2.

Solution: In a B.D., we have

$$\text{Mean}=np=6 \quad ...(i)$$

$$\text{Variance}=npq=2 \quad ...(ii)$$

Dividing (*ii*) by (*i*)

$$\frac{npq}{np}=\frac{2}{6}$$

$$\therefore \quad q=\frac{1}{3}$$

$$p=1-q=1-\frac{1}{3}=\frac{2}{3}$$

Substituting the value of p in (*i*)

$$n\times\frac{2}{3}=6$$

$$\therefore \quad n=9$$

Here, $n=9,\ p=\frac{2}{3}, q=\frac{1}{3}$

Now, $P(X=5)={}^9C_5\left(\frac{1}{3}\right)^4\cdot\left(\frac{2}{3}\right)^5=126\times\frac{32}{19683}=0.2048$

▶ **(C) To find $\overline{X}$ and σ when n, p and q are given.**

Example 20. If the probability of a defective bolt is 0.1, find (*i*) the mean and (*ii*) standard deviation for the distribution of defective bolts in a total of 500. Also find the coefficient of skewness and kurtosis

Solution: Given, $p=0.1 \ \therefore\ q=1-0.1=0.9,\ n=500$

(*i*) $\text{Mean}=np=500\times0.1=500\times\frac{1}{10}=50$

(*ii*) $\text{S.D.}=\sigma=\sqrt{npq}=\sqrt{500\times0.1\times0.9}=6.70$

(*iii*) Coefficient of Skewness $(\sqrt{\beta_1})=\frac{q-p}{\sqrt{npq}}=\frac{0.9-0.1}{\sqrt{500\times0.1\times0.9}}=\frac{0.8}{6.70}=0.119$

(*iv*) Coefficient of Kurtosis $(\beta_2)=3+\frac{1-6pq}{npq}=3+\frac{1-6(0.1)(0.9)}{500\times0.1\times0.9}=3.010$

Example 21. Find the mean and the standard deviation of the number of heads in 100 tosses of a fair coin.

Solution: Given, $n=100,\ P(H)=p=\frac{1}{2},\ q=\frac{1}{2}$

$$\therefore \quad \text{Mean}=np=100\times\frac{1}{2}=50$$

$$S.D.=\sqrt{npq}=\sqrt{100\times\frac{1}{2}\times\frac{1}{2}}=\sqrt{25}=5$$

EXERCISE 8.2

1. The mean and standard deviation of a binomial distribution are 2 and 1 respectively. Calculate n, p and q. **[Ans.** $n=4, p=1/2, q=1/2$**]**
2. Find the parameters of a binomial distribution for which mean = 4 and variance = 3. **[Ans.** $n=16, p=1/4$**]**
3. Is there any inconsistency in the statement. "The mean of a B.D. is 80 and S.D. is 8." If no inconsistency is found, what shall be the values of p, q and n? **[Ans.** $\frac{1}{5},\frac{4}{5}$, 400**]**
4. Find the probability of 3 successes in a binomial distribution whose mean and variance are respectively 2 and $\frac{3}{2}$. **[Ans.** 0.2076**]**
5. For a binomal distribution the mean is 6 and the standard deviation is $\sqrt{2}$. Write the terms of the binomial distribution.

   [Hint: $n=9, p=2/3, q=1/3$**]** **[Ans.** $\left(\frac{1}{3}\right)^9, 9\left(\frac{2}{3}\right)\cdot\left(\frac{1}{3}\right)^8$,.....**]**
6. A discrete random variable X has mean equal to 6 and variance equal to 2. If it is assumed that the underlying distribution X is binomial, what is the probability that $5\le X\le 7$? **[Ans.** 0.712**]**
7. If the probability of a defective bolt is 10%. Find (*i*) mean (*ii*) standard deviation (*iii*) moment coefficient of skewness and (*iv*) moment coefficient of kurtosis for the distribution of defective bolts in a total of 400. **[Ans.** $\overline{X}=40,\ \sigma^2=36, \sqrt{\beta_1}=0.133, \beta_2=3.013$**]**
8. For a B.D., the parameters n and p are 16 and $\frac{1}{2}$. Find the mean and S.D. **[Ans.** $\overline{X}=8,\ \sigma=2$**]**
9. An unbiased coin is tossed ten times. Find the mean and the standard deviation. **[Ans.** $\overline{X}=5, \sigma=\sqrt{2.5}$**]**
10. The mean and variance of a binomial distribution are 3 and 2 respectively find the probability that the variate takes values: (*i*) less than or equal to 2 (*ii*) greater than or equal to 7. **[Ans.** $p=\frac{1}{3}, q=\frac{2}{3}, n=9$, 0.377, 0.0083**]**
11. If on an average 8 ships out of 10 arrive safely at a port, find the mean and S.D. of the number of ships arriving safely out of a total of 1600 ships. **[Ans.** $\overline{X}=1280, \sigma=16$**]**

▶ (D) Fitting of Binomial Distribution

The following procedure is adopted while fitting a binomial distribution to the observed data:

(*i*) Determine the value of p and q from the given information.

(*ii*) Note the value of n and N, where n is the number of trials in an experiment and N is the total number of trials in all the experiments.

(*iii*) Find the probability of all possible number of successes coming out of a given experiment.

(*iv*) Multiply these probabilities by N and the result will the required expected frequencies.

The following examples illustrate the fitting of binomial distribution:

Example 22. Four coins were tossed 160 times and the following results were obtained:

No. of heads:	0	1	2	3	4
Frequency:	17	52	54	31	6

Fit a binomial distribution under the assumption that the coins are unbiased.

Solution: Under the assumption that the coins are unbiased; the probability of head (p) and tail (q) are $\frac{1}{2}$ and $\frac{1}{2}$.

In this case, $n=4,\ N=160$

The probability of 0, 1, 2, 3, 4 heads will be given by

$$P(X=x)={}^{n}C_{x}\ q^{n-x}\cdot p^{x}$$

In order to obtain the expected frequencies, we will have to multiply each probability by N.

The expected frequencies will be obtained as follows:

Number of heads (n)	**Expected Frequency** $N\times {}^{n}C_{x}\ q^{n-x}\cdot p^{x}$
0	$160\times {}^{4}C_{0}\left(\frac{1}{2}\right)^{4}\cdot\left(\frac{1}{2}\right)^{0}=10$
1	$160\times {}^{4}C_{1}\left(\frac{1}{2}\right)^{3}\cdot\left(\frac{1}{2}\right)^{1}=40$
2	$160\times {}^{4}C_{2}\left(\frac{1}{2}\right)^{2}\cdot\left(\frac{1}{2}\right)^{2}=60$
3	$160\times {}^{4}C_{3}\left(\frac{1}{2}\right)^{1}\cdot\left(\frac{1}{2}\right)^{3}=40$
4	$160\times {}^{4}C_{4}\left(\frac{1}{2}\right)^{0}\cdot\left(\frac{1}{2}\right)^{4}=10$

Example 23. Fit a binomial distribution to the following data:

X :	0	1	2	3	4
f:	28	62	46	10	4

Solution: To fit a binomial distribution to the data, we need the values of n, N, p and q

X :	0	1	2	3	4	
f:	28	62	46	10	4	$\Sigma f=150$
fX:	0	62	92	30	16	$\Sigma fX=200$

Here, $n=4,\ N=150$ (given)

$$\overline{X}=\frac{\Sigma fX}{N}=\frac{200}{150}=\frac{4}{3}$$

$$\because \quad np=\overline{X}=\frac{4}{3} \Rightarrow 4p=\frac{4}{3} \Rightarrow p=\frac{1}{3}$$

$$\therefore \quad q=1-\frac{1}{3}=\frac{2}{3}$$

The probability of 0, 1, 2, 3 and 4 will be given by

$$P(X=x)={}^{n}C_{x}\ q^{n-x}\cdot p^{x}$$

The expected frequencies are obtained by multiplying the probability by N as follows:

X	fe (X)
0	$150\times {}^{4}C_{0}\left(\frac{2}{3}\right)^{4}\cdot\left(\frac{1}{3}\right)^{0}=29.62\approx 30$
1	$150\times {}^{4}C_{1}\left(\frac{2}{3}\right)^{3}\cdot\left(\frac{1}{3}\right)^{1}=59.26\approx 59$
2	$150\times {}^{4}C_{2}\left(\frac{2}{3}\right)^{2}\cdot\left(\frac{1}{3}\right)^{2}=44.44\approx 44$
3	$150\times {}^{4}C_{3}\left(\frac{2}{3}\right)^{1}\left(\frac{1}{3}\right)^{3}=14.81\approx 15$
4	$150\times {}^{4}C_{4}\left(\frac{2}{3}\right)^{0}\cdot\left(\frac{1}{3}\right)^{4}=1.85\approx 2$

Example 24. Four perfect dice were thrown 112 times and the number of times 1, 3 or 5 were thrown as under:

Number of dice Showing 1, 3 or 5:	0	1	2	3	4
Frequency:	10	25	40	30	7

Fit a Binomial Distribution.

Solution: Under the assumption that dice are perfect, the probability of getting 1, 3 or 5 $(p)=\frac{3}{6}=\frac{1}{2}$ and $q=\frac{1}{2}$.

Here, $n=4$, N = 112.

The probability of 0, 1, 2, 3, 4 successes will be given by

$$P(X=x)={}^nC_x\, q^{n-x}\,.\,p^x$$

The expected frequencies are obtained by multiplying $P(X)$ with N, *i.e.*, $N.P(X)$. These are shown below:

Numbers 1, 3 or 5 (X)	Expected Frequency $N\times {}^nC_x\, q^{n-x}\,.\,p^x$
0	$112\times {}^4C_0\left(\frac{1}{2}\right)^4.\left(\frac{1}{2}\right)^0=112\times\frac{1}{16}=7$
1	$112\times {}^4C_1\left(\frac{1}{2}\right)^3.\left(\frac{1}{2}\right)^1=112\times\frac{4}{16}=28$
2	$112\times {}^4C_2\left(\frac{1}{2}\right)^2.\left(\frac{1}{2}\right)^2=112\times\frac{6}{16}=42$
3	$112\times {}^4C_3\left(\frac{1}{2}\right)^1.\left(\frac{1}{2}\right)^3=112\times\frac{4}{16}=28$
4	$112\times {}^4C_4\left(\frac{1}{2}\right)^0.\left(\frac{1}{2}\right)^4=112\times\frac{1}{16}=7$

Example 25. A survey of 800 families with four children each revealed the following distribution:

No. of Boys:	0	1	2	3	4
No. of Families:	42	178	290	226	64

Fit a Binomial Distribution under the hypothesis that male and female births are equally probable.

Solution: Under the assumption that male and female births are equally probable, the probability of male birth, $p=\frac{1}{2}$

$\therefore\quad q=1-\frac{1}{2}=\frac{1}{2}$

In this case, $n=4$, $N=800$

The probability of 0, 1, 2, 3, 4 boy's will be given by

$$P(X=x)={}^nC_x\, q^{n-x}\,.\,p^x$$

The expected frequencies will be obtained by multiplying $P(X)$ with N, *i.e.*, $N\times P(X)$. These are given as follows:

Number of Boys (X)	Expected frequency $N\times {}^nC_x\, q^{n-x}\,.\,p^x$
0	$800\times {}^4C_0\left(\frac{1}{2}\right)^4.\left(\frac{1}{2}\right)^0=800\times\frac{1}{16}=50$
1	$800\times {}^4C_1\left(\frac{1}{2}\right)^3.\left(\frac{1}{2}\right)^1=800\times\frac{4}{16}=200$
2	$800\times {}^4C_2\left(\frac{1}{2}\right)^2.\left(\frac{1}{2}\right)^2=800\times\frac{6}{16}=300$
3	$800\times {}^4C_3\left(\frac{1}{2}\right)^1.\left(\frac{1}{2}\right)^3=800\times\frac{4}{16}=200$
4	$800\times {}^4C_4\left(\frac{1}{2}\right)^0.\left(\frac{1}{2}\right)^4=800\times\frac{1}{16}=50$

Example 26. 8 unbiased coins are tossed 256 times. Find the expected frequencies of success (getting a head) and tabulate the results obtained. Calculate the mean and standard deviations of the number of heads.

Solution: We are given:

$n=8$, N = 256

$$p=\text{Probability of success (head)}=\frac{1}{2}$$

$$\therefore\quad q=1-\frac{1}{2}=\frac{1}{2}$$

According to B.D.

$$P(X=x)={}^nC_x\, q^{n-x}\,.\,p^x$$

$$f_e(x)=N.P(X)$$

Thus the expected frequencies are tabulated as:

Number of heads	Expected Frequency
0	$256.\ {}^8C_0\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0=1$
1	$256.\ {}^8C_1\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^1=8$
2	$256.\ {}^8C_2\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2=28$
3	$256.\ {}^8C_3\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3=56$
4	$256.\ {}^8C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^4=70$

5	$256.\ {}^8C_5\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5 = 56$
6	$256.\ {}^8C_6\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 = 28$
7	$256.\ {}^8C_7\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7 = 8$
8	$256.\ {}^8C_8\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^8 = 1$

$$\text{Mean} = np = 8\times\frac{1}{2} = 4$$

$$\text{S.D.} = \sqrt{npq} = \sqrt{8\times\frac{1}{2}\times\frac{1}{2}}$$

$$= \sqrt{2} = 1.4142$$

Example 27. Six dice are thrown 729 times. How many times do you expect at least 3 dice to show a five or six?

Solution: Given $N = 729$, $n = 6$

The probability of getting either 5 or 6 $= p = \frac{1}{6}+\frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

The probability of not getting 5 or 6 $= q = 1-\frac{1}{3} = \frac{2}{3}$

Thus $p = \frac{1}{3}$, $q = \frac{2}{3}$

P (at least 3 dice to show 5 or 6) = P(3) + P(4) + P(5) + P(6)

$$= {}^6C_3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^3 + {}^6C_4\left(\frac{2}{3}\right)^2\cdot\left(\frac{1}{3}\right)^4 + {}^6C_5\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^5 + {}^6C_6\left(\frac{2}{3}\right)^0\cdot\left(\frac{1}{3}\right)^6$$

$$= 20.\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^3 + 15.\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^4 + 6.\left(\frac{2}{3}\right)^1\cdot\left(\frac{1}{3}\right)^5 + 1.\left(\frac{1}{3}\right)^6$$

$$= 20\times\frac{8}{729} + 15\times\frac{4}{729} + 6\times\frac{2}{729} + \frac{1}{729}$$

$$= \frac{1}{729}[160+60+12+1]$$

$$= \frac{233}{729}$$

Hence, out of 729, the number of times we expect at least 3 dice to show five or six

$$= 729\times\frac{233}{729} = 233$$

EXERCISE 8.3

1. Five perfect dice are thrown together for 96 times. The number 4, 5 or 6 was actually thrown in the experiment are given below:

No. of dice showing 4, 5 or 6:	0	1	2	3	4	5
Frequency:	2	8	22	35	24	5

Fit a Binomial Distribution and calculate the expected frequencies. **[Ans.** 3, 15, 30, 30, 15, 3]

2. The following data gives the number of seeds germinating out of 1000 damp filters for 80 sets of seeds. Fit a BD to these data :

X:	0	1	2	3	4	5	6	7	8	9	10
f:	6	20	28	12	8	6	0	0	0	0	0

[Ans.: Expected frequencies: 6.9. 19.1, 24.0, 17.8, 8.6, 2.9, 0.7, 0.1, 0, 0, 0.]

3. Four coins were tossed 200 times. The number of tosses showing 0, 1, 2, 3 and 4 heads were observed as under:

No. of heads:	0	1	2	3	4
No. of tosses:	15	35	90	40	20

Fit a Binomial Distribution to these observed results. **[Ans.** 12.5, 50, 75, 50, 12.5]

4. The screws produced by a certain machine were checked by examining samples of 128. The following table shows the distribution of 128 samples according to the number of defective items they contained:

No. of defectives in a sample of 128 :	0	1	2	3	4	5	6	7	Total
No. of samples:	7	6	19	35	30	23	7	1	128

Fit a binomial distribution and find the expected frequencies if the chance of machine being defective is $\frac{1}{2}$. Find the mean and variance of the fitted distribution.

[Ans. 1,7,21,35,21,7,1; $\overline{X} = 3.5, \sigma^2 = 1.75$]

5. 5 coins are tossed 128 times. What is the probability of getting 3 or more heads and find out the expected frequencies of 3 or more heads. **[Ans.:** (*i*) 16/32, (*ii*) 64]

(2) POISSON DISTRIBUTION

Poisson distribution is a discrete probability distribution and it is widely used in statistical work. This distribution was developed by a French Mathematician Dr. Simon Denis Poisson in 1837 and the distribution is named after him. The Poisson distribution is used in those situations where the probability of the happening of an event is very small, *i.e.*, the event rarely occurs. For example, the probability of defective items in a manufacturing company is very small, the probability of occuring earthquake in a year is very small, the probability of the accidents on a road is very small, etc. All these are examples of such events where the probability of occurrence is very small.

Poisson Distribution as Limiting Form of Binomial Distribution

Poisson distribution is derived as a limiting form of binomial distribution under certain conditions:

(1) n, the number of trials is infinitely large, *i.e.*, $n \to \infty$.

(2) p, the probability of success is very small and q, the probability of failure is very large, *i.e.*, $p \to 0, q \to 1$.

(3) The average number of successes (np) is equal to a positive finite quantity (m) *i.e.*, $np = m$, where, m is the parameter of the distribution.

Definition of Poisson Distribution

From binomial equation:

$$^nC_x = \frac{n(n-1)(n-2)...(n-x+1)}{x!} p^x q^{n-x}$$

Since, $np = m \Rightarrow p = \frac{m}{n} \therefore q = \left(1 - \frac{m}{n}\right)$

$$= \frac{n(n-1)(n-2)...(n-x+1)}{n^x . x!} n^x p^x q^{n-x}$$

$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{x-1}{n}\right) q^{n-x} . \frac{(np)^x}{x!}$$

$$= \left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{x-1}{n}\right)\left(1-\frac{m}{n}\right)^{n-x}\right] . \frac{m^x}{x!}$$

When, $n \to \infty$

$$= e^{-m} . \frac{m^x}{x!} = \frac{e^{-m} m^x}{x!}$$

Poisson distribution is defined and given by the following probability function:

$$P(X = x) = e^{-m} . \frac{m^x}{x!}$$

Where, $P(X = x)$ = probability of obtaining x number of success

$m = np$ = parameter of the distribution

$e = 2.7183$ (base of natural logarithms)

By substituting the different values of X in the above probability function of the Poisson distribution, we can obtain the probability of 0, 1, 2..... X successes as follows:

Number of Success (X)	Probability P(X)
0	$e^{-m} . \frac{m^0}{0!} = e^{-m}$
1	$e^{-m} . \frac{m^1}{1!} = me^{-m}$
2	$e^{-m} . \frac{m^2}{2!} = \frac{m^2}{2} . e^{-m}$
3	$e^{-m} . \frac{m^3}{3!} = \frac{m^3}{3!} . e^{-m}$
x	$e^{-m} . \frac{-m^x}{x!}$

Properties/Characteristics of Poisson Distribution

The following are the important properties or characteristics of Poisson distribution:

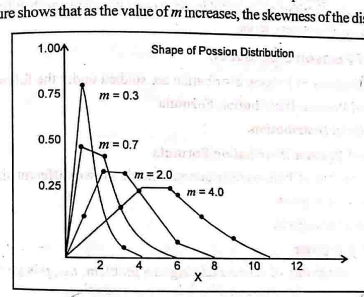
(1) Discrete Probability Distribution: The Poisson distribution is a discrete probability distribution in which the number of successes are given in whole numbers such as 0, 1, 2..., etc.

(2) Value of p and q: The Poisson distribution is used in those situations where the probability of occurrence of an event is very small (*i.e.*, $p \to 0$) and the probability of the non-occurrence of the event is very large (*i.e.*, $q \to 1$) and the value of n is also infinitely large.

(3) Main Parameter: It has only one parameter m and its value is equal to np, *i.e.*, $m = np$. The entire distribution can be known from this parameter.

(4) Shape of Poisson Distribution: The Poisson distribution is always positively skewed but the skewness diminishes as the value of m increases. The distribution shifts to the right and degree of skewness falls as m increases. The following figure illustrate the point:

The above figure shows that as the value of m increases, the skewness of the distribution diminishes.



(5) Constant of Poisson Distribution: The constants of the Poisson Distribution can be obtained from the following formula:

Mean $= \bar{X} = m = np$

Moment coeff. of skewness $= \sqrt{\beta_1} = \frac{1}{\sqrt{m}}$

Variance $= \sigma^2 = m$

S.D. $= \sigma = \sqrt{m}$

Moment coeff. of Kurtosis $= \beta_2 = 3 + \frac{1}{m}$

(6) Equality of Mean and Variance: An important characteristic of the Poisson distribution is that its mean and variance are equal, *i.e.*,

$\bar{X} = \sigma^2$ or Mean = Variance.

(For Proof, See Example 53)

Role/Uses/Examples/Importance of Poisson Distribution

Poisson distribution is widely used in the study of many problems. Few practical situations in which the Poisson distribution can be used are given below:

(1) It is used in statistical quality control to count the number of defects of an item.
(2) In Biology to count the number of bacteria.
(3) In Insurance problems to count the number of casualities.
(4) To count the number of typing errors per page in a typed material.
(5) To count the number of incoming telephone calls in a town.
(6) To count the number of defective blades in a lot of manufactured blades in a factory.
(7) To count the number of deaths at a particular crossing in a town as a result of road accident.
(8) To count the number of suicides committed by lovers point in a year.

In general, the Poisson distribution is useful in rare events where the probability of success (p) is very small and the value of n is very large.

Applications of Poisson Distribution

The practical applications of Poisson distribution are studied under the following heads:

(A) Application of Poisson Distribution Formula

(B) Fitting of Poisson Distribution.

(A) Applications of Poisson Distribution Formula

We study the applications of Poisson distribution formula in two different situations:

(1) When the value of p is given
(2) When the value of m is given

(1) When the value of p is given

When we are given probability of success relating to a problem, *i.e.*, p is given, the uses of the Poisson distribution formula can be illustrated by following examples:

Example 28. It is given that 2% of the screws manufactured by a company are defective. Use Poisson Distribution to find the probability that a packet of 100 screws contains: (*i*) no defective screws (*ii*) one defective and (*iii*) two or more defectives. [Given, $e^{-2} = 0.135$]

Solution: Let p = probability of a defective screw = 2% = $\frac{2}{100}$

In the usual notation, we are given

$$p = \frac{2}{100}, \quad n = 100$$

$$\therefore \quad m = np = 100 \times \frac{2}{100} = 2$$

The Poisson Distribution is given as:

$$P(X = x) = P(X = 0) = \frac{e^{-2} . 2^0}{0!}$$

$$= e^{-2} = 0.135 \qquad \text{[given: } e^{-2} = 0.135]$$

(*ii*) P (One defective) = $P(X = 1) = \frac{e^{-2} . 2^1}{1!}$

$$= e^{-2} \times 2 = (0.135)(2) = 0.270$$

(3) P (Two or more defectives) = 1 − [P(0) + P(1)]

$$= 1 - [0.135 + 0.270] = 1 - 0.405 = 0.595$$

Example 29. A manufacturer of pins knows that on an average 5% of his product is defective. He sells pins in a packet of 100 and guarantees that not more than 4 pins will be defective. What is the probability that a packet will meet the guaranteed quality? [Given: $e^{-5} = 0.0067$]

Solution: Let p = probability of a defective pin = 5% = $\frac{5}{100}$

We are given: $n = 100, p = \frac{5}{100}$

$$\therefore \quad m = np = 100 \times \frac{5}{100} = 5$$

The Poisson distribution is given as:

$$P(X = x) = \frac{e^{-m} . m^x}{x!}$$

Required probability = P [packet will meet the guarantee]

= P [packet contains upto 4 defectives]

= P(0) + P(1) + P(2) + P(3) + P(4)

$$= e^{-5}\frac{5^0}{0!} + e^{-5}\frac{5^1}{1!} + e^{-5}\frac{5^2}{2!} + e^{-5}\frac{5^3}{3!} + e^{-5}\frac{5^4}{4!}$$

$$= e^{-5}\left[1 + \frac{5}{1} + \frac{25}{2} + \frac{125}{6} + \frac{625}{24}\right]$$

$$= 0.0067 \times 65.374 = 0.438$$

Example 30. Assume that the probability of a fatal accident in a factory during the year is $\frac{1}{1200}$. Calculate the probability that in a factory employing 300 workers, there will be at least 2 fatal accidents in a year (Given $e^{-0.25} = 0.7788$).

Solution: Let p = probability of a fatal accident = $\frac{1}{1200}$

n = number of workers = 300

$\therefore \quad m = np = 300 \times \frac{1}{1200} = \frac{1}{4} = 0.25$

The Poisson distribution is given as:

$$P(X = x) = \frac{e^{-m} \cdot m^x}{x!}$$

$$P\text{ (0 accident)} = P(X=0) = \frac{e^{-m} \cdot m^0}{0!} = e^{-m} = e^{-0.25} = 0.7788$$

$$P\text{ (1 accident)} = P(X=1) = \frac{e^{-m} \cdot m^1}{1!} = \frac{(0.25)}{1!}(e^{-0.25})$$

$$= (0.25)(0.7788) = 0.1947$$

$$P\text{ (at least 2 fatal accidents)} = 1 - [P(0) + P(1)]$$

$$= 1 - [0.7788 + 0.1947] = 1 - 0.9735 = 0.0265$$

Example 31. It is known that from the past experience that in a certain factory 3% products are defective. A sample of 100 items are taken at random. Find the probability that exactly 5 products are defective (Given: $e^{-3} = 0.04979$).

Solution: Let p = probability of a defective product = $3\% = \frac{3}{100}$

We are given $n = 100, p = \frac{3}{100}$

$\therefore \quad m = np = 100 \times \frac{3}{100} = 3$

The Poisson Distribution is given as:

$$P(X = x) = \frac{e^{-m} \cdot m^x}{x!}$$

$$P\text{ (Exactly 5)} = P(X = 5) = \frac{e^{-3} \cdot 3^5}{5!}$$

$\because \quad e^{-3} = 0.04979$ (given)

$\therefore \quad P(5) = \frac{0.04979 \times (3)^5}{5 \times 4 \times 3 \times 2 \times 1} = 0.100$

Note: When the value of e^{-m} is not given in the question, we compute the value of e^{-m} by using the formula:

$$e^{-m} = \text{Reciprocal [Antilog } (m \times .4343)]$$

Suppose, $m = 3$

$\therefore \quad e^{-3}$ = Rec. [Antilog (3 × 0.4343)]

= Rec. [Antilog (1.3029)]

= Rec. [20.08] = 0.04979

Example 32. The probability of a defective bolt is 3%. Find (*i*) mean (*ii*) standard deviation, (*iii*) moment coefficient of skewness and (*iv*) moment coefficient of kurtosis in a total of 50 packets.

Solution: We have, $p = 3\% = \frac{3}{100} = 0.03, n = 50$

$$m = np = \frac{3}{100} \times 50 = 1.5$$

(*i*) Mean = m = 1.5

(*ii*) S.D. = $\sqrt{m} = \sqrt{1.5} = 1.225$

(*iii*) Moment coeff. of Skewness $(\sqrt{\beta_1}) = \frac{1}{\sqrt{m}} = \frac{1}{\sqrt{1.5}} = \frac{1}{1.225} = 0.82$

(*iv*) Moment coeff. of Kurtosis $(\beta_2) = 3 + \frac{1}{m} = 3 + \frac{1}{1.5} = 3.67$

Example 33. If X is a Poisson variate such that $P(X = 1) = P(X = 2)$. Find the mean and variance of the distribution.

Solution: Poisson distribution is given by

$$P(X = x) = \frac{e^{-m} \cdot m^x}{x!}$$

Putting X = 1, 2

$\therefore \quad P(X = 1) = \frac{e^{-m} \cdot m^1}{1!} = m e^{-m}$

$$P(X = 2) = \frac{e^{-m} \cdot m^2}{2!} = \frac{m^2 \cdot e^{-m}}{2!}$$

By the given condition

$$P(X = 1) = P(X = 2)$$

$\therefore \quad e^{-m} \cdot m = e^{-m} \cdot \frac{m^2}{2 \times 1}$

Cancelling me^{-m} both sides, we get

$\therefore \quad 1 = \frac{m}{2}$

$\therefore \quad m = 2$

Hence, mean of the distribution = 2

As $\quad \overline{X} = \sigma^2$,

$\therefore \quad$ Variance $\sigma^2 = 2$

EXERCISE 8.4

1. It is given that 3% of the electric bulbs manufactured by a company are defective. Using the Poisson distribution, find the probability that a sample of 100 bulbs will contain (*i*) no defective (*ii*) exactly one defective. (Given: $e^{-3} = 0.04979$) **[Ans.** (*i*) 0.05 (*ii*) 0.15]
2. Find the probability that at most 5 defective bolts will be found in a box of 200 bulbs if it is known that 2 per cent of such bolts are expected to be defective (You may take the distribution to be Poisson). (Take $e^{-4} = 0.0183$) **[Ans.** 0.784]
3. Assuming one in 80 births is a case of twins, calculate the probability of 2 or more sets of twins on a day when 30 births occur.
   [Hint: $m = [0.0125 \times 30 = 0.3750]$ **[Ans.** 0.055]
4. A manufacturer of pins known that on an average 2% of his product is defective. He sells pins in a boxes of 200 and guarantees that not more than 3 pins will be defective. What is the probability that a box will fail to meet the guaranteed quality? (Given: $e^{-4} = 0.0183$) **[Ans.** 0.567]
5. One fifth per cent of the blades produced by a blade manufacturing factors turn out to be defective. The blades are supplied in packets of 10. Use Poisson distribution to calculate the approximate number to packets containing no defective, one defective and two defective blades respectively in a consignment of 1,00,000 packets. (Given: $e^{-0.02} = 0.9802$) **[Ans.** 98020, 1960.4, 19.604]
6. In a factory manufacturing pens, of which 0.5 percent are defective pens each carton is packet of 100 pens. Which is the percentage of such cartons of which (*i*) not a single pen is defective, (*ii*) at least each one pen is defective and (*iii*) two or more pens are defective? (Given: $e^{-0.5} = 0.6065$) **[Ans.** (*i*) 60.65% (*ii*) 39.35% (*iii*) 9.025%]
7. The probability of a defective bolt is 2%. Find (*i*) mean (*ii*) standard deviation (*iii*) moment coefficient of skewness and (*iv*) moment coefficient of kurtosis in a total of 200 bolts. **[Ans.** $\bar{X} = 4,\ \sigma = 2, \sqrt{\beta_1} = 0.5, \beta_2 = 3.25$]
8. In a frequency distribution, frequency corresponding to 3 successes is 2/3 times frequency corresponding to 4 successes. Find the $\bar{X}$ and σ. **[Ans.** $\bar{X} = 6, \sigma = 2.44$]
9. If X is a poisson variable such that
   $P(X = 2) = 9\,P(X = 4) + 90P(X = 6)$, find the mean and variance of X.
   [Hint: $m^4 + 3m^2 - 4 = 0 \Rightarrow m^2 = 1 \Rightarrow m = +1$] **[Ans.** $\bar{X} = 1, V(X) = 1$]

(2) When the value of *m* is given

When we are given the value of the parameter *m* relating to a problem, the uses of Poisson distribution formula can be illustrated by the following examples:

Example 34. Between the hours of 2 P.M. and 4 P.M., the average number of phone calls per minute coming into a switch board of a company is 2.5. Find the probability that during one particular minute there will be (*i*) no phone call at all (*ii*) exactly 3 calls, (*iii*) at least 2 calls. (Given $e^{-2} = 0.1353$, $e^{-0.5} = 0.6065$).

Solution: This is a problem of Poisson distribution

$$P(X) = \frac{e^{-m}\,.\,m^x}{x!} \quad \text{where, X} = 1, 2, 3......$$

Given, Average number of phone calls = $\bar{X} = m = 2.5$

Poisson Distribution is given by

$$P(X = x) = \frac{e^{-m}\,.\,m^x}{x!}$$

$$(i)\ \text{P (No call)} = \text{P (X = 0)} = \frac{e^{-2.5}\,.(2.5)^0}{0!} = e^{-2.5}$$

$$= e^{-2}\,.\,e^{-0.5} \qquad (\text{Given } e^{-2} = 0.1353,\ e^{-0.5} = 0.6065)$$

$$= 0.1353 \times 0.6065 = 0.0821$$

Hence the probability that during one particular minute there will be no phone call at all is 0.0821.

$$(ii)\ \text{P (Exactly 3 calls)} = \text{P(X = 3)} = \frac{e^{-2.5}\,.(2.5)^3}{3!}$$

$$= \frac{(0.0821)(15.625)}{3 \times 2 \times 1} = 0.2138$$

$$(iii)\ \text{P (At least 2 calls)} = 1 - [\text{P(X = 0)} + \text{P(X = 1)}]$$

$$= 1 - [e^{-2.5} + (2.5)\,e^{-2.5}]$$

$$= 1 - e^{-2.5}\,[1 + 2.5] = 1 - [(0.0821)\,(3.5)]$$

$$= 1 - 0.28735 = 0.71265$$

Example 35. It is known from the past experience that the average number of industrial accidents in a factory per month in a plant is 4. Find the probability that during a particular month, there will be lower than 4 accidents. Use Poisson distribution to explain your answer (Given $e^{-4} = 0.0183$).

Solution: Given, Average No. of Accidents = $\bar{X} = m = 4$.

$$\text{P(X)} = e^{-m}\,.\,\frac{m^x}{x!}$$

$$\text{P(0)} = e^{-m}\,.\,\frac{m^0}{0!} = e^{-m} = e^{-4}$$

$$\text{P(1)} = e^{-m}\,.\,\frac{m^1}{1!} = me^{-m} = (4)\,e^{-4}$$

$$\text{P(2)} = e^{-m}\,.\,\frac{m^2}{2!} = \frac{m^2}{2!}\,.\,e^{-m} = \frac{(4)^2}{2!}\,e^{-4}$$

$$\text{P(3)} = e^{-m}\,.\,\frac{m^3}{3!} = \frac{m^3}{3!}\,.\,e^{-m} = \frac{(4)^3}{3!} \times e^{-4}$$

The probability that there will be lower than 4 accidents

$= P(0) + P(1) + P(2) + P(3)$

$= e^{-4} \cdot \left[1 + 4 + \frac{4^2}{2!} + \frac{4^3}{3!}\right]$

$= 0.0183\ [1 + 4 + 8 + 10.67]$ $\quad [\because e^{-4} = 0.0183]$

$= 0.0183 \times 23.67 = 0.4332$

$\therefore$ Probability of less than 4 accidents is 0.4332 or 43.32%.

Example 36. A car hire firm has two cars which it hires out daily. The number of demand for a car on each day is distributed as Poisson distribution with mean 1.5. Calculate the number of days out of 100 days on which (*i*) neither car is used and (*ii*) some demand is refused. (Given $e^{-1.5} = 0.2231$).

Solution: Given, Mean = $m = 1.5$, N = 100

$$P(X = x) = \frac{e^{-m} . m^x}{x!}$$

(*i*) P (Neither car is used) = P (demand for car is zero)

$= P(X = 0) = \frac{e^{-1.5} . (1.5)^0}{0!}$

$= e^{-1.5} = 0.2231$ $\quad$ (Given $e^{-1.5} = 0.2231$)

Thus, the required number of days on which neither car is used

$= 100 \times 0.2231 = 22.31 \approx 22$

(*ii*) P (Demand for car is refused) = P (Demand for car is more than 2)

$= 1 - P\ (0, 1 \text{ or } 2)$

$= 1 - [P(0) + P(1) + P(2)]$

$= 1 - \left[e^{-1.5} + (1.5)\, e^{-1.5} + \frac{(1.5)^2}{2!} . e^{-1.5}\right]$

$= 1 - e^{-1.5}\left[1 + 1.5 + \frac{2.25}{2}\right]$

$= 1 - (0.2231)\left(\frac{7.25}{2}\right)$ $\quad [\because e^{-1.5} = 0.2231]$

$= 1 - 0.8087 = 0.1913$

Thus, the number of days on which some demand is refused

$= 100 \times 0.1913 = 19.13 \approx 19$

Example 37. Consider a Poisson probability distribution with 2 as the average number of occurrences per time period:

(*i*) Write the appropriate Poisson probability function.

(*ii*) What is the average number of occurrence in 3 time periods?

(*iii*) Find the probability of 6 occurrences in 3 time periods.

Solution: Given, average no.of occurrences for 1 time period = $m = 2$

(*i*) Poisson probability function = $P(X = x) = \frac{e^{-2} 2^x}{x!}$

(*ii*) Average no.of occurrences for 3 time period = $2 \times 3 = 6$

(*iii*) $P[X = 6] = \frac{e^{-6} . (6)^6}{6!} = 0.1575$

IMPORTANT TYPICAL EXAMPLE

Example 38. In a town 10 accidents took place in a period of 50 days. Assume that the number of accidents per day follows Poisson Distribution, find the probability that there will be three or more accidents per day. (Given: $e^{-0.2} = 0.8187$)

Solution: The average number of accidents per day = $m = \frac{10}{50} = 0.2$

P (3 or more accidents) = 1 – P [2 or less accidents]

$= 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$

$= 1 - \left[e^{-0.2} + 0.2\, e^{-0.2} + \frac{(0.2)^2}{2!} e^{-0.2}\right]$

$= 1 - e^{-0.2} . [1 + 0.2 + 0.02]$

$= 1 - 0.8187\ (1.22) = 1 - 0.9988$

$= 0.0012$

EXERCISE 8.5

1. Suppose that a manufacturing product has 4 defects per unit of product inspected. Using Poisson distribution calculate the probability of finding a product with 2 defects. (Given: $e^{-4} = 0.01832$) **[Ans. 0.146624]**
2. The number of accidents is a year attributes to taxi drivers in a city follows Poisson distribution with mean 3. Out of 1,000 taxi drivers, find approximately the number of drivers with (*i*) no accidents in a year, and (*ii*) more than 3 accidents in a year. (Given: $e^{-1} = 0.3679, e^{-2} = 0.1353, e^{-3} = 0.0498$). **[Ans. (*i*) 50 (*ii*) 353]**
3. Given that the switchboard of consultant's office receives on the average 0.9 calls per minute. Find the probabilities
   (*i*) In a given minute there will be at least one incoming call.
   (*ii*) Between 10.00 AM and 10.02 AM there will be exactly 2 incoming calls.
   (*iii*) During an interval of four minutes there will be at most 3 incoming calls.
   (Given: $e^{-0.9} = 0.4066, e^{-1} = 0.36788, e^{-0.6} = 0.5488, e^{-0.8} = 0.4493, e^{-3} = 0.04979$)
   [**Hint:** See Example 51] **[Ans. (*i*) 0.5934, (*ii*) 0.2677, (*iii*) 0.5147]**

4. A television company estimates that average demand for engineers for repairing TV sets on each day is 1.5. Assuming this as a Poisson distribution it appoints two engineers. Calculate the proportion of days in a year in which both engineers are unemployed and the proportion of days in which the some demand for engineers is refused.
(Given: e^{-1}=0.3678, $e^{-0.5}$=0.6065). [Ans. (*i*) 0.2231, (*ii*) 0.1913]
5. A telephone exchange receives on the average 4 calls per minute. Find the probability on the basis of Poisson distribution if (*i*) 2 or less calls per minute (*ii*) upto 4 calls per minute and (*iii*) more than 4 calls per minute. (Given: e^{-4}=0.0183) [Ans. (*i*) 0.2379, (*ii*) 0.6283, (*iii*) 0.3717]

▶ (B) Fitting of Poisson Distribution

The following procedure is adopted for fitting a Poisson distribution to the observed data:

(1) Firstly, we compute mean ($\overline{X}$) from the observed frequency data by using the formule:

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

We use the value of this mean as the parameter of the Poisson distribution, *i.e.*, $\overline{X} = m$.

(2) The value of e^{-m} is obtained. If the value of e^{-m} is not given in the question, then the following formule is used to compute:

$$e^{-m} = \text{Reciprocal [Antilog } (m \times 0.4343)]$$

(3) Then, we compute the probability of 0, 1, 2, 3 or x success by using the Poisson distribution formule:

$$P(X = x) = \frac{e^{-m} . m^x}{x!}$$

(4) The expected or theoretical frequencies are then obtained by multiply each probability with N, total frequencies. Thus,

No. of Successes (X)	Probability P(X)	Expected Frequencies fe(x)
0	$P(0) = e^{-m} . \frac{m^0}{0!} = e^{-m}$	$N.P(0) = Ne^{-m}$
1	$P(1) = e^{-m} . \frac{m^1}{1!} = e^{-m} . m$	$N.P(1) = N.e^{-m} . m$
2	$P(2) = e^{-m} . \frac{m^2}{2!} = e^{-m} . \frac{m^2}{2!}$	$N.P(2) = N.e^{-m} . \frac{m^2}{2!}$
⋮	⋮	⋮
x	$P(x) = \frac{e^{-m} . m^x}{x!}$	$N.P(x) = \frac{N.e^{-m} . m^x}{x!}$

Alternative Method:

The expected frequencies can also be calculated in an easy way as follows:

(*i*) First we calculate the $fe(0) = N . P(0) = N . e^{-m}$

(*ii*) Other expected frequencies can be calculated as follows:

$$fe(0) = N . P(0) = N . e^{-m}$$

$$fe(1) = \frac{m}{1} . fe(0)$$

$$fe(2) = \frac{m}{2} . fe(1)$$

$$fe(3) = \frac{m}{3} . fe(2)$$

$$fe(4) = \frac{m}{4} . fe(3) \text{ and so on.}$$

The following examples would clarify the procedure of fitting a Poisson distribution:

Example 39. Fit a Poisson distribution to the following data and calculate the theoretical frequencies:

Deaths:	0	1	2	3	4
Frequency:	109	65	22	3	1

Also find mean and variance of the above distribution. (Given $e^{-0.66}$ = 05432)

Solution: **Fitting of Poisson Distribution**

Deaths (X)	Frequency (f)	fX
0	109	0
1	65	65
2	22	44
3	3	9
4	1	4
	$\Sigma f = 200$	$\Sigma fX = 122$

$$\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{122}{200} = 0.61$$

$\therefore$ $m = 0.61$

Now we obtain the value of $e^{-0.61}$ either from the table or by using the formula

$$e^{-m} = \text{Rec. [Antilog } (m \times 0.4343)]$$

Putting $m = 0.61$

$$e^{-0.61} = \text{Rec. [Antilog } (0.61 \times 0.4343)]$$
$$= \text{Rec. [Antilog } (0.26492)]$$
$$= \text{Rec. } [1.841] = 0.5432$$

Now, $$P(0) = e^{-0.61} . \frac{(0.61)^0}{0!}$$
$$= e^{-0.61} = 0.5432$$

Calculation of Expected Frequencies

$fe(0) = N.P(0) = 200 \times (0.5432) = 108.64 \approx 109$

$fe(1) = fe(0) \times \frac{m}{1} = 108.64 \times \frac{0.61}{1} = 66.27 \approx 66$

$fe(2) = fe(1) \times \frac{m}{2} = 66.27 \times \frac{0.61}{2} = 20.21 \approx 20$

$fe(3) = fe(2) \times \frac{m}{3} = 20.21 \times \frac{0.61}{3} = 4.11 \approx 4$

$fe(4) = fe(3) \times \frac{m}{4} = 4.11 \times \frac{0.61}{4} = 0.63 \approx 1$

Thus,

X:	0	1	2	3	4
fe:	109	66	20	4	1

Mean $= \bar{X} =$ Variance $= \sigma^2 = 0.61$

Example 40. After correcting the proofs of the first 50 pages of a book, it is found that an average there are 3 errors per 5 pages. Use Poisson distribution to estimate the number of pages with 0, 1, 2, 3, ... errors in the whole book of 1000 pages.

(You are given $e^{-0.6} = 0.5488$)

Solution: The average number of mistake $= m = \frac{3}{5} = 0.6$

Calculation of Expected Number of Pages

X	P(X)	$fe(X) = N \cdot P(X)$
0	0.5488	$1000 \times 0.5488 = 548.8 = 549$
1	0.32928	$1000 \times 0.32928 = 329.28 = 329$
2	0.098784	$1000 \times 0.098784 = 98.74 = 98$
3	0.0197568	$1000 \times 0.0197568 = 19.7568 = 20$
more than 3	0.0033792	$1000 \times 0.0033792 = 3.37 = 3$
		N = 1000

where, $P(0) = e^{-m}\frac{m^0}{0!} = e^{-0.6} \times \frac{0.6^0}{0!} = e^{-0.6} = 0.5488$ (Given)

$P(0) = \frac{e^{-0.6} \cdot (0.6)^1}{1!} = \frac{0.5488 \times 0.6}{1} = 0.32928$

$P(2) = \frac{e^{-.6} \cdot (0.6)^2}{2!} = \frac{0.5488 \times 0.36}{2 \times 1} = 0.098784$

$P(3) = \frac{e^{-0.6} \cdot (0.6)^3}{3!} = \frac{0.5488 \times 0.216}{3 \times 2 \times 1} = 0.0197568$

$P[X > 3] = 1 - [P(0) + P(1) + P(2) + P(3)]$

$\therefore$ $P[X > 3] = 1 - [0.5488 + 0.32928 + 0.098784 + 0.0197568]$

$= 1 - [0.9966208]$

$= 0.0033792$

Example 41. The following table gives the number of days in a 50 days period during which automobile accidents occured in a certain part of a city. Fit a Poisson distribution to the data:

No. of accidents:	0	1	2	3	4
No. of days:	19	18	8	4	1

Solution:

Fitting of Poisson Distribution

No. of accidents (X)	No. of days (f)	fX
0	19	0
1	18	18
2	8	16
3	4	12
4	1	4
	$N = \Sigma f = 50$	$\Sigma fX = 50$

$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{50}{50} = 1$

$\therefore$ $m = 1$

Since the value of e^{-m} is not given, we obtain the value of e^{-1} by using the formula:

$e^{-m} =$ Rec. [Antilog $(m \times 0.4343)$]

Here $m = 1$

$\therefore$ $e^{-m} =$ Rec. [Antilog (1×0.4343)]

$=$ Rec. [Antilog (0.4343)]

$=$ Rec. [2.719]

$= \frac{1}{2.719} = 0.3678$

Now, $P(0) = e^{-1}\frac{(1)^0}{0!}$

$= e^{-1} = 0.3678$

Calculation of Expected Frequencies

$$fe(0) = N.P(0) = 50 \times 0.3678 = 18.39$$

$$fe(1) = fe(0) \times \frac{m}{1} = 18.39 \times \frac{1}{1} = 18.39$$

$$fe(2) = fe(1) \times \frac{m}{2} = 18.39 \times \frac{1}{2} = 9.195$$

$$fe(3) = fe(2) \times \frac{m}{3} = 9.195 \times \frac{1}{3} = 3.065$$

$$fe(4) = fe(3) \times \frac{m}{4} = 3.065 \times \frac{1}{4} = 0.76625$$

Thus the expected frequencies of 0, 1, 2, 3, 4 accidents as per Poisson distribution are:

No. of accidents:	0	1	2	3	4
No. of deaths:	18.39	18.39	9.195	3.065	0.76625

EXERCISE 8.6

1. The following mistakes per page were observed in a book:

No. of mistakes per page:	0	1	2	3	4
No. of pages:	211	90	19	5	0

Fit a Poisson distribution for the data. **[Ans.** 209.40, 92.14, 20.27, 2.97, 0.33]

2. One hundred car radios are inspected as they come off the production line and number of defects per radio set is recorded below:

No. of defects:	0	1	2	3
No. of Radio sets	79	18	2	1

Estimate the average number of defects per radio and expected frequencies of 0, 1, 2, 3. **[Ans.** $m = 0.25$, 77.88, 19.47, 2.43, 0.21]

3. Fit a Poisson distribution of the following data and calculate the theoretical frequencies:

Death:	0	1	2	3	4
Frequency:	122	60	15	2	1

(Given: $e^{-0.5} = 0.60657$) **[Ans.** 121.3, 60.65, 15.16, 2.53, 0.32]

4. Below are given the number of vacancies of judges occurring in a High Court over a period of 96 years:

No. of vacancies:	0	1	2	3
Frequency:	59	27	9	1

Fit a Poisson Distribution and calculate the mean and variance of the above distribution. **[Ans.** 58.22, 29.11, 7.278, 1.21, $\overline{X} = \sigma^2 = 0.5$]

MISCELLANEOUS SOLVED EXAMPLES

Example 42. The probability that India wins a Cricket Test Match against Pakistan is given to be 1/4. If India and England play three test matches, find the probability that:

(*i*) India will lose all the 3 matches.

(*ii*) India wins atleast 1 test match.

(*iii*) India wins two matches.

Solution:

$$p = \text{probability of winning the test match} = \frac{1}{4}$$

$$\therefore \quad q = 1 - p = 1 - \frac{1}{4} = \frac{3}{4} \text{ and } n = 3$$

(*i*) The probability of losing all matches is given by:

$$P(X=0) = {}^3C_0 \left(\frac{3}{4}\right)^3 . \left(\frac{1}{4}\right)^0 = 1 \times \frac{27}{64} = \frac{27}{64}$$

(*ii*) The probability of winning at least 1 test match is given by:

$$P(X \geq 1) = 1 - P(X=0) = 1 - \frac{27}{64} = \frac{64-27}{64} = \frac{37}{64}$$

(*iii*) The probability of winning two matches is given by:

$$P(X=2) = {}^3C_2 \left(\frac{3}{4}\right)^1 . \left(\frac{1}{4}\right)^2 = 3 \times \frac{3}{64} = \frac{9}{64}$$

Example 43. Assuming that sex ratio of male children is 1/2. Find the probability that in a family of 5 children, (*i*) all children will be of same sex, (*ii*) three of them will be boys and two girls.

Solution: Let p = probability of a male child = 1/2

$$\therefore \quad q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}, \; n = 5$$

(*i*) P (Same Sex) = P (either 5B or 5G) = P(5B) + P(5G) = $P(X = 5) + P(X = 0)$

$$= {}^5C_5 \left(\frac{1}{2}\right)^5 . \left(\frac{1}{2}\right)^0 + {}^5C_0 \left(\frac{1}{2}\right)^0 . \left(\frac{1}{2}\right)^5$$

$$= 1.\frac{1}{32} + 1.\frac{1}{32} = \frac{2}{32} = \frac{1}{16}$$

(*ii*)

$$P(3B) = {}^5C_3 \left(\frac{1}{2}\right)^3 . \left(\frac{1}{2}\right)^2$$

$$= 10 \times \frac{1}{32} = \frac{5}{16}$$

Example 44. If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 bolts (*i*) no bolt (*ii*) one bolt and (*iii*) at most 2 bolts will be defective.

Solution: Let p = probability of a defective bolt $= 20\% = \frac{20}{100} = \frac{1}{5}$

$$\therefore\ q = 1 - \frac{1}{5} = \frac{4}{5},\ n = 4$$

(*i*) $P(X=0) = {}^4C_0 \cdot \left(\frac{4}{5}\right)^4 \cdot \left(\frac{1}{5}\right)^0 = \frac{256}{625}$

(*ii*) $P(X=1) = {}^4C_1 \cdot \left(\frac{4}{5}\right)^3 \cdot \left(\frac{1}{5}\right)^1 = 4 \times \frac{64}{625} = \frac{256}{625}$

(*iii*) $P(X=0,1,2) = {}^4C_0\left(\frac{4}{5}\right)^4 \cdot \left(\frac{1}{5}\right)^0 + {}^4C_1\left(\frac{4}{5}\right)^3 \cdot \left(\frac{1}{5}\right)^1 + {}^4C_2\left(\frac{4}{5}\right)^2 \left(\frac{1}{5}\right)^2$

$$= \frac{256}{625} + \frac{256}{625} + \frac{96}{625} = \frac{608}{625}$$

Example 45. There are 64 beds in a garden and 3 seeds of a particular type of flower are sown in each bed. The probability of a flower being white is 1/4. Find the number of beds with 3, 2, 1 and 0 white flowers.

Solution: Let p = probability of a white flower $= \frac{1}{4}$

$$q = 1 - p = 1 - \frac{1}{4} = \frac{3}{4};\ n = 3$$

White flower ($X=x$)	P(X)	Total No. of beds N.P(X)
3	${}^3C_3\left(\frac{3}{4}\right)^0 \cdot \left(\frac{1}{4}\right)^3 = \frac{1}{64}$	$64 \times \frac{1}{64} = 1$
2	${}^3C_2\left(\frac{3}{4}\right)^1 \cdot \left(\frac{1}{4}\right)^2 = \frac{9}{64}$	$64 \times \frac{9}{64} = 9$
1	${}^3C_1\left(\frac{3}{4}\right)^2 \cdot \left(\frac{1}{4}\right)^1 = \frac{27}{64}$	$64 \times \frac{27}{64} = 27$
0	${}^3C_0\left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^0 = \frac{27}{64}$	$64 \times \frac{27}{64} = 27$

Example 46. The overall result of a college in arts faculty is 60%. Find the probability that out of a group of 5 candidates, atleast 4 passed the examination.

Solution: Let p = probability of passing $= 60\% = \frac{60}{100} = \frac{3}{5}$

$$\therefore\ q = 1 - \frac{3}{5} = \frac{2}{5},\quad n = 5$$

P (at least 4 passes) = P(4) + P(5)

$$= {}^5C_4\left(\frac{2}{5}\right)^1 \cdot \left(\frac{3}{5}\right)^4 + {}^5C_5\left(\frac{2}{5}\right)^0 \cdot \left(\frac{3}{5}\right)^5$$

$$= \frac{5 \times 2 \times 81}{3125} + 1 \times \frac{243}{3125} = \frac{810}{3125} + \frac{243}{3125} = \frac{1053}{3125}$$

Example 47. A multiple choice test consists of 8 questions with 3 answers (of which only one is correct). A student answers each question by rolling a balanced die and checking the first answer if he gets 1 or 2, the second answer if he gets 3 or 4 and the third answer if he gets 5 or 6. To get a distinction, the student must secure at least 75% correct answers. If there is no negative marking, what is the probability that the student secures a distinction?

Solution: Probability of getting correct answer $= p = \frac{1}{3}$

Probability of getting incorrect answer $= q = 1 - \frac{1}{3} = \frac{2}{3}$ and $n = 8$

The minimum number of questions a student is supposed to answer correctly to get at least 75% marks.

$$= 8 \times 75\% = 8 \times \frac{3}{4} = 6 \text{ questions}$$

Probability of getting at least 75% marks:

$$P(X=6) + P(X=7) + P(X=8) = {}^8C_6\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^2 + {}^8C_7\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right)^1 + {}^8C_8\left(\frac{1}{3}\right)^8$$

$$= \frac{28 \times 4}{6561} + \frac{8 \times 2}{6561} + \frac{1}{6561}$$

$$= \frac{112 + 16 + 1}{6561} = \frac{129}{6561}$$

$$= 0.0197$$

Example 48. Suppose the probability that an item produced by a particular machine is defective is 0.2. If 10 items produced by this machine are selected at random, what is the probability that not more than one defective item is found? Use the binomial and Poisson distributions and compare the answers. (Use $e^{-2} = 0.1353$)

Solution: **Using B.D.**

Here $n = 10,\ p = 0.2,\ q = 1 - 0.2 = 0.8$

$$P(X = x) = {}^nC_x\, q^{n-x} \cdot p^x$$

$\therefore$ P (not more than one defective) $= P[X = x = 0] + P[X = x = 1]$

$$= {}^{10}C_0\,(0.8)^{10} \cdot (0.2)^0 + {}^{10}C_1\,(0.8)^9 \cdot (0.2)^1$$

$$= 0.1074 + 0.2684 = 0.3758$$

Using P.D.

Here, $n = 10, p = 0.2, \therefore\ m = np = 10 \times 0.2 = 2$

$$P(X = x) = \frac{e^{-m} . m^x}{x!},\ x = 0, 1, 2$$

$\therefore$ P (not more than one defective) = P (X = 0) + P (X = 1)

$$= e^{-2} . \frac{(2)^0}{0!} + e^{-2} . \frac{(2)^1}{1!} = e^{-2}(1+2)$$

$$= 0.1353 \times 3 = 0.4059$$

$\therefore$ We find that the required probability is more when PD is used.

Example 49. If the probability that an individual suffers from reaction of a given medicine is 0.001, determine the probability that out of 2,000 individuals (*i*) exactly 3 individuals and (*ii*) more than 2 individuals will suffer from reaction. (Given: $e^{-2} = 0.1353$)

Solution: Probability of suffering from reaction = 0.001

$n = 2{,}000$

$$\therefore\ m = np = 0.001 \times 2000 = \frac{1}{1000} \times 2000 = 2$$

The table value of $e^{-2} = 0.1353$

(*i*) $P(\text{Exactly } 3) = P(X = 3) = \frac{e^{-m}\ m^3}{3!} = e^{-2} . \frac{2^3}{3!}$

$$= \frac{8}{6} \times 0.1353 = 0.1804$$

(*ii*) P (More than 2) = 1 – [P (0) + P (1) + P (2)]

$$= 1 - \left[e^{-2} . \frac{2^0}{0!} + e^{-2} . \frac{2^1}{1!} + e^{-2} . \frac{2^2}{2!} \right]$$

$$= 1 - e^{-2} . [1+2+2] = 1 - e^{-2} . [5]$$

$$= 1 - (0.1353)(5) = 1 - 0.6765 = 0.3235$$

Example 50. Comment on the following:

For a poisson distribution, Mean = 8 and Variance = 7.

Solution: In a poisson distribution, Mean = variance, *i.e.*, $\overline{X} = \sigma^2$

We are given: $\overline{X} = 8, \sigma^2 = 7$

Since $\overline{X} > \sigma^2$, the statement is incorrect.

Example 51. Given that the switch board of consultant's office receives on the average 0.9 calls per minute. Find the probabilities: (*i*) In a given minute there will be at least one incoming call, (*ii*) Between 10.00 AM and 10.02 AM there will be exactly 2 incoming calls, (*iii*) During an interval of 4 minutes there will be at most 3 incoming calls. (Given: $e^{-0.9} = 0.4066, e^{-1} = 0.36788, e^{-0.6} = 0.5488, e^{-0.8} = 0.4493, e^{-3} = 0.04979$)

Solution: Here, $m = 0.9$

(*i*) $P(X \geq 1) = 1 - P(0) = 1 - e^{-0.9} = 1 - 0.4066 = 0.5934$

(*ii*) Here, $m = 0.9 \times 2 = 1.8$

$$\text{Required probability} = \frac{e^{-1.8}(1.8)^2}{2!} = 0.2677$$

(*iii*) Here, $m = 0.9 \times 4 = 3.6$

$$\text{Required probability} = P(0) + P(1) + P(2) + P(3)$$

$$= e^{-3.6} + e^{-3.6}(3.6) + \frac{e^{-3.6}(3.6)^2}{2!} + \frac{e^{-3.6}(3.6)^3}{3!}$$

$$= e^{-3.6}\left[1 + 3.6 + \frac{(3.6)^2}{2!} + \frac{(3.6)^3}{3!}\right]$$

$$= 0.5147$$

Example 52. Find the Mean and Variance of Binomial distribution with parameter n and p.

Solution: The binomial distribution is given as: $P(X = x) = {}^nC_x q^{n-x} . p^x$

$X = x$:	0	1	2	—	n
p:	q^n	$n q^{n-1} p^1$	$\frac{n(n-1)}{2 \times 1} . q^{n-2} p^2$	—	p^n

Calculation of Mean ($\overline{X}$)

$$\overline{X} = \frac{\Sigma px}{\Sigma p}$$

$\because\ \Sigma p = 1$

$$\therefore\ \overline{X} = \Sigma px = 0 . q^n + 1 . nq^{n-1} p^1 + 2 . \frac{n(n-1)}{2 \times 1} . q^{n-2} p^2 + ... + np^n$$

$$= nq^{n-1} p^1 + n(n-1) q^{n-2} p^2 + ... + np^n$$

Taking np as common

$= np\,[q^{n-1} + (n-1) q^{n-2} . p + ... + p^{n-1}]$

$= np\,[q + p]^{n-1}$ $\quad [\because (q+p)^{n-1} = q^{n-1} + ... + p^{n-1}]$

$= np\,[1]^{n-1} = np$ $\quad [\because q + p = 1]$

$\therefore\ \overline{X} = np$

Thus, the mean of the binomial distribution is np.

Calculation of Variance (σ^2)

$$\sigma^2 = \frac{\Sigma x^2 p}{\Sigma p} - \left(\frac{\Sigma xp}{\Sigma p}\right)^2$$

$\because \Sigma p = 1$

$\therefore \quad \sigma^2 = \Sigma x^2 p - (\overline{X})^2$

Now,

$$\Sigma x^2 p = 0^2.q^n + 1^2.nq^{n-1}p^1 + 2^2.\frac{n(n-1)}{2\times1}q^{n-2}p^2 + 3^2.\frac{n(n-1)(n-2)}{3\times2\times1}.q^{n-3}p^3 + \ldots\ldots\ldots. + n^2 p^n$$

$$= nq^{n-1}.p + 2n(n-1)q^{n-2}p^2 + \frac{3(n)(n-1)(n-2)}{2\times1}.q^{n-3}.p^3 + \ldots\ldots\ldots. + n^2 p^n$$

$$= np\left[q^{n-1} + 2(n-1)q^{n-2}p + \frac{3(n-1)(n-2)}{2\times1}.q^{n-3}.p^2 + \ldots\ldots\ldots. + np^{n-1}\right]$$

Breaking second, third and following terms into parts, we get

$$2(n-1)q^{n-2}p = (n-1)q^{n-2}p + (n-1)q^{n-2}p$$

$$\frac{3(n-1)(n-2)}{2\times1}.q^{n-3}p^2 = \frac{(n-1)(n-2)}{2\times1}.q^{n-3}p^2 + \frac{2(n-1)(n-2)}{2\times1}.q^{n-3}p^2$$

$$np^{n-1} = [(n-1)+1]p^{n-1} = [1+(n-1)]p^{n-1}$$

Substituting the values, we get

$$\Sigma x^2 p = np[(q^{n-1} + (n-1)q^{n-2}.p + \frac{(n-1)(n-2)}{2\times1}.q^{n-3}p^2 + \ldots\ldots\ldots. + p^{n-1})$$

$$+((n-1)q^{n-2}p + \frac{2(n-1)(n-2)}{2\times1}.q^{n-3}p^2 + \ldots\ldots\ldots. + (n-1)p^{n-1})]$$

$$= np[(q+p)^{n-1} + (n-1)pq^{n-2} + (n-2)q^{n-3}.p + \ldots\ldots\ldots. + p^{n-2})]$$

$$= np[(q+p)^{n-1} + (n-1)p((q+p)^{n-2})]$$

$$= np[(1)^{n-1} + (n-1)p(1)^{n-2}]$$

$$= np[1+(n-1)p] = np + np^2(n-1)$$

$$= np + n^2p^2 - np^2$$

$\therefore \Sigma x^2 p = np + n^2p^2 - np^2$

Substituting the values of $\Sigma x^2 p$ and Σxp in the formula of σ^2, we get

$$\sigma^2 = \Sigma x^2 p - (\overline{X})^2 = np + n^2p^2 - np^2 - n^2p^2 \quad [\because \Sigma xp = np]$$

$$= np - np^2$$

$$= np(1-p) \quad [\because q = 1-p]$$

$$= npq$$

Thus, the variance of the binomial distribution is npq.

Hence, in a Binomial Distribution, Mean = np; Variance = npq

Example 53. Show that the mean and variance are identical in a Poisson distribution.

Solution: The Poisson distribution is given as: $P(X=x) = \frac{e^{-m}.m^x}{x!}$

x :	0	1	2	3	4
p :	e^{-m}	$\frac{me^{-m}}{1!}$	$\frac{m^2e^{-m}}{2!}$	$\frac{m^3e^{-m}}{3!}$	$\frac{m^4e^{-m}}{4!}$

Calculation of Mean ($\overline{X}$)

$$\overline{X} = \frac{\Sigma px}{\Sigma p}$$

$\because \quad \Sigma p = 1$

$$\therefore \quad \overline{X} = \Sigma px = 0.e^{-m} + 1.\frac{me^{-m}}{1!} + 2.\frac{m^2e^{-m}}{2!} + 3.\frac{m^3e^{-m}}{3!} + 4.\frac{m^4e^{-m}}{4!} + \ldots..\infty$$

Take out me^{-m} as common

$$\overline{X} = me^{-m}\left(1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots.. + \infty\right)$$

$$\overline{X} = me^{-m}.e^m = m.e^{-m+m} = me^0 = m \quad [\because e^m = 1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots.. + \infty]$$

Hence, Mean of the Poisson distribution is m.

Calculation of Variance (σ^2)

$$\sigma^2 = \frac{\Sigma x^2 p}{\Sigma p} = \frac{(\Sigma xp)^2}{\Sigma p}$$

$\because \quad \Sigma p = 1$

$\therefore \quad \sigma^2 = \Sigma x^2 p - (\overline{X})^2$

$$\Sigma x^2 p = 0^2 e^{-m} + 1^2.\frac{me^{-m}}{1!} + 2^2.\frac{m^2}{2!}e^{-m} + 3^2.\frac{m^3}{3!}e^{-m} + 4^2.\frac{m^4}{4!}.e^{-m} + \ldots\infty$$

Take out me^{-m} as common

$$= me^{-m}.\left(1 + 2m + 3.\frac{m^2}{2!} + 4.\frac{m^3}{3!} + \ldots..\infty\right)$$

Breaking second, third and following terms into parts, we get

$$2m = m + m, \quad \frac{3m^2}{2!} = \frac{m^2}{2!} + \frac{2m^2}{2!} \text{ and}$$

$$\frac{4m^3}{3!} = \frac{m^3}{3!} + 3.\frac{m^3}{3!}$$

$$\Sigma x^2 p = me^{-m}.\left\{\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\dots\infty\right)+\left(m+\frac{2m^2}{2!}+\frac{3m^3}{3!}+\dots\infty\right)\right\}$$

$$= me^{-m}.\left\{e^m + m\left(1+m+\frac{m^2}{2!}+\dots\infty\right)\right\} \qquad \left[\because e^m = 1+m+\frac{m^2}{2!}+\dots\infty\right]$$

$$= me^{-m}.\{e^m + m.(e^m)\}$$

$$= me^{-m}.e^m + me^{-m}.me^m$$

$$= m + m^2$$

$$\Sigma x^2 p = m + m^2$$

Thus,
$$\sigma^2 = \Sigma x^2 p - (\bar{X})^2$$
$$= m + m^2 - (m)^2$$
$$= m + m^2 - m^2$$
$$= m$$

Hence, the variance of the Poisson distribution is m.

Thus, the mean and variance of Poisson Distribution are identical.

IMPORTANT FORMULAE

1. Binomial Distribution:

(*i*) The probability of x successes in n trials is given by:

$P(X=x) = {}^nC_x\, q^{n-x}.p^x$ where, $p = 1-q$

(*ii*) The parameter of the distribution is p.

(*iii*) The mean and S.D. of the distribution are np and $\sqrt{npq}$ respectively.

(*iv*) The theoretical frequencies of x successes are given by

$$f(x) = N.P(x) = N.({}^nC_x\, q^{n-x}.p^x)$$

where, $x = 0,1,2,\dots$

2. Poisson Distribution:

(*i*) The probability of x successes in n trials is given by

$$P(X=x) = \frac{e^{-m}.m^x}{x!}$$

where, $x = 0,1,2,\dots\infty$

(*ii*) The parameter of the distribution is m which is equal to np, *i.e.*, $m = np$

(*iii*) The mean and variance of the Poisson distribution are equal.

(*iv*) The theoretical frequencies of x successes are given by

$$f(x) = N.P(x) = N.\left(\frac{e^{-m}.m^x}{x!}\right)$$

where, $x = 0,1,2,\dots\infty$

Poisson Distribution is limiting form of Binomial Distribution when

(*a*) $n \to \infty$ (*ii*) $p \to 0$, $q \to 1$ (*iii*) $np = m$

QUESTIONS

1. What do you understand by theoretical frequency distribution? Explain the properties of Binomial, Poisson and Normal Distributions.
2. What is Binomial distribution? Discuss the conditions for application of the Binomial distribution. What are its important properties?
3. What is Poisson distribution? Explain the characteristics of Poisson distribution. Point out its role.
4. Discuss the important properties of Binomial and Poisson Distribution.
5. What is Poisson distribution? Give examples where it can be applied.
6. What is Binomial distribution? Under what conditions will Binomial distribution tend to Poisson distribution?
7. Find the Mean and Variance of Binomial distribution with parameter n and p.
8. Show that the mean and variance are identical in a Poisson distribution.
9. What are the two parameters of a binomial distribution? Define mean and variance of a binomial distribution in terms of these parameters.
10. Explain briefly the procedure of fitting (*i*) binomial distribution and (*ii*) Poisson distribution.

9 Probability Distribution—Normal

INTRODUCTION

Normal distribution is one of the most important and widely used continuous probability distribution. It is mainly used to study the behaviour of continuous random variables like height, weight and intelligence of a group of students. Normal distribution was first discovered by an English Mathematician Abraham De Moivre in 1733. But it was later rediscovered and applied by Laplace and Karl Gauss. Normal distribution is also known as Gaussian distribution after the name of Karl Gauss.

NORMAL DISTRIBUTION AS A LIMITING FORM OF BINOMIAL DISTRIBUTION

Normal distribution may be looked upon as the limiting form of binomial distribution under certain conditions:

(1) n, the number of trials is infinitely large, *i.e.*, $n \to \infty$

(2) Neither p nor q is very small.

DEFINITION OF NORMAL DISTRIBUTION

Normal distribution is defined and given by the following probability function:

$$P(X = x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{1}{2}\left(\frac{X-\bar{X}}{\sigma}\right)^2} \quad -\infty < X < +\infty$$

Where $\bar{X}$ = Mean, σ = Standard deviation, e (base of natural logarthim) = 2.7183 and $\pi = 3.1415$

Normal distribution in its standard normal variate form is given by:

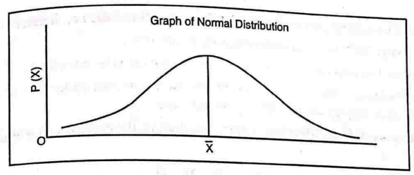
$$P(Z) = \frac{1}{\sqrt{2\pi}} . e^{-\frac{1}{2}Z^2} \quad -\infty < Z < \infty \text{ where } Z = \text{S.N.V.} = \frac{X-\bar{X}}{\sigma}$$

The mean of Z is zero and standard deviation of σ is 1.

Standard Normal Distribution (S.N.D.) is that normal distribution whose mean is zero and variance is unity.

GRAPH OF NORMAL DISTRIBUTION

The graph of the normal distribution is called normal curve. Normal curve is the graphic presentation of normal distribution. The following figure illustrates the normal curve:



The shape of normal curve depends on the values of mean ($\bar{X}$) and standard deviation (σ). There will be different shapes of normal curve for different values of mean and standard deviation.

Assumptions of Normal Distribution

The normal distribution is based on the following set of assumptions:

(1) Independent Causes: The forces affecting the event must be independent of one another, *i.e.*, they are independent of each other.

(2) Condition of Symmetry: The operation of causal forces must be such that the deviations from mean on either side is equal in number and size.

(3) Multiple Causation: The causal forces must be numerous and of approximately equal weight or importance.

Characteristics/Properties of Normal Distribution/Normal Curve

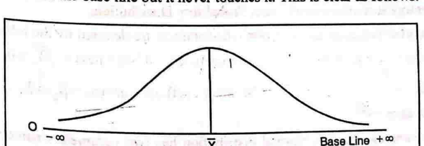
(1) Perfectly Symmetrical and Bell Shaped: The normal curve is perfectly symmetrical and bell shaped about mean. This means that if we fold the curve along its vertical axis at the centre, the two halves would coincide.

(2) Unimodal Distribution: It has only one mode, *i.e.*, it is unimodal distribution.

(3) Equality of Mean, Median and Mode: In a normal distribution, mean, median and mode are equal, *i.e.*,

$$\bar{X} = M = Z$$

(4) Asymptotic to the Base Line: Normal curve is asymptotic to the base line on either sides, *i.e.*, it has a tendency to touch the base line but it never touches it. This is clear as follows:



(5) **Range:** The normal curve extends to infinity on either side, *i.e.*, it extends $-\infty$ to $+\infty$.

(6) **Total Area:** The total area under the normal curve is 1.

(7) **Ordinate:** The ordinate of the normal curve at the mean is maximum.

(8) **Mean Ordinate:** The mean ordinate divides the whole area under the curve into two equal parts, *i.e.*, 50% on the right side and 50% on the left side.

(9) **Equidistance of Quartiles:** In a normal distribution, the quartiles Q_1 and Q_3 are equidistant from the median, *i.e.*,

$$Q_3 - M = M - Q_1$$

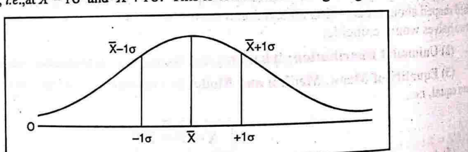
(10) **Quartile Deviation:** In a normal distribution, the quartile deviation is 2/3 times the standard deviation, *i.e.*,

$$Q.D. = \frac{2}{3} S.D.$$

(11) **Mean Deviation:** In a normal distribution, the mean deviation is 4/5 times the standard deviation, *i.e.*,

$$M.D. = \frac{4}{5} S.D.$$

(12) **Points of Inflexion:** The normal curve has two points of inflexion (*i.e.*, the points where the curve changes its curvature) at $\bar{X} - 1\sigma$ and $\bar{X} + 1\sigma$. In other words, the points of inflexion occurs at $\bar{X} \pm 1\sigma$, *i.e.*, at $\bar{X} - 1\sigma$ and $\bar{X} + 1\sigma$. This is clear from the figure given below:



(13) **Continuous Probability Distribution:** Normal distribution is a distribution of continuous variables. Therefore it is called continuous Probability Distribution.

(14) **Constants:** The constants of normal distribution are denoted by the following symbols:

Mean = $\bar{X}$ or μ or m	Moment coeff. of Skewness = $\sqrt{\beta_1} = 0$
S.D. = σ	Moment coeff. of Kurtosis = $\beta_2 = 3$
Variance = σ^2	

(15) **Main Parameters:** The normal distribution has two parameters namely mean ($\bar{X}$) and standard deviation (σ). The entire distribution can be known from these two parameters.

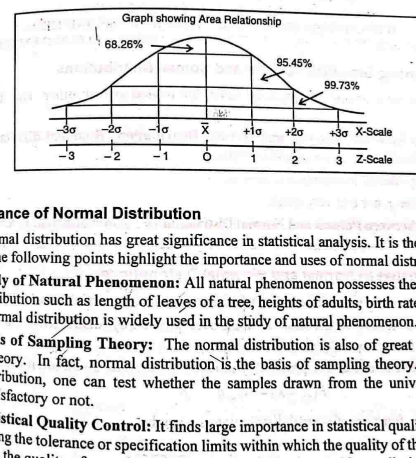
(16) **Areas Property:** One of the most important property of normal curve is the area relationship property. The total area under the normal curve is 1. It has been found that:

(*i*) Area under the normal curve between $\bar{X} - 1\sigma$ and $\bar{X} + 1\sigma$ is 0.6826, *i.e.*, Mean $\pm 1\sigma$ covers 68.26% area under the normal curve.

(*ii*) Area under the normal curve between $\bar{X} - 2\sigma$ and $\bar{X} + 2\sigma$ is 0.9545, *i.e.*, Mean $\pm 2\sigma$ covers 95.45% area under the normal curve.

(*iii*) Area under the normal curve between $\bar{X} - 3\sigma$ and $\bar{X} + 3\sigma$ is 0.9973, *i.e.*, Mean $\pm 3\sigma$ covers 99.73% area under the normal curve.

The following figure illustrate the area property:



Importance of Normal Distribution

The normal distribution has great significance in statistical analysis. It is the basis of modern statistics. The following points highlight the importance and uses of normal distribution:

(1) **Study of Natural Phenomenon:** All natural phenomenon possesses the characteristics of normal distribution such as length of leaves of a tree, heights of adults, birth rates and death rates, etc. The normal distribution is widely used in the study of natural phenomenon.

(2) **Basis of Sampling Theory:** The normal distribution is also of great importance in the sampling theory. In fact, normal distribution is the basis of sampling theory. With the help of normal distribution, one can test whether the samples drawn from the universe represent the universe satisfactory or not.

(3) **Statistical Quality Control:** It finds large importance in statistical quality control. It helps in determining the tolerance or specification limits within which the quality of the product lies. The variations in the quality of a product are acceptable within these tolerance limits.

(4) **Useful for Large Sample Tests:** The normal distribution is also widely used in case of large samples. Large sample tests are based on the properties of normal distribution.

(5) **Approximation to Binomial and Poisson Distribution:** The normal distribution serves as a good approximation to many theoretical distributions such as Binomial, Poisson, etc. As the number of observations increases, the importance of normal distribution to solve the problems relating to Binomial, Poisson, etc., increases.

Prof. Youden has expressed the importance of normal distribution in the shape of a normal curve which is shown below:

THE
NORMAL
LAW OF ERROR
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALISATIONS OF NATURAL
PHILOSOPHY. IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES.
IN THE PHYSICAL AND SOCIAL SCIENCES AND
IN MEDICINE AGRICULTURE AND ENGINEERING
IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT.

Relation among Binomial, Poisson and Normal Distributions

The binomial, poisson and normal distributions are related to each other. The relationship is shown below:

(A) Relation between Binomial and Normal Distribution: Binomial distribution tend to become normal distribution under certain conditions:

(*i*) n approaches to infinity, *i.e.*, $n \to \infty$

(*ii*) Neither p nor q is very small.

(B) Relation between Poisson and Normal Distributions: Poisson distribution tends to become normal distribution of its parameter 'm' becomes very large, *i.e.*, if $m \to \infty$, than PD tends to ND.

Difference between Normal and Binomial Distributions

The following are the main differences between normal and binomial distributions:

(1) Nature: Binomial distribution is a discrete probability distribution whereas normal distribution is a continuous probability distribution.

(2) Probability Function: The probability function of binomial distribution is given by:

$$P(X=x) = {}^nC_x \, q^{n-x} \cdot p^x$$

The probability function of normal distribution is given by:

$$P(X=x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{X-\bar{X}}{\sigma}\right)^2}$$

(3) Value of n: In a binomial distribution, the value of n, *i.e.*, number of trials is finite whereas in normal distribution, n approaches to infinity, *i.e.*, $n \to \infty$

(4) Values of p and q: In a binomial distribution, the values of p and q are approximately equal to 0.5 whereas under normal distribution, neither p nor q is very small.

(5) Parameters: The binomial distribution has two parameters n and p, whereas normal distribution has also two parameters, namely $\bar{X}$ and σ.

(6) Shape: The binomial distribution can be symmetrical and asymmetrical. It depends on the values of p and q. On the other hand, the normal distribution is always perfectly symmetrical.

Difference between Normal and Poisson Distributions

The followings are the main differences between normal and poisson distributions:

(1) Nature: Poisson distribution is a discrete probability distribution whereas normal distribution is a continuous probability distribution.

(2) Probability Function: The probability function of poisson distribution is given by:

$$P(X=x) = e^{-m} \cdot \frac{m^x}{x!}$$

Whereas the probability function of normal distribution given by:

$$P(X=x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{X-\bar{X}}{\sigma}\right)^2}$$

(3) Value of n: In both poisson and normal distributions, the value of n is very large, *i.e.*, $n \to \infty$.

(4) Values of p and q: In a poisson distribution, $p \to 0$ and $q \to 1$ whereas in normal distribution neither p nor q is very small.

(5) Parameters: The poisson distribution has only one parameter m whereas normal distribution has two parameters namely $\bar{X}$ and σ.

(6) Shape: The graph of poisson distribution is always positively skewed whereas the graph of normal distribution is perfectly symmetrical.

A Comparative Study of Binomial, Poisson and Normal Distributions

A comparative study of Binomial, Poisson and Normal distributions can be made on the basis of the following properties:

S.No.	Properties	Binomial	Poisson	Normal
1.	Nature	Discrete	Discrete	Continuous
2.	Probability Function	$P(X=x) = {}^nC_x \, q^{n-x} \cdot p^x$	$P(X=x) = e^{-m} \cdot \frac{m^x}{x!}$	$P(X=x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{X-\bar{X}}{\sigma}\right)^2}$
3.	Parameter Restrictions	n, p $0 < p < 1$	m $m > 0$	$\bar{X}, \sigma$ $-\infty \le X \le \infty$
4.	Limiting Form of BD	—	$p \to \infty$ $p \to 0$ $np \to m$	$n \to \infty$ Neither p nor q is small
5.	Mean and Variance	$\bar{X} = np$ $\sigma^2 = npq$	$\bar{X} = m$ $\sigma^2 = m$	$\bar{X}$ or m σ^2
6.	Shape	Symmetrical or Asymmetrical	Positively Skewed	Perfectly Symmetrical

How to Measure Area under the Normal Curve?

The following steps are to be followed to measure area under the normal curve:

(1) Firstly, the given value of the normal variate is transformed into standard units by substitution. The formula of Z-tranformation is given by:

$$Z = \frac{X - \overline{X}}{\sigma}$$

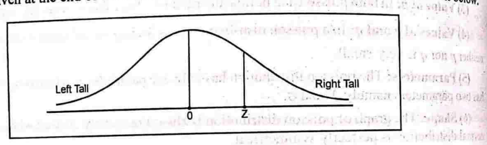
For example, if $\overline{X} = 30, \sigma = 5$ and $X = 35$, then the standard normal variate corresponding to 35 will be:

$$Z = \frac{35 - 30}{5} = 1$$

Thus, the Z-transformation of 35 will be 1.

(2) Then area is obtained from the area tables (given at the end of the book) for any particular value of Z.

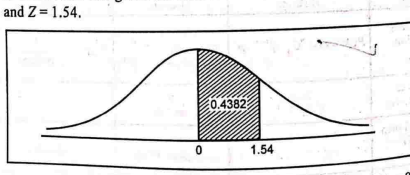
The table given at the end of the book shows the area between O to Z, which is shown below:



The following examples will show how the table of area under normal curve is consulted to find the area under the normal curve:

Example 1. Find the area under the normal curve between Z = 0 and Z = 1.54.

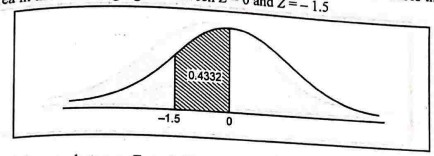
Solution: If we look to the table given at the end of the book, the entry corresponding to Z = 1.54 is 0.4382 and this gives the shaded area in the following figure between Z = 0 and Z = 1.54.



Example 2. Find the area under the normal curve between $Z = -1.5$ and $Z = 0$

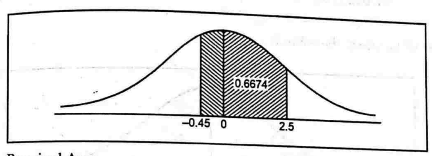
Solution: The table given at the end of the book does not contain entries corresponding to negative values of Z. But since the curve is symmetrical, we can find the area between $Z = 0$ and $Z = -1.5$ by looking the area corresponding to $Z = 0$ and $Z = -1.5$.

Therefore, the entry corresponding to $Z = 1.5$ is 0.4332 and it measures the shaded area in the following figure between $Z = 0$ and $Z = -1.5$



Example 3. Find the area between $Z = -0.45$ and $Z = 2.5$.

Solution:



Required Area

= (Area between Z = –0.45 and Z = 0) + (Area between Z = 0 and Z = 2.5)

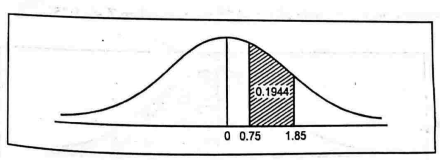
= (Area between Z = 0 and Z = 0.45) + (Area between Z = 0 and Z = 2.5)

= 0.1736 + 0.4938

= 0.6674

Example 4. Find the area under the normal curve between Z = 0.75 and Z = 1.85.

Solution:



Required Area

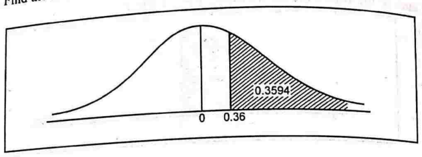
= (Area between Z = 0 and Z = 1.85) – (Area between Z = 0 and Z = 0.75)

= 0.4678 – 0.2734

= 0.1944

Example 5. Find the area to the right of Z = +0.36.

Solution:



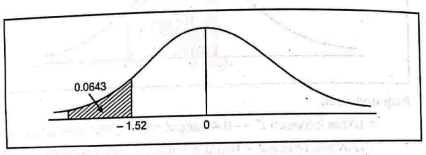
Required Area

= (Area to the right of Z = 0) – (Area between Z = 0 and Z = 0.36)

= 0.5000 – 0.1406 = 0.3594

Example 6. Find the area to the left of Z = – 1.52.

Solution:



Required Area

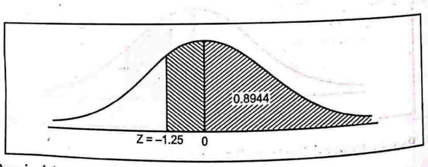
= (Area to the left of Z = 0) – (Area between Z = –1.52 and Z = 0)

= (Area to the left of Z = 0) – (Area between Z = 0 and Z = 1.52)

= 0.5000 – 0.4357 = 0.0643

Example 7. Find the area to the right of Z = –1.25 or greater than Z = –1.25.

Solution:



Required Area

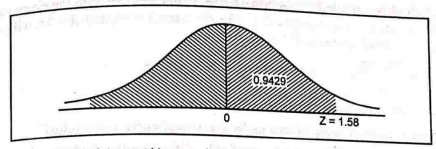
= (Area between Z = –1.25 and Z = 0) + (Area to the right of Z = 0)

= 0.3944 + 0.5000

= 0.8944

Example 8. Find the area to the left of Z = +1.58.

Solution:



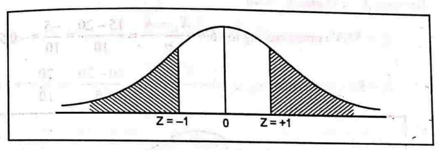
Required Area = (Area to the left of Z = 0) + (Area between Z = 0 and Z = 1.58)

= 0.5000 + 0.4429

= 0.9429

Example 9. Find the area to the right of Z = +1 and to the left of Z = –1.

Solution:



Required Area

= Total Area – (Area between Z = –1 and Z = 0) – (Area between Z = 0 and Z = +1)

= 1 – 0.3413 – 0.3413

= 1 – 0.6826 = 0.3174

EXERCISE 9.1

1. Find the area under the normal curve in the following cases using table:
   (*i*) Between Z = 0 and Z = 1.3 (*ii*) Between Z = 0.75 and Z = 0.
   (*iii*) Between Z = – 0.56 and Z = 2.45 (*iv*) Between Z = 0.85 and Z = 1.96.
   [Ans. (*i*) 0.4032 (*ii*) 0.2734 (*iii*) 0.7052 (*iv*) 0.1727**]**

Applications of Normal Distribution

The applications relating to normal distribution are studied under the following heads:

(1) Finding areas when $\overline{X}$ and σ of normal variate are given.

(2) Finding mean and standard deviation when the area is given.

(3) Finding minimum and maximum score amongst highest and lowest group.

(4) Fitting of Normal Curve.

▶ **(1) Finding areas when $\overline{X}$ and σ of normal variate are given**

In order to find the area under the normal curve, firstly we transform the given value of normal variate in to the Z-variate. For example, if $\overline{X} = 30, \sigma = 5$ and $X = 35$, then $X = 35$, will be transformed into the standard normal variate as follows:

$$Z = \frac{35-30}{5} = 1 \qquad \text{where, } Z = \frac{X - \overline{X}}{\sigma}$$

Thus, for X=35, the standard normal variate (SNV) is 1.

After Z-transformation, table of area under the normal curve is consulted.

The following examples illustrate how the **Table of Area under the Normal curve** is consulted to find the area under the normal curve.

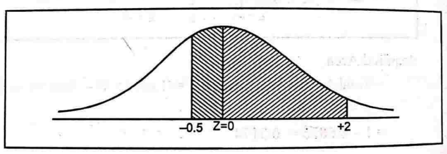
Example 10. A normal curve has $\overline{X} = 20$ and $\sigma = 10$. Find the area between $X_1 = 15$ and $X_2 = 40$

Solution: Given $\overline{X} = 20$, $\sigma = 10$

Between $X_1 = 15$ and $X_2 = 40$

$$Z_1 = \text{SNV corresponding to } 15 = \frac{X_1 - \overline{X}}{\sigma} = \frac{15-20}{10} = \frac{-5}{10} = -0.5$$

$$Z_2 = \text{SNV corresponding to } 40 = \frac{X_2 - \overline{X}}{\sigma} = \frac{40-20}{10} = \frac{20}{10} = +2.0$$



Required Area = Area between (Z = –0.5 and Z = 0) + Area between (Z = 0 and Z = +2)

= 0.1915 + 0.4772 = 0.6687

Example 11. An aptitude test for selecting officers in a bank was conducted on 1,000 candidates, the average score is 42 and the standard deviation of scores is 24.

Assume normal distribution for the scores, find

(*i*) the number of candidates whose score exceed 60

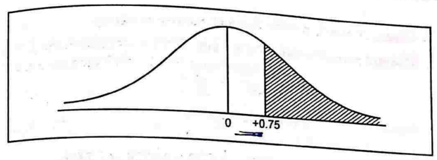
(*ii*) the number of candidates whose score lie between 30 and 66.

Solution: Given, $\overline{X} = 42, \sigma = 24$, N = 1000

(*i*) Exceeding 60

Z = SNV corresponding to 60

$$= \frac{X - \overline{X}}{\sigma} = \frac{60-42}{24} = \frac{18}{24} = \frac{3}{4} = +0.75$$



Required Proportion

= Area to the right of (Z = 0) – Area between (Z = 0 and Z = +0.75)

= 0.5000 – 0.2734 = 0.2266

Number of candidates whose score exceeds 60

= 1,000 × 0.2266 = 226.6 ≈ 227

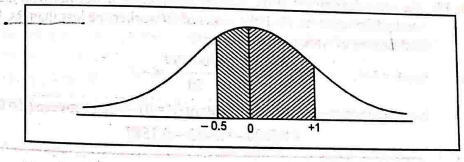
(*ii*) Between 30 and 66

Z_1 = SNV corresponding to 30

$$= \frac{X_1 - \overline{X}}{\sigma} = \frac{30-42}{24} = \frac{-12}{24} = -05$$

Z_2 = SNV corresponding to 66

$$= \frac{X_2 - \overline{X}}{\sigma} = \frac{66-42}{24} = \frac{24}{24} = +1$$



Required Proportion

= Area between (Z = – 0.5 and Z = 0) + Area between (Z = 0 and Z = +1)

= 0.1915 + 0.3413

= 0.5328

Number of candidates whose score lie between 30 and 66

= 1,000 × 0.5328 = 532.8 or 533

Example 12. The monthly income distribution of workers in a certain factory was found to be normal with mean Rs 500 and standard deviation equal to Rs 50. There were 228 persons getting income above Rs 600 per month. How many workers were there in all? Extract of area under standard normal is given below:

Z:	1	2	2.5	3
Area:	0.3413	0.4772	0.4938	0.4987

Solution: Given, $\overline{X} = 500$, $\sigma = 50$, $X = 600$

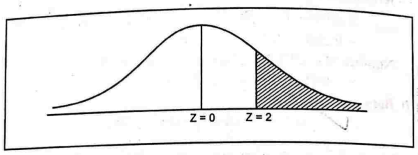
Standard normal variance for $X = 600$

$$Z = \frac{600-500}{50} = \frac{100}{50} = 2$$

Area between $Z = 0$ and $Z = 2$ is 0.4772

Area to the right of $Z = 2$

$$= 0.5000 - 0.4772 = 0.0228 \text{ or } 2.28\%$$



For area 0.0228, the number of workers are 228

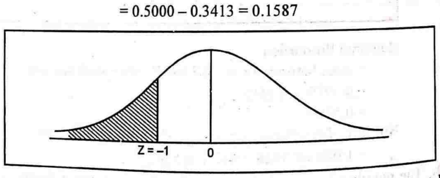
For area 1.00; the number of workers are $= \frac{228}{0.0228} \times 1 = 10{,}000$ workers

Example 13. The wage distribution of the workers in a factory is normal with mean Rs 400 and standard deviation Rs. 50. If the wages of 40 workers be less than Rs. 350, what is the total number of workers in the factory?

Solution: For $X = 350$, $Z = \frac{X-400}{\sigma} = \frac{350-400}{50} = -1$

Required Proportion = Area to the left of $(Z = 0)$ – Area between $(Z = 0$ and $Z = -1)$

$$= 0.5000 - 0.3413 = 0.1587$$



Number of workers drawing less than Rs. 350 = 40 = 0.1587 of total number of workers.

$\therefore$ Total number of workers $= \frac{40}{0.1587}$

$$= 252.04 \approx 252$$

Example 14. Assume mean height of soldiers to be 68.22 inches with a variance of 10.8 inches, how many soldiers in a regiment of 1000 would you expect to be over six feet tall?

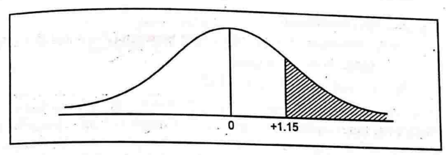
Solution: We are given, $\overline{X} = 68.22$, $\sigma^2 = 10.8 \rightarrow \sigma = \sqrt{10.8} = 3.28$

Above 6 feet (*i.e.*, 72 inches)

For $X = 72$, $Z = \frac{X-\overline{X}}{\sigma} = \frac{72-68.22}{3.28} = \frac{3.78}{3.28} = 1.15$

Required Proportion = (Area to the right of $Z = 0$) – (Area between $Z = 0$ and $Z = 1.15$)

$$= 0.5000 - 0.3749$$
$$= 0.1251$$



Thus, the expected number of soldiers having height above 6 feet

$$= 1000 \times 0.1251$$
$$= 125.1 \approx 125$$

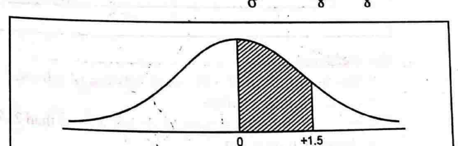
Example 15. Find the probability that an item drawn at random from a normal distribution with mean = 70 and S.D. = 8 will be (*i*) between 70 and 82, and (*ii*) more than 82.

Solution: Given, $\overline{X} = 70$, $\sigma = 8$

(*i*) Between 70 and 82

Z_1 = SNV corresponding to $70 = \frac{X_1 - \overline{X}}{\sigma} = \frac{70-70}{8} = 0$

Z_2 = SNV corresponding to $82 = \frac{X_2 - \overline{X}}{\sigma} = \frac{82-70}{8} = \frac{12}{8} = +1.5$



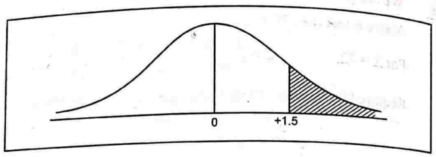
Required Probability = Area between ($Z = 0$ and $Z = 1.5$)

$$= 0.4332$$

Hence, 43.32% items lie between 70 and 82.

(*ii*) **Beyond 82** (*i.e.*, exceeding 82)

SNV(Z) corresponding to 82 $= \dfrac{X - \overline{X}}{\sigma} = \dfrac{82 - 70}{8} = \dfrac{12}{8} = +1.5$



Required Probability

= Area to the right of (Z = 0) – Area between (Z = 0 and Z = 1.5)

= 0.5000 – 0.4332 = 0.0668

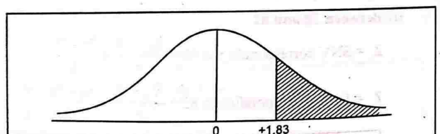
Hence, 6.68% items lie beyond 82.

Example 16. As a result of tests on 20,000 electric bulbs manufactured by a company it was found that the life time of a bulb was normally distributed with an average life of 2040 hours and standard deviation of 60 hours. On the basis of the information, estimate the number of bulbs that is expected to burn for (*i*) more than 2150 hours, and (*ii*) less than 1960 hours.

Solution: Given, N = 20,000, $\overline{X} = 2040$, $\sigma = 60$

(*i*) **More than 2150 hours**

For, X = 2150, $Z = \dfrac{X - \overline{X}}{\sigma} = \dfrac{2150 - 2040}{60} = \dfrac{110}{60} = +1.83$



Required Proportion

= Area to the right of (Z = 0) – Area between (Z = 0 and Z = 1.83)

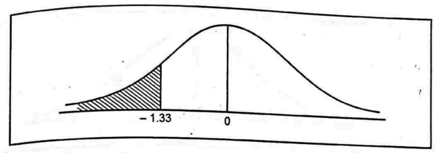
= 0.5000 – 0.4664 = 0.0336

Thus the number of bulbs expected to burn for more than 2150 hours

= 20,000 × 0.0336 = 672

(*ii*) **Less than 1960 hours**

For, X = 1960, $Z = \dfrac{X - \overline{X}}{\sigma} = \dfrac{1960 - 2040}{60} = \dfrac{-80}{60} = -1.33$



Required Proportion

= Area to the left of (Z = 0) – Area between (Z = –1.33 and Z = 0)

= 0.5000 – 0.4082

= 0.0918

Thus, the number of bulbs expected to burn for less than 1960 hours

= 20,000 × 0.0918

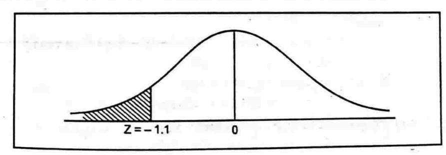
= 1836

Example 17. Net profit of 400 companies is normally distributed with a mean profit of Rs. 150 lakhs and a standard deviation of Rs. 20 lakhs. Find the number of companies whose profits (Rs. lakhs) are (*i*) less than 128, (*ii*) more than 175 and (*iii*) between 100 and 138. Also find the minimum profit of top 15% companies.

Solution: Given, N = 400, $\overline{X} = 150$, $\sigma = 20$

(*i*) **Less than 128**

For X = 128, $Z = \dfrac{128 - 150}{20} = \dfrac{-22}{20} = -1.1$



Required Proportion

= Area to the left of (Z = 0) – Area between (Z = – 1.1 and Z = 0)

= 0.5000 – 0.3643

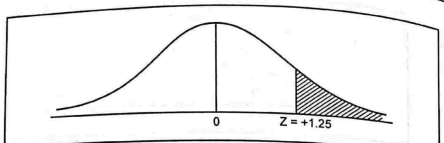
= 0.1357

Number of companies = 0.1357 × 400

= 54.28 = 54 approx.

(*ii*) **More than 175**

For X = 175, $Z = \dfrac{175 - 150}{20} = \dfrac{25}{20} = +1.25$



Required Proportion

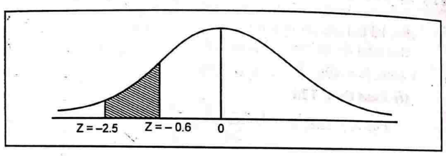
= Area to the right of (Z = 0) – Area between (Z = 0 and Z = 1.25)
= 0.5000 – 0.3944 = 0.1056

Number of companies = 0.1056 × 400 = 42.24 = 42 app.

***(iii)* Between 100 and 138**

For X = 100, $Z_1 = \dfrac{100-150}{20} = -\dfrac{50}{20} = -2.5$

For X = 138, $Z_2 = \dfrac{138-150}{20} = -\dfrac{12}{20} = -0.6$



Required Proportion

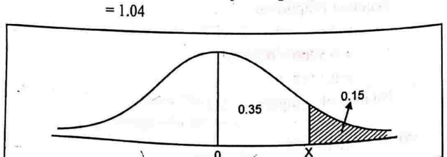
= Area between (Z = –2.5 and Z = 0) – Area between (Z = – 0.6 and Z = 0)
= 0.4938 – 0.2257 = 0.2681

No. of Companies = 0.2681 × 400
= 107.24 = 107 approx.

***(iv)* Proportion of Top Companies = 0.15**

Value of Z having 0.15 area of its right

= Value of Z corresponding to (0.50 – 0.15), *i.e.,* 0.35 area
= 1.04



$$\therefore \quad Z = \frac{X - \bar{X}}{\sigma}$$

$$1.04 = \frac{X - 150}{20}$$ [From the table, Z = 1.04 be 0.35 area]

$1.04 \times 20 = X - 150$

or $20.8 = X - 150$

or $X = 170.8 = 171$ approx.

Hence, the minimum profit of top 15% companies is Rs. 171 lakhs.

EXERCISE 9.2

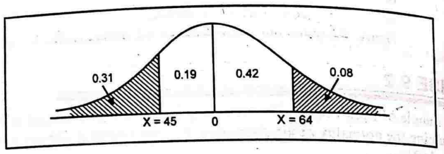
1. In a sample of 1000 workers, the mean weight is 45 kg with a standard deviation of 15 kg. Assuming the normality of the distribution, find the number of workers weighing between 40 and 60 kgs. **[Ans. 471]**
2. A sample of 100 dry battery cells was tested and found the mean life 12 hours and standard deviation 3 hours. Assume that the data to be normally distributed, what percentage of battery cells are expected to have (*i*) more than 15 hours, (*ii*) Between 10 and 14 hours and (*iii*) less than 6 hours. **[Ans. (*i*) 15.87% (*ii*) 49.72% (*iii*) 2.28%]**
3. In an entrance test for admission, 1000 students appeared. Their average marks were 45 and standard deviation 10. Find (*i*) the number of students securing beween 40 and 50 (*ii*) number of students exceeding the score 60 (*iii*) the value of score exceeded by top 100 students. **[Ans. (*i*) 383 (*ii*) 67 (*iii*) 58]**
4. Find the probability that an item drawn at random from the normal distribution with mean 5 and S.D. 3 will be between 2.57 and 4.34. **[Ans. 0.2039]**
5. A normal distribution has mean (μ)=12 and standard deviation (σ) = 2. Find the area between X=9.6 and X= 13.8. **[Ans. 70.08%]**
6. For a normal distribution, mean = 12 and standard deviation = 2, find the area under the curve from X= 6 to X= 18. **[Ans. 99.74%]**
7. The Ambala Municipality installed 3,000 electric tubes in various streets at a particular moment of time. If the average life of electric tube is 1,200 burning hours with S.D. of 250 hours, find the expected number of electric tubes: (*i*) that might be expected to be fused in first 700 burning hours and (*ii*) expected to be good after 1950 burning hours. **[Ans. (*i*) 68 (*ii*) 4]**
8. A Sales Tax Officer has reported that the average sales of 500 businessmen that he has to deal with during a year amount to Rs. 36,000 with a S.D. of Rs. 10,000. Find out: (*i*) the number of businessmen, the sales of which are over Rs. 40,000 (*ii*) the percentage of businessmen the sales of which are likely to range between Rs. 30,000 and Rs. 40,000. **[Ans. (*i*) 172.3 (*ii*) 38.11%]**

▶ **(2) Finding $\overline{X}$ and σ when the area under normal curve is given**

When the area under the normal curve is given, then we can find the mean ($\overline{X}$) and standard deviations (σ) of the normal distribution. The following examples illustrate the procedure:

Example 18. In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find $\overline{X}$ and σ of the distribution.

Solution:



$$Z = \frac{X - \overline{X}}{\sigma}$$

Value of Z corresponding to 0.50 – 0.31 = 0.19 area = – 0.5 (From the table)

$$\therefore \quad -0.5 = \frac{45 - \overline{X}}{\sigma}$$

or $\quad -05\sigma = 45 - \overline{X}$

or $\quad \overline{X} - 0.5\sigma = 45 \qquad$ ...(i)

Value of Z corresponding to 0.5 – 0.08 = 0.42 area = +1.41 (From the table)

$$\therefore \quad 1.41 = \frac{64 - \overline{X}}{\sigma}$$

or $\quad 1.41\sigma = 64 - \overline{X}$

or $\quad \overline{X} + 1.41\sigma = 64 \qquad$ ...(ii)

Solving the two equations

$\overline{X} - 0.5\sigma = 45$

$\overline{X} + 1.41\sigma = 64$

– – –

$-1.91\sigma = -19$

⇒ $\quad \sigma = 10$

Substituting the value of σ in equation (i)

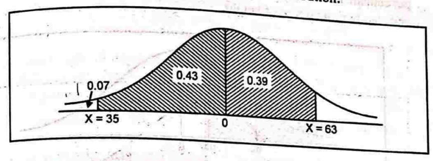
$\overline{X} - 0.5(10) = 45$

$\overline{X} - 5 = 45$

or $\quad \overline{X} = 50$

$\therefore \overline{X} = 50, \ \sigma = 10$

Example 19. In a distribution exactly normal 7% of the items are under 35 and 89% are under 63. Find the mean and standard deviation of the distribution:



Solution: Value of Z corresponding to 0.43 area = – 1.48

$$\therefore \quad -1.48 = \frac{35 - \overline{X}}{\sigma}$$

or $\quad -1.48\sigma = 35 - \overline{X}$

or $\quad \overline{X} - 1.48\sigma = 35 \qquad$ ...(i)

Value of Z corresponding to 0.39 area = +1.23

$$\therefore \quad 1.23 = \frac{63 - \overline{X}}{\sigma}$$

or $\quad -1.23\sigma = 63 - \overline{X}$

or $\quad \overline{X} + 1.23\sigma = 63 \qquad$ ...(ii)

Solving the two equations

$\overline{X} - 1.48\sigma = 35$

$\overline{X} + 1.23\sigma = 63$

– – –

$-2.71\sigma = -28$

or $\quad 2.71\sigma = 28$

or $\quad \sigma = \frac{28}{2.71}$

$= 10.33$

Substituting the value of σ in (i)

$\overline{X} - 1.48(10.33) = 35$

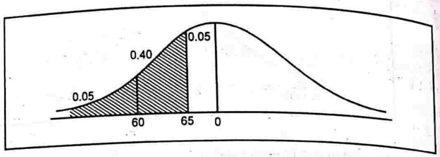
$\overline{X} - 15.3 = 35$

⇒ $\quad \overline{X} = 50.3$

Thus, $\overline{X} = 50.3, \ \sigma = 10.33$

Example 20. In a large group of men, it is found that 5 per cent are under 60 inches in height and 40 per cent are between 60 and 65 inches. Assuming a normal distribution, find the mean and standard deviation of height.

Solution:



The value of Z corresponding to 0.45 (0.5000 – 0.05) area = –1.65. The negative sign with the value of Z is taken as the value on the LHS of the mean of the distribution.

$\therefore \quad -1.65 = \dfrac{60-\overline{X}}{\sigma}$

or $\quad -1.65\sigma = 60-\overline{X}$

or $\quad \overline{X}-1.65\sigma = 60$ ...(i)

The value of Z corresponding to 0.05 area = – 0.13

$\therefore \quad -0.13 = \dfrac{63-\overline{X}}{\sigma}$

or $\quad -0.13\sigma = 65-\overline{X}$

or $\quad \overline{X}-0.13\sigma = 65$ ...(ii)

Solving the two equations

$$\begin{array}{r} \overline{X}-1.65\sigma = 60 \\ \overline{X}-0.13\sigma = 65 \\ -\quad +\quad - \\ \hline -1.52\sigma = -5 \end{array}$$

or $\quad 1.52\sigma = 5 \qquad \therefore \ \sigma = \dfrac{5}{1.52} = 3.29$

Substituting the value of σ in (i)

$\overline{X}-1.65(3.29) = 60 \qquad \overline{X}-5.4285 = 60$

$\overline{X} = 65.4285 \approx 65.42 \qquad \therefore \ \overline{X} = 65.42,\ \sigma = 3.29$

EXERCISE 9.3

1. The marks obtained by the students in an examination are known to be normally distributed. If 10% of the students got less than 40 marks while 15% got over 80, what are the mean and standard deviation of marks? **[Ans.** $\overline{X} = 62.15, \sigma = 17.16$**]**

2. In a certain examination, 15% of the candidates passed with distinction while 25% of them failed. It is known that a candidate fails if he obtains less than 40 marks (out of 100) while he must obtain at least 75 marks in order to pass with distinction. And the mean and standard deviation of the distribution of marks assuming this to be normal. **[Ans.** $\overline{X} = 53.79, \sigma = 20.46$**]**
3. Assuming that height of a group of men is normal, find the mean and standard deviation, given that 84% of men have heights less than 65.2 inches and 68% have heights between 65.2 and 62.8 inches. **[Ans.** $\overline{X} = 64, \sigma = 1.2$**]**
4. In a certain examination the percentage of passes and distinctions were 46 and 9 respectively. Estimate the average marks obtained by the candidates and their standard deviation, the minimum pass and distinction marks being 40 and 75 respectively (Assume the distribution of marks to be normal). **[Ans.** $\overline{X} = 37.18, \sigma = 28.22$**]**

(3) Finding Minimum and Maximum Score Amongst the Highest and Lowest Group

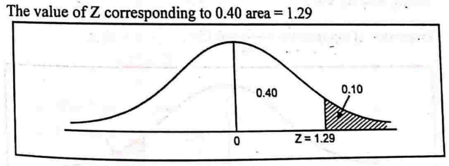
When the $\overline{X}$, σ and proportion of highest and lowest groups are given, then we can find the minimum and maximum score amongst the highest and lowest group. The following examples illustrate the procedure:

Example 21. The wages of 5,000 workers were found to be normally distributed with mean Rs. 2,000 and standard deviation Rs. 120. What was the lowest wages amongst the richest 500 workers?

Solution: Given, N = 5,000, $\overline{X}$ = 2000, σ = 120

Proportion of richest workers $= \dfrac{500}{5000} = \dfrac{1}{10} = 0.10$

The value of Z corresponding to 0.40 area = 1.29



We know that: $\quad Z = \dfrac{X-\overline{X}}{\sigma}$

$$1.29 = \frac{X-2000}{120}$$

$\Rightarrow \quad X - 2000 = 154.8$

$\Rightarrow \quad X = \text{Rs. } 2154.8$

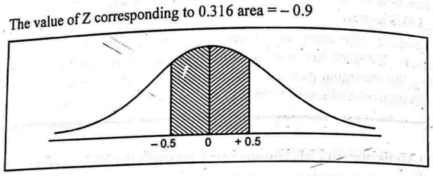
Thus, the lowest wages among the richest 500 workers is Rs. 2154.8.

Example 22. The mean and standard deviation of a graduation examination, following normal distribution are 500 and 100 respectively. If 550 students are to be passed out of 674 students, what would be the minimum passing marks?

Solution: Given, N = 674, $\overline{X} = 500, \sigma = 100$

Proportion of passed students $= \dfrac{550}{674} = 0.816$

The value of Z corresponding to 0.316 area = – 0.9



$$Z = \frac{X - \overline{X}}{\sigma} \Rightarrow -0.9 = \frac{X - 500}{100}$$

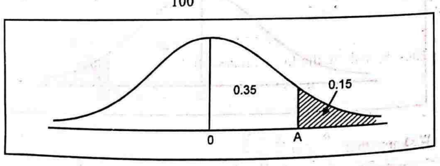
or $\quad -90 = X - 500 \quad$ or $X = 410$

Hence the minimum passing marks are 410.

Example 23. The marks of students in a class are normally distributed with $\overline{X} = 70$ and S.D. = 5. If the instructor decides to give 'A' grade to top 15% students, how many marks a student must get to be able to get 'A' grade.

Solution: Given, $\overline{X} = 70, \sigma = 5$

Proportion of top students $= \dfrac{15}{100} = 0.15$



The value of Z corresponding to 0.35 area = 1.04.

We know that

$$Z = \frac{X - \overline{X}}{\sigma} \Rightarrow 1.04 = \frac{X - 70}{5} \Rightarrow 5.20 = X - 70$$

$\therefore X = 75.2$

Thus, to get 'A' grade, one must secure 75 marks or more.

Example 24. A set of examination marks is approximately normally distributed with a mean of 75 and standard deviation of 5. If the top 5% of the students get grade A and the bottom 25% get grade F, what is the lowest A and what mark is the highest F?

Solution: Given, $\overline{X} = 75, \sigma = 5$

$$Z = \frac{X - \overline{X}}{\sigma}$$

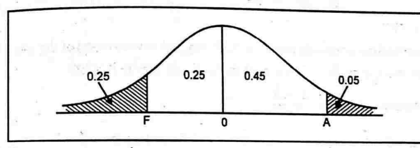
Value of Z corresponding to 0.5 – 0.25 = 0.25 area = – 0.68 (From the table)

$\therefore \quad -0.68 = \dfrac{X - 75}{5}$

or $\quad -0.68 \times 5 = X - 75$

or $\quad -3.4 = X - 75$

or $\quad X = 71.6$ or 72.



Thus, 72 will the highest marks of the bottom 25% of the students.

Value of Z corresponding to (0.50 – 0.05) = 0.45 area = 1.65 (From the table)

$\therefore \quad 1.65 = \dfrac{X - 75}{5}$

or $\quad 1.65 \times 5 = X - 75$

or $\quad 8.25 = X - 75$

or $\quad X = 83.25$ or 83

Thus, 83 will be the lowest marks of the top 5% of the students.

Thus, the lowest marks of the top 5% would be 83 and the highest marks of the bottom 25% students would be 72.

EXERCISE 9.4

1. The monthly incomes of 500 workers were found to be normally distributed with the mean Rs 2,000 and a standard deviation of Rs 200. What was the lowest income among the richest 125 workers? **[Ans. Rs. 2,134]**
2. The incomes of a group of 5,000 persons were found to be normally distributed with mean = Rs. 900 and S.D. = Rs. 75. What was the highest income among the poorest 200? (Given: Area under the standard normal curve from $Z = 0$ to $Z = 1.75$ is 0.46) **[Ans. Rs. 768.75]**

3. The marks of students in a class are normally distributed with $\overline{X}$= 6.7 and S.D. = 1.2. Assuming the marks to be normally distributed, determine the maximum marks of the lowest 10% of the class. [Ans. 5.164 or 5]
4. In an intelligence test administered to 1,000 students, the average score was 42 and standard deviation is 24. If the top 10% of the students get grade A, how many minimum marks a student must get to be able to get grade 'A'. [Ans. 72.72 or 73]

(4) Fitting of Normal Curve

There are two methods for fitting the normal curve:

(1) Ordinate Method

(2) Area Method

(1) Ordinate Method

This method uses the Table of Ordinates of the Standard Normal Curve. This method involves the following steps:

(*i*) First, we find arithmetic mean ($\overline{X}$) and standard deviation (σ) of the given distribution.

(*ii*) Find the mid points of each class interval and denote it by *X*.

(*iii*) For each X, find $Z = \dfrac{X - \overline{X}}{\sigma}$

(*iv*) Find ordinates at each of these value of Z from the table of ordinates.

(*v*) Multiply each of these values by $N \times \dfrac{i}{\sigma}$ and we find the expected frequencies.

Here, N = number of items, *i* = size of class interval, σ = S.D.

The following example illustrate the procedure of fitting the normal curve.

Example 25. Fit a normal curve to the following data by the method of ordinates:

Variable :	0—10	10—20	20—30	30—40	40—50
Frequency:	3	5	8	3	1

Solution: For fitting the normal curve, we compute $\overline{X}$ and σ

Computation of $\overline{X}$ and σ

Variable	*f*	M.V. (X)	*d*	$d' = d/i$	fd'	fd'^2
0—10	3	5	− 20	− 2	− 6	12
10—20	5	15	− 10	− 1	− 5	5
20—30	8	25	0	0	0	0
30—40	3	35	+ 10	+ 1	+ 3	3
40—50	1	45	+ 20	+ 2	+ 2	4
	N = 20				$\Sigma fd' = -6$	$\Sigma fd'^2 = 24$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 25 + \frac{-6}{20} \times 10 = 22$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i$$

$$= \sqrt{\frac{24}{20} - \left(\frac{-6}{20}\right)^2} \times 10 = 10.53$$

After finding the $\overline{X}$ and σ, we adopt the following procedure.

Variable (1)	M.V (X) (3)	$Z = \dfrac{X - \overline{X}}{\sigma}$ (4)	Value of Ordinate from Ordinate Table (5)	$fe = \dfrac{\text{Ordinate} \times N \times i}{\sigma}$ (6)
0—10	5	− 1.61	0.1092	2.07 = 2
10—20	15	− 0.66	0.3209	6.09 = 6
20—30	25	0.28	0.3836	7.28 = 7
30—40	35	1.23	0.1872	3.55 = 4
40—50	45	2.18	0.0371	0.7046 = 1
				N = 20

(2) Area Method

This method uses the table of area under the Standard Normal Curve. It involves the following steps:

(*i*) First, we find $\overline{X}$ and σ of the given distribution.

(*ii*) Write the lower limit of the each class interval and denote it by '*X*'.

(*iii*) For each lower class limit *X*, find $Z = \dfrac{X - \overline{X}}{\sigma}$.

(*iv*) Find the area at each of these values of Z from the area table.

(*v*) Then the successive difference between the two area values are computed. These are obtained by subtracting the successive area obtained when the corresponding Z's have the same sign and adding them when the Z's have opposite sign.

(*vi*) Multiple each of these by N to find the expected frequencies.

The following example illustrate the procedure.

Example 26: Fit a normal curve to the following data:

Variable :	0—10	10—20	20—30	30—40	40—50
Frequency:	3	5	8	3	1

Solution: From the above example, we find that

$\overline{X}$ = 22, σ = 10.53, N = 20,

After finding the $\overline{X}$ and σ, we adopt the following procedure:

Variable	Lower Class Limit (X)	$Z = \frac{X - \overline{X}}{\sigma}$	Area from 0 to Z	Area of each Class Interval	$fe = N \times Area$
(1)	(2)	(3)	(4)	(5)	(6)
0—10	0	−2.09	0.4817	0.1088	2.17 ≈ 2
10—20	10	−1.14	0.3729	0.2975	5.95 ≈ 6
20—30	20	−0.19	0.0753	0.3518	7.036 ≈ 7
30—40	30	0.76	0.2764	0.1800	3.6 ≈ 4
40—50	40	1.71	0.4564	0.0397	0.794 ≈ 1
50—60	50	2.66	0.4961		N = 20

EXERCISE 9.5

1. (*i*) Name the two methods available to fit the normal curve.

   (*ii*) Fit a normal curve to the following data by Ordinate Method.

Class Interval :	0—10	10—20	20—30	30—40	40—50
f:	5	8	12	8	7

[**Ans.** f = 3, 9, 14, 10, 4]

2. Fit a normal curve to the following date by area method:

Class Interval :	10.5—20.5	20.5—30.5	30.5—405	40.5—50.5	50.5—60.5	60.5—70.5	70.5—80.5
f:	12	28	40	60	32	20	8

[**Ans.** f = 9, 26, 45, 54, 40, 20, 6]

3. Fit a normal curve to the following data:

Mid Values :	61	64	67	70	73
f:	5	18	42	27	8

[**Ans.** f = 4, 20, 41, 28, 7]

4. Fit a normal curve to the following data:

Height (cm):	60—62	63—65	66—68	69—71	72—74
No. of Students	5	18	42	27	8

Given that $\overline{X}$ = 67.45 cm and σ = 2.92 cm., N = 100 [**Ans.** f = 4, 20, 41, 28, 7]

Normal Distribution as an Approximation to Binomial Distribution

The normal distribution can be used as an approximation to binomial distribution when it is difficult or almost impossible to calculate the probabilities of the events. When such approximation is used, it is desirable to make correction for continuity.

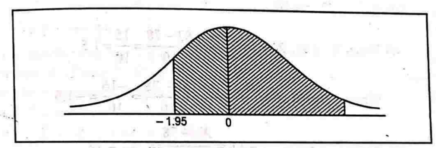
Example 27. A fair coin is tossed 400 times. Using normal approximation to the binomial, find the probability that a head will occur (*i*) more than 180 times and (*ii*) less than 195 times.

Solution: $\overline{X} = np = 400\left(\frac{1}{2}\right) = 200$ and $\sigma = \sqrt{npq} = \sqrt{400 \times \frac{1}{2} \times \frac{1}{2}} = 10$

If X is used to denote the number of heads, X is a discrete variable. The use of normal approximation, therefore, requires application of continuity factor. Thus, we have the following:

(*i*) More than 180

SNV corresponding to 180 = $\frac{180.5 - 200}{10} = -1.95$



Required Probability

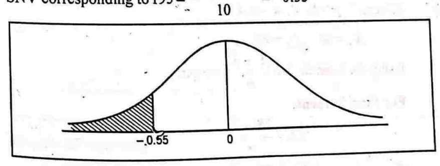
= Area between (Z = −1.95 and Z = 0) + Area to the right of (Z = 0)

= 0.4744 + 0.5000

= 0.9744

(*ii*) Less than 195

SNV corresponding to 195 = $\frac{194.5 - 200}{10} = -0.55$



Required Probability

= Area to the left of (Z = 0) − Area between (Z = − 0.55 and Z = 0)

= 0.5000 − 0.2088 = 0.2912

EXERCISE 9.6

1. A coin is tossed 400 times. Using normal approximation to the binomial, find the probability that the number of heads lies between 190 and 210. **[Ans. 0.7062]**
2. How would you use the normal distribution to find approximately the frequency of exactly 5 successes in 100 trials, the probability of success in each trial being $p = 0.1$.
[**Hint:** $\bar{X} = 100 \times 0.01 = 10, \sigma = \sqrt{100 \times 0.01 \times 0.9} = 3$] **[Ans. 3.32]**

MISCELLANEOUS SOLVED EXAMPLES

Example 28. On a Statistics examination, the mean score was 78 and S.D. was 10. Determine (*i*) standard score in terms of standard units of school boys whose score were 93 and 62 respectively, (*ii*) the score of two students whose standard scores were –0.6 and 1.4 respectively.

Solution: Given, $\bar{X} = 78, \sigma = 10$

(*i*) When $X_1 = 93$, $Z_1 = \dfrac{X_1 - \bar{X}}{\sigma} = \dfrac{93 - 78}{10} = \dfrac{15}{10} = 1.5$

When $X_2 = 62$, $Z_2 = \dfrac{X_2 - \bar{X}}{\sigma} = \dfrac{62 - 78}{10} = \dfrac{-16}{10} = -1.6$

(*ii*) When $Z_1 = -0.6$, $\therefore -0.6 = \dfrac{X_1 - 78}{10} \Rightarrow X_1 = 72$

When $Z_2 = 1.4$, $\therefore 1.4 = \dfrac{X_2 - 78}{10} \Rightarrow X_2 = 92$

Example 29. Two students were informed that they received standard scores of 0.8 and –0.4 respectively on a multiple choice examination in commerce. If their marks were 88 and 64 respectively, find the mean and standard deviation of the examination marks.

Solution: Given: $Z_1 = 0.8$, $Z_2 = -0.4$

$X_1 = 88$, $X_2 = 64$

Using the formula $Z = \dfrac{X - \bar{X}}{\sigma}$, we get:

For First Student:

$$0.8 = \frac{88 - \bar{X}}{\sigma}$$

or $0.8\sigma = 88 - \bar{X}$

or $\bar{X} + 0.8\sigma = 88$ ...(*i*)

For Second Student:

$$-0.4 = \frac{64 - \bar{X}}{\sigma}$$

or $-0.4\sigma = 64 - \bar{X}$

or $\bar{X} - 0.4\sigma = 64$ ...(*ii*)

Solving the two equations

$$\begin{aligned} \bar{X} + 0.8\sigma &= 88 \\ \bar{X} - 0.4\sigma &= 64 \\ - \quad + \quad &\quad - \\ \hline 1.2\sigma &= 24 \end{aligned}$$

$$\sigma = \frac{24}{1.2} = 20$$

Substituting the value of σ in equation (*i*), we get

$\bar{X} + 0.8(20) = 88$

$\bar{X} = 88 - 16 = 72$

$\therefore \bar{X} = 72, \sigma = 20$

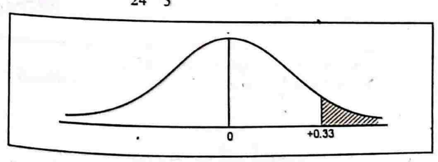
Example 30. In an intelligence test administered to 1000 students, the average score was 42 and standard deviation 24. Find: (*i*) the number of students exceeding a score 50 (*ii*) the number of students lying between 30 and 54 (*iii*) the value of the score exceeded by top 100 students.

Solution: Given, $\bar{X} = 42, \sigma = 24, N = 1000$

(*i*) Exceeding 50

For $X = 50$, $Z = \dfrac{X - \bar{X}}{\sigma} = \dfrac{50 - 42}{24}$

$= \dfrac{8}{24} = \dfrac{1}{3} = +0.33$



Required Proportion

= (Area to the right of Z = 0) – (Area between Z = 0 and Z = 0.33)

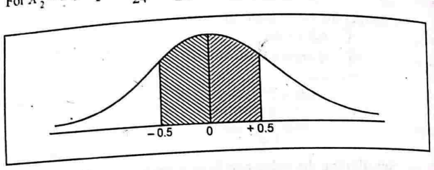
= 0.5000 – 0.1293

= 0.3707

Number of students getting more than 50 marks = 0.3707 × 1000 = 370.7 ≈ 371 *(ii)* **Between 30 and 54**

For $X_1 = 30$, $Z_1 = \frac{30-42}{24} = \frac{-12}{24} = -0.5$

For $X_2 = 54$, $Z_2 = \frac{54-42}{24} = \frac{12}{24} = +0.5$



Required Proportion

= (Area between Z = – 0.5 and Z = 0) + (Area between Z = 0 and Z = +0.5)

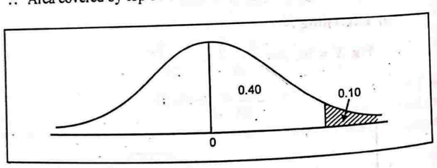
= 0.1915 + 0.1915 = 0.3830

No. of students getting score between 30 and 54

= 0.3830 × 1,000 = 383

(iii) Proportion of 100 top students $= \frac{100}{1000} = 0.10$

∴ Area covered by top 100 students = 0.10



The value of Z having 0.10 area to its right.

= Value of Z corresponding to (0.5 – 0.1) *i.e.* 0.40 area = 1.28 approx.

∴ $Z = \frac{X - \bar{X}}{\sigma}$

$1.28 = \frac{X-42}{24}$

$1.28 \times 24 = X - 42$

∴ $X = 1.28 \times 24 + 42 = 72.72 = 73$

Example 31. You are the incharge of the rationing department of a state efffected by food shortage. The following information is received from your local investigators:

Area	Mean calories	Standard deviation of calories
X	2500	500
Y	2200	300

The estimated requirement at an adult is taken at 3,000 calories daily and absolute minimum at 1,250. Comment on the reported figures and determine which area needs more urgent action.

Solution: In a population $\bar{X} \pm 3\sigma$ covers 99.73% almost all cases.

The limits on the basis of the information given to us should be

Area X : $\bar{X} \pm 3\sigma = 2500 \pm 3 \times 500$

= 1,000 to 4,000

Area Y : $\bar{Y} \pm 3\sigma = 2,200 \pm 3 \times 300$

= 1,300 to 3,100

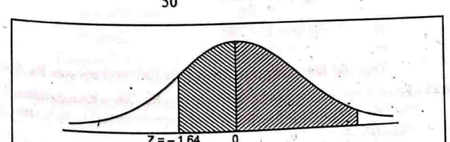
The absolute daily minimum calories requirement for a person is 1250. From the above figures we observe that almost all the persons in the area Y are getting more than the minimum calories requirement as the lower limit in this area is 1300. However, since in the area X, the lower 3-σ limit is 1000 which is less than 1250, quite a number of people in area X are not getting the minimum requirement of 1250 calories. Hence, as the incharge of the rationing department, it becomes my duty to take urgent action for the people of area X.

Example 32. The income of a group of 10,000 persons was found to be normally distributed with mean Rs. 750 per month and standard deviation Rs. 50. Show that of this group 95% had income exceeding Rs. 668 and only 5% had income exceeding Rs. 832. What was the lowest income among the richest 100?

Solution: Given, N = 10,000, $\bar{X} = 750$, $\sigma = 50$.

(i) **Exceeding Rs. 668**

For, X = 668, $Z = \frac{668-750}{50} = -1.64$



Required Proportion

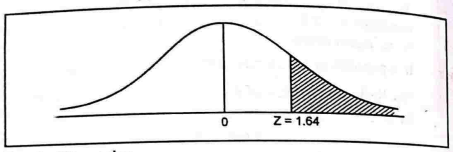
= Area between (Z = –1.64 and Z = 0) + Area to the right of (Z = 0).

= 0.4495 + 0.5000 = 0.9495

Hence, the required percentage = 0.9495 × 100 = 94.95% = 95%.

(ii) **Exceeding Rs. 832**

For X = 832, $Z = \dfrac{832 - 750}{50} = 1.64$



Required Proportion

= Area to the rignt of (Z = 0) – Area between (Z = 0 and Z = 1.64)

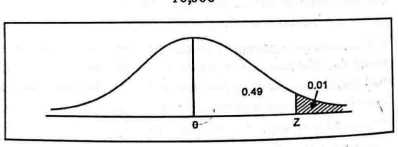
= 0.5000 – 0.4495

= 0.0505

Hence, the required percentage = 0.0505 × 100

= 5.05% = 5%

(iii) Proportion of the richest 100 = $\dfrac{100}{10{,}000} = 0.01$



Value of Z corresponding to 0.50 – 0.01 = 0.49 area = 2.33

∴ $2.33 = \dfrac{X - 750}{50}$

or $2.33 \times 50 = X - 750$

or $X = 866.50$

Thus, the lowest wage of the richest 100 workers was Rs. 866.50.

Example 33. The average daily sales of a shopkeeper is Rs. 200 with a standard deviation of Rs. 40. For how many days in a leap year his sales is expected to be worth *(i)* less than Rs. 150 and *(ii)* over Rs. 300?

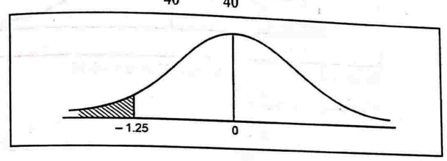
Solution: Given, $\bar{X} = 200$, σ = 40

Number of days in a leap year = N = 366

(i) **Sales less than Rs. 150:**

For X = 150, $Z = \dfrac{X - \bar{X}}{\sigma}$

$= \dfrac{150 - 200}{40} = \dfrac{-50}{40} = -1.25$



Required Proportion

= Area to the left of (Z = 0) – Area between (Z = – 1.25 and Z = 0)

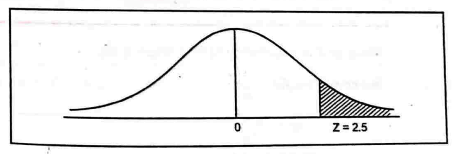
= 0.5000 – 0.3944 = 0.1056

∴ Expected days = 366 × 0.1056

= 38.64 = 39 days approx.

(ii) **Sales over Rs. 300:**

For X = 300, $Z = \dfrac{300 - 200}{40} = \dfrac{100}{40} = 2.5$



Required Proportion

= Area to the right of (Z = 0) – Area between (Z = 0 and Z = 2.5)

= 0.5000 – 0.4938

= 0.0062

∴ Expected days = 366 × 0.0062

= 2.2 = 2 days.

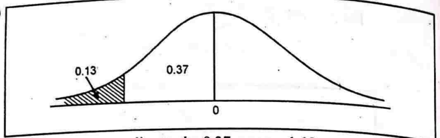
Example 34. Given a normal distribution with $\bar{X} = 50$ and σ = 10. Find the value of X that has

(i) 13% of the area of its left, and

(ii) 14% of the area to its right.

Solution: Given $\overline{X} = 50,\ \sigma = 10$

(*i*)



Value of Z corresponding to the 0.37 area = −1.13

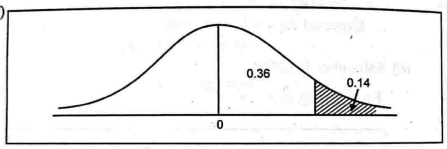
We know, $Z = \dfrac{X - \overline{X}}{\sigma}$

$$-1.13 = \frac{X - 50}{10}$$

$$-11.3 = X - 50$$

or $50 - 11.3 = X$ or $X = 38.7$

(*ii*)



Value of Z corresponding to 0.36 area = 1.08

We know that, $Z = \dfrac{X - \overline{X}}{\sigma}$

$$1.08 = \frac{X - 50}{10}$$

$$10.8 = X - 50$$

$$50 + 10.8 = X$$

$$\Rightarrow X = 60.8$$

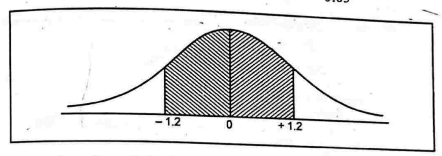
Example 35. The mean diameter of a sample of 500 washers produced by a machine is 5.02 mm with a standard deviation of 0.05 mm. The purpose for which the washers are manufactured allows a maximum tolerance in the diameter of 4.96 and 5.08 mm, otherwise the washers are considered defective. Determine the percentage of defective washers produced on the machine, assuming that the diameter variations are normally distributed.

Solution: Given $\overline{X} = 5.02,\ \sigma = 0.05,\ N = 500$

Between 4.96 and 5.08

$$\text{SNV}\ (Z_1) \text{ corresponding to } 4.96 = \frac{X_1 - \overline{X}}{\sigma} = \frac{4.96 - 5.02}{0.05} = -1.2$$

$$\text{SNV}\ (Z_2) \text{ corresponding to } 5.08 = \frac{X_2 - \overline{X}}{\sigma} = \frac{5.08 - 5.02}{0.05} = +1.2$$



The proportion of non-defective washers

= Area between (Z = −1.2 and Z = 0) + Area between (Z = 0 and Z = 1.2)

= 0.3849 + 0.3849 = 0.7698 or 77%

Hence, the percentage of the defective washers

= 100 − 77 = 23%

Example 36. The results of a given selection test exercise are summarised below:

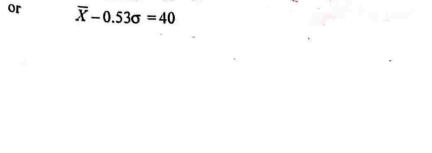
(*i*) Cleared with distinction = 10 percent

(*ii*) Cleared without distribution = 60 per cent

(*iii*) Those who failed = 30 per cent

It is known that a candidate fails if he/she obtains less than 40 per cent marks, while one must obtain at least 75 per cent marks in order to pass with distinction. Determine the mean and standard deviation of the distribution of marks assuming the same to be normal.

Solution:



Value of Z corresponding to 0.50 − 0.30

= 0.20 area = − 0.53 (From the table)

$$-0.53 = \frac{40 - \overline{X}}{\sigma}$$

or $-0.53\sigma = 40 - \overline{X}$

or $\overline{X} - 0.53\sigma = 40$ ...(*i*)

Value of Z corresponding to 0.50 − 0.10 = 0.40 area = 1.29 (From the table)

$$1.29 = \frac{75 - \bar{X}}{\sigma}$$

or $1.29\sigma = 75 - \bar{X}$

or $\bar{X} + 1.29\sigma = 75$ ...(ii)

Solving the two equations

$$\bar{X} - 0.53\sigma = 40$$
$$\bar{X} + 1.29\sigma = 75$$
$$- \quad - \quad -$$
$$-1.82\sigma = -35$$

$$\Rightarrow \quad \sigma = \frac{35}{1.82} = 19.23$$

Substituting the value of σ in equation (i)

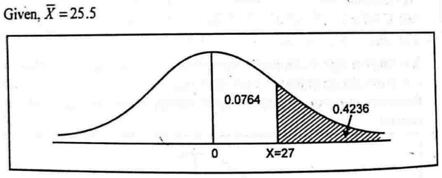
$$\bar{X} - 0.53(19.23) = 40$$

$\bar{X} - 10.19 = 40$ or $\bar{X} = 50.19$

$\therefore \quad \bar{X} = 50.19, \ \sigma = 19.23$

Example 37. A normal variate X has a mean of 25.5. It is known that 42.36 per cent of the X values are more than X = 27. Find the standard deviation of X.

Solution: Given, $\bar{X} = 25.5$



$$Z = \frac{X - \bar{X}}{\sigma}$$

Value of Z corresponding to (0.5000 − 0.4236) = 0.0764 area = + 0.19 (From the table)

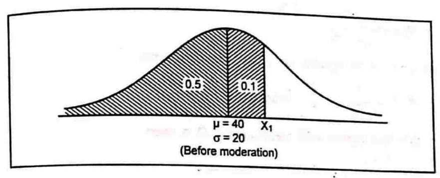
$$0.19 = \frac{27 - 25.5}{\sigma}$$

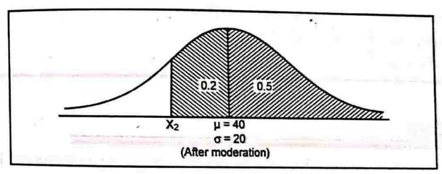
$$\Rightarrow \quad \sigma = \frac{1.5}{0.19} = 7.89 \approx 7.9$$

$$\therefore \quad \sigma(S.D.) = 7.9$$

Example 38. The marks of the students in a certain examination are normally distributed with mean marks as 40% and standard deviation marks as 20%. On this basis, 60% students failed. The result was moderated and 70% students passed. Find the pass marks before and after the moderation.

Solution: Let X_1% be the pass marks before moderation and X_2% be the pass marks after moderation.





Value of Z corresponding to area 0.1 from the mean = 0.253

Value of Z corresponding to area 0.2 from the mean = − 0.525

$$\therefore \quad \frac{X_1 - 40}{20} = 0.253 \quad \Rightarrow \quad X_1 = 45.06\% \text{ or } 45\%$$

$$\frac{X_2 - 40}{20} = -0.525 \quad \Rightarrow \quad X_2 = 29.5\%$$

Pass marks before moderation = 45%

and pass marks after moderation = 29.5%.

IMPORTANT FORMULAE

Normal Distribution:

(i) It is a continuous probability distribution.

(ii) The equation of the normal curve in general form:

$$P(X)=\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{X-\bar{X}}{\sigma}\right)^2}$$

(iii) The equation of the normal curve in its standard form:

$$P(Z)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{1}{2}Z^2} \quad \text{where, } Z=\frac{X-\bar{X}}{\sigma}$$

(iv) If X is a normal variate with mean μ and S.D. σ, then

$Z=\frac{X-\mu}{\sigma}$ is a standard normal variate.

(v) Area between $\bar{X}-1\sigma$ and $X+1\sigma=0.6827$

Area between $\bar{X}-2\sigma$ and $X+2\sigma=0.9545$

Area between $\bar{X}-3\sigma$ and $X+3\sigma=0.9974$

QUESTIONS

1. What is normal distribution? Explain its properties. Bring out its importance in statistics.
2. Describe normal distribution and discuss its properties. Why is it so important in statistics?
3. What is meant by theoretical frequency distribution? Discuss the salient features of Binomial, Poisson and Normal Probability Distribution.
4. How does a normal distribution differ from a binomial distribution? Mention the properties of normal distribution?
5. Discuss briefly the importance of normal distribution in statistical analysis.
6. Write short notes on any two of the following:
   (i) Assumptions to apply Binomial Distribution.
   (ii) Properties of Normal Distribution
   (iii) Importance of Poisson Distribution.
7. State the conditions under which Binomial distribution tends to normal distribution.
8. When Poisson distribution tends to be normal?

ADVANCED STATISTICS

1

Partial and Multiple Correlation and Regression

INTRODUCTION

Partial and multiple correlation and regression are extension of the technique of simple correlation and regression under which we study the interrelationship between three or more variables.

(1) Multiple Correlation

Multiple correlation is the study of the relationship among three or more variables. Multiple correlation measures the combined influence of two or more independent variables on a single dependent variable. For example, if we study the combined influence of amount of fertiliser (x_2) and rainfall (x_3) on the yield of wheat (x_1), then it is called the problem of multiple correlation. We shall denote the multiple correlation coefficient between x_1, the dependent variables x_2 and x_3 independent variables by $R_{1.23}$. Similarly, we shall denote the other multiple correlation coefficients by $R_{2.13}$ and $R_{3.12}$.

Calculation of Coefficient of Multiple Correlation: The formulae for calculating the multiple correlation coefficients $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$ are as follows :

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

Where, $R_{1.23}$ = Multiple correlation coefficient

r_{12}, r_{13}, r_{23} = Simple or zero order correlation coefficient.

$$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{21} \cdot r_{23} \cdot r_{13}}{1 - r_{13}^2}}$$

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{31} \cdot r_{32} \cdot r_{12}}{1 - r_{12}^2}}$$

Limits of Multiple Correlation Coefficients: The value of multiple correlation coefficient ($R_{1.23}$) lies between 0 and 1. It can never be negative.

$$0 \le R_{1.23} \le 1$$

The following examples illustrate the calculations of multiple correlation coefficients:

Example 1. Calculate $R_{1.23}$, $R_{3.12}$ and $R_{2.13}$ for the following data:

$r_{12} = 0.60,\ r_{13} = 0.70,\ r_{23} = 0.65$

Solution. Given, $r_{12} = 0.60, r_{13} = 0.70, r_{23} = 0.65$

(i) $$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.6)^2 + (0.7)^2 - 2\,(0.6)\,(0.7)\,(0.65)}{1-(0.65)^2}}$$

$$= \sqrt{\frac{0.36 + 0.49 - 0.546}{0.5775}}$$

$$= \sqrt{0.526} = 0.725$$

(ii) $$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{21} \cdot r_{23} \cdot r_{13}}{1 - r_{13}^2}}$$

$$= \sqrt{\frac{(0.6)^2 + (0.65)^2 - 2\,(0.6)\,(0.65)\,(0.7)}{1-(0.70)^2}}$$

$$= \sqrt{\frac{0.36 + 0.4225 - 0.546}{1 - 0.49}}$$

$$= \sqrt{0.4638} = 0.6809$$

(iii) $$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{31} \cdot r_{32} \cdot r_{12}}{1 - r_{12}^2}}$$

$$= \sqrt{\frac{(0.70)^2 + (0.65)^2 - 2\,(0.70)\,(0.65)\,(0.60)}{1-(0.60)^2}}$$

$$= \sqrt{\frac{0.49 + 0.4225 - 0.546}{1 - 0.36}} = \sqrt{0.5726} = 0.756.$$

Example 2. For a large group of students x_1 = Score in Economics, x_2 = Score in Maths, x_3 = Score in Statistics, $r_{12} = 0.69$, $r_{13} = 0.45$, $r_{23} = 0.58$. Determine the coefficient of multiple correlation $R_{3.12}$.

Solution. $$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{31} \cdot r_{32} \cdot r_{12}}{1 - r_{12}^2}}$$

$$= \sqrt{\frac{(0.45)^2 + (0.58)^2 - 2\,(0.45)\,(0.58)\,(0.69)}{1-(0.69)^2}}$$

$$= \sqrt{\frac{0.2025 + 0.3364 - 0.3601}{1 - 0.4761}} = \sqrt{0.3412} = 0.584$$

Example 3. The following zero order correlation coefficient are given:

$r_{12} = 0.98,\ r_{13} = 0.44$ and $r_{23} = 0.54$

Calculate multiple correlation coefficient treating the first variable as dependent and second and third variables as independent.

Solution. We have to calculate the multiple correlation coefficient treating first variable as dependent and second and third variable as independent *i.e.,* we have to find $R_{1.23}$.

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

Substituting the given values,

$$R_{1.23} = \sqrt{\frac{(0.98)^2 + (0.44)^2 - 2\,(0.98)\,(0.44)\,(0.54)}{1-(0.54)^2}}$$

$$= \sqrt{\frac{0.9604 + 0.1936 - 0.4657}{1 - 0.2916}} = \sqrt{\frac{.6883}{.7084}}$$

$$= \sqrt{0.9716} = 0.985$$

Example 4. If $R_{1.23} = 1$, prove that $R_{2.13} = 1$.

Solution. $$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

and $$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{21} \cdot r_{23} \cdot r_{13}}{1 - r_{23}^2}}$$

putting $R_{1.23} = 1$ and squaring both sides,

$$1 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}$$

$$\Rightarrow\quad r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23} = 1 - r_{23}^2$$

$$\Rightarrow\quad r_{12}^2 + r_{23}^2 - 2r_{12} \cdot r_{13} \cdot r_{23} = 1 - r_{13}^2$$

$$\Rightarrow\quad \frac{r_{12}^2 + r_{23}^2 - 2r_{12} \cdot r_{23} \cdot r_{23}}{1 - r_{13}^2} = 1$$

$$\Rightarrow\quad R_{2.13}^2 = 1 \ \text{ or } \ R_{2.13} = 1$$

Since, the coefficient of multiple correlation is considered non-negative.

(2) Partial Correlation

Partial Correlation is the simple correlation between two variables after eliminating the influence of the third variable from them. For example, if we measure the relationship between yield of wheat (x_1) and the amount of fertiliser (x_2), eliminating the effect of climate (x_3) from both (having the same climate), then it is called the problem of partial correlation. For three variables

$(x_1, x_2$ and $x_3)$, there are three partial correlation coefficients. They are denoted by $r_{12.3}, r_{13.2}$ and $r_{23.1}$. The partial correlation coefficient $r_{12.3}$ indicates the relationship between x_1 and x_2 when the effect of x_3 is eliminated from both.

Calculation of Partial Correlation Coefficients : The formulae for calculating the partial correlation coefficients $r_{12.3}, r_{13.2}$ and $r_{23.1}$ are as follows :

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

Where, $r_{12.3}$ = Partial correlation between x_1 and x_2

r_{12}, r_{13} and r_{23} = Simple or zero order correlation coefficient.

Similarly, we have

$$r_{23.1} = \frac{r_{23} - r_{21} \cdot r_{31}}{\sqrt{1-r_{21}^2}\sqrt{1-r_{31}^2}}$$

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{32}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{32}^2}}$$

Limits of Partial Correlation Coefficient : The value of $r_{12.3}$ lies betwen -1 and $+1$.

$$-1 \le r_{12.3} \le 1$$

The following examples illustrate the calculations of partial correlation coefficient.

Example 5. **Given that $r_{12} = 0.7$, $r_{13} = 0.61$, $r_{23} = 0.4$. Find the values of $r_{12.3}, r_{13.2}, r_{23.1}$.**

Solution.

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

Substituting the values,

$$r_{12.3} = \frac{0.7 - (0.61)(0.4)}{\sqrt{1-(0.61)^2}\sqrt{1-(0.4)^2}}$$

$$= \frac{0.456}{0.792 \times 0.916} = 0.629$$

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{32}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{32}^2}}$$

$$= \frac{0.61 - (0.7)(0.4)}{\sqrt{1-(0.7)^2}\sqrt{1-(0.4)^2}}$$

$$= \frac{0.61 - 0.28}{\sqrt{1-.49}\sqrt{1-.16}}$$

$$= \frac{0.33}{0.714 \times 0.916} = \frac{0.33}{0.654} = 0.505$$

$$r_{23.1} = \frac{r_{23} - r_{21} \cdot r_{31}}{\sqrt{1-r_{21}^2}\sqrt{1-r_{31}^2}}$$

$$= \frac{0.4 - (0.7)(0.61)}{\sqrt{1-(0.7)^2}\sqrt{1-(0.61)^2}}$$

$$= \frac{0.4 - 0.427}{\sqrt{1-(0.49)}\sqrt{1-0.3721}} = \frac{-0.027}{0.714 \times 0.792} = -0.048$$

Example 6. **On the basis of observations made on 30 cotton plants the total correlation of yield of cotton (x_1) the number of balls *i.e.*, seed vessels (x_2) and height (x_3) are found to be:**

$$r_{12} = 0.8,\ r_{13} = 0.65,\ r_{23} = 0.7$$

Compute the partial correlation between yield of cotton and number of balls, eliminating the effect of height.

Solution. We have to find the partial correlation between yield of cotton (x_1) and the number of balls (x_2), eliminating the effect of height (x_3) *i.e.*, we have to find $r_{12.3}$.

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

Substituting the given values,

$$r_{12.3} = \frac{0.8 - (0.65)(0.7)}{\sqrt{1-(.65)^2}\sqrt{1-(0.7)^2}}$$

$$= \frac{0.8 - 0.455}{\sqrt{1-0.4225}\sqrt{1-0.49}}$$

$$= \frac{0.345}{0.76 \times 0.714} = \frac{0.345}{0.543} = 0.635$$

Example 7. **For a large group of students x_1 = Score in theory, x_2 = Score in method, x_3 = Score in field work. The following results were found :**

$$r_{12} = 0.69,\ r_{13} = 0.45,\ r_{23} = 0.58$$

Determine the partial correlation coefficient between score in field work and score in theory keeping the score in method constant and interpret the result.

Solution. We have to find partial correlation coefficient between score in field (x_3) and score in theory (x_1) keeping the scores in method constant *i.e.*, we have to find $r_{31.2}$.

$$r_{31.2} = \frac{r_{31} - r_{32} \cdot r_{12}}{\sqrt{1-r_{32}^2}\sqrt{1-r_{12}^2}}$$

$$= \frac{0.45 - (0.58)(0.69)}{\sqrt{1-(0.58)^2}\sqrt{1-(0.69)^2}}$$

$$= \frac{0.45 - 0.4002}{\sqrt{1-0.3364}\sqrt{1-0.4761}}$$

$$= \frac{0.0498}{\sqrt{0.6636}\ \sqrt{0.5239}}$$

$$= \frac{0.0498}{0.81 \times 0.72} = \frac{0.0498}{0.5832} = 0.085$$

Thus, there is low degree of correlation between score in field work and score in theory.

Example 8. **Is it possible to have the following set of experimental data:**

$$r_{12} = 0.6,\ r_{23} = 0.8,\ r_{31} = -0.5.$$

Solution. In order to see whether there is inconsistency in the given data, we should calculate $r_{12.3}$. If the value of $r_{12.3}$ exceeds one, there is inconsistency, otherwise not.

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2}\ \sqrt{1 - r_{23}^2}}$$

Substituting the given values,

$$= \frac{0.6 - (-0.5)(0.8)}{\sqrt{1 - (0.5)^2}\ \sqrt{1 - (0.8)^2}}$$

$$= \frac{0.6 + 0.4}{\sqrt{1 - 0.25}\ \sqrt{1 - 0.64}}$$

$$= \frac{1}{\sqrt{0.75}\ \sqrt{0.36}} = \frac{1}{0.866 \times 0.6} = \frac{1}{0.52} = 1.92$$

Since, the value of $r_{12.3}$ is greater than one, there is some inconsistency in the given data.

Aliter : We can also check the inconsistency in the data by calculating $R_{1.23}$. If the value of $R_{1.23}$ exceeds 1, there is some inconsistency otherwise not

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.6)^2 + (-0.5)^2 - 2(0.6)(-0.5)(0.8)}{1 - (0.8)^2}}$$

$$= \sqrt{\frac{0.36 + 0.25 + 0.48}{1 - .64}} = \sqrt{\frac{1.09}{0.36}} = \sqrt{3.0277} = 1.74$$

Since, the value of $R_{1.23}$ is greater than one, there is some inconsistency in the data.

Example 9. **Suppose a computer has found, for a given set of values of x_1, x_2 and x_3: $r_{12} = 0.96$, $r_{13} = 0.36$ and $r_{23} = 0.78$.**
Explain whether these computations may be said to be free from errors.

Solution. For determining whether the given computed values are free from errors or not, we compute the value of $r_{12.3}$. If $r_{12.3}$ comes out to be greater than one, the computed values cannot be regarded as free from errors.

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2}\ \sqrt{1 - r_{23}^2}}$$

Substituting the given values,

$$r_{12.3} = \frac{0.96 - (0.36)(0.78)}{\sqrt{1 - (0.36)^2}\ \sqrt{1 - (0.78)^2}}$$

$$= \frac{0.96 - 0.2808}{\sqrt{0.8704}\ \sqrt{0.3916}} = \frac{0.6792}{0.9329 \times 0.6258} = \frac{0.6792}{0.5838} = 1.163$$

Since, $r_{12.3}$ is greater than one, the given computed values do contain some errors.

Relationship between Simple, Partial and Multiple Correlation Coefficients

There exists relationship between simple, partial and multiple correlation coefficients which is clear from the following equation:

(i) $1 - R_{1.23}^2 = (1 - r_{12}^2)(1 - r_{13.2}^2)$

(ii) $1 - R_{2.13}^2 = (1 - r_{21}^2)(1 - r_{23.1}^2)$ and

(iii) $1 - R_{3.12}^2 = (1 - r_{31}^2)(1 - r_{32.1}^2)$

Example 10. **In a trivariate distribution, $r_{12} = 0.60$, $r_{13} = 0.70$, $r_{23} = 0.65$, find $R_{1.23}^2$ from r_{12} and $r_{13.2}$.**

Solution. Given : $r_{12} = 0.60, r_{13} = 0.70, r_{23} = 0.65$

Multiple, Simple and Partial Correlation coefficients are related as :

$$R_{1.23}^2 = 1 - (1 - r_{12}^2)(1 - r_{13.2}^2)$$

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{32}}{\sqrt{1 - r_{12}^2}\ \sqrt{1 - r_{32}^2}} = \frac{0.70 - 0.60 \times 0.65}{\sqrt{1 - (0.60)^2}\ \sqrt{1 - (0.65)^2}}$$

$$= \frac{0.70 - 0.39}{0.8 \times 0.760} = 0.509$$

$$\therefore \quad r_{13.2}^2 = 0.259,\ r_{12}^2 = 0.36$$

Substituting values r_{12}^2 and $r_{13.2}^2$ for $R_{1.23}^2$, we have

$$R_{1.23}^2 = 1 - (1 - 0.36)(1 - 0.259)$$

$$= 1\ (0.64)(0.74) = 0.526.$$

MISCELLANEOUS SOLVED EXAMPLES

Example 11. **x_1, x_2 and x_3 are measured from their means with :**

$$N = 10,\ \Sigma x_1^2 = 90,\ \Sigma x_2^2 = 160,\ \Sigma x_3^2 = 40$$

$$\Sigma x_1 x_2 = 60,\ \Sigma x_2 x_3 = 60,\ \Sigma x_3 x_1 = 40$$

Calculate $r_{12.3}$ and $R_{2.31}$.

Solution.

$$r_{12} = \frac{\Sigma x_1 x_2}{\sqrt{\Sigma x_1^2}\ \sqrt{\Sigma x_2^2}} = \frac{60}{\sqrt{90 \times 160}} = \frac{60}{120} = 0.5$$

$$r_{13} = \frac{\Sigma x_1 x_3}{\sqrt{\Sigma x_1^2 \times \Sigma x_3^2}} = \frac{40}{\sqrt{90 \times 40}} = \frac{40}{60} = 0.67$$

$$r_{23} = \frac{\Sigma x_2 x_3}{\sqrt{\Sigma x_2^2 \times \Sigma x_3^2}} = \frac{60}{\sqrt{160 \times 40}} = \frac{60}{80} = 0.75$$

Now, $$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

Substituting the values, we have

$$r_{12.3} = \frac{0.5 - 0.67 \times 0.75}{\sqrt{1-(0.67)^2}\sqrt{1-(0.75)^2}} = -\frac{0.0025}{0.4910} = -0.0051$$

$$R_{2.31} = \sqrt{\frac{r_{23}^2 + r_{21}^2 - 2r_{23} \cdot r_{21} \cdot r_{31}}{1 - r_{31}^2}}$$

$$= \sqrt{\frac{(0.75)^2 + (0.5)^2 - 2\,(0.75)\,(0.5)\,(0.67)}{1-(.67)^2}}$$

$$= \sqrt{\frac{0.5625 + 0.25 - 0.5025}{0.5511}} = \sqrt{\frac{0.31}{0.5511}} = 0.75$$

Example 12. **Calculate $r_{12.3}$ and $R_{1.23}$ from the following data :**

X :	3	4	5	6	7	8	9
Y :	2	5	6	4	3	2	4
Z :	5	6	4	5	6	5	8

Solution. Calculation of $r_{12.3}$ and $R_{1.23}$

X	X^2	Y	Y^2	Z	Z^2	XY	XZ	YZ
3	9	2	4	5	25	6	15	10
4	16	5	25	6	36	20	24	30
5	25	6	36	4	16	30	20	24
6	36	4	16	5	25	24	30	20
7	49	3	9	6	36	21	42	18
8	64	2	4	5	25	16	40	10
9	81	4	16	8	64	36	72	32
$N = 7$, $\Sigma X = 42$	$\Sigma X^2 = 280$	$\Sigma Y = 26$	$\Sigma Y^2 = 110$	$\Sigma Z = 39$	$\Sigma Z^2 = 227$	$\Sigma XY = 153$	$\Sigma XZ = 243$	$\Sigma YZ = 144$

$$r_{12} = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{[\Sigma X^2.N - (\Sigma X)^2][\Sigma Y^2.N - (\Sigma Y)^2]}} = \frac{7 \times 153 - (42 \times 26)}{\sqrt{[280 \times 7 - (42)^2][110 \times 7 - (26)^2]}} = -0.155$$

$$r_{13} = \frac{N.\Sigma XZ - \Sigma X.\Sigma Z}{\sqrt{[\Sigma X^2.N - (\Sigma X)^2][\Sigma Z^2.N - (\Sigma Z)^2]}} = \frac{7 \times 243 - (42 \times 39)}{\sqrt{[280 \times 7 - (42)^2][227 \times 7 - (39)^2]}} = 0.546$$

$$r_{23} = \frac{N.\Sigma YZ - \Sigma Y.\Sigma Z}{\sqrt{[\Sigma Y^2.N - (\Sigma Y)^2][\Sigma Z^2.N - (\Sigma Z)^2]}} = \frac{144 \times 7 - 26 \times 39}{\sqrt{[110 \times 7 - (26)^2][227 \times 7 - (39)^2]}} = -0.075$$

Partial Correlation Coefficient

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2} \cdot \sqrt{1 - r_{23}^2}}$$

$$= \frac{-0.155 - (0.546 \times -0.075)}{\sqrt{1-(0.546)^2}\sqrt{1-(-0.075)^2}} = -0.1366$$

Multiple Correlation Coefficient

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(-0.155)^2 + (0.546)^2 - (2 \times -0.155 \times 0.546 \times -0.075)}{1-(-0.075)^2}}$$

$$= \sqrt{\frac{0.024 + 0.298 - (.01269)}{1 - .006}}$$

$$= \sqrt{\frac{0.1951}{0.994}}$$

$$= \sqrt{0.1962} = 0.443$$

Example 12 A. **In a trivariate distribution, $r_{12} = 0.80$, $r_{23} = -0.56$, $r_{31} = -0.40$, compute $r_{23.1}$ and $R_{1.23}$.**

Solution. (i)

$$r_{23.1} = \frac{r_{23} - r_{21} \cdot r_{31}}{\sqrt{1 - r_{21}^2}\sqrt{1 - r_{31}^2}}$$

$$= \frac{-0.56 - (0.8) \cdot (-0.40)}{\sqrt{1-(0.8)^2}\sqrt{1-(-0.4)^2}}$$

$$= \frac{-0.56 + 0.32}{\sqrt{1 - 0.64}\sqrt{1 - 0.16}}$$

$$= \frac{-0.24}{\sqrt{0.36 \times 0.84}} = \frac{-0.24}{05499} = -0.436$$

(ii)

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.8)^2 + (-0.4)^2 - 2\,(0.8)\,(-0.4)\,(-0.56)}{1-(-0.56)^2}}$$

$$= \sqrt{\frac{0.64 + 0.16 - 0.3584}{1 - 0.3136}} = \sqrt{\frac{0.4416}{0.6864}} = 0.802$$

Example 13. **The linear correlation coefficient between x_1 (Yield), x_2 (Irrigation) and x_3 (Fertiliser) are as follows :**

$$r_{12} = 0.81,\ r_{13} = 0.90,\ r_{23} = 0.65$$

Calculate the partial correlation coefficient of:

(i) yield with irrigation

(ii) yield with fertiliser.

Solution. (i) We have to find $r_{12.3}$

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\ \sqrt{1-r_{23}^2}}$$

Substituting the given values,

$$r_{12.3} = \frac{(0.81)-(0.90)\ (0.65)}{\sqrt{1-(0.90)^2}\ \sqrt{1-(0.65)^2}}$$

$$= \frac{0.81-0.585}{0.4358 \times 0.7599} = \frac{0.225}{0.3311} = 0.679$$

(ii) We have to find $r_{13.2}$

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{32}}{\sqrt{1-r_{12}^2}\ \sqrt{1-r_{32}^2}}$$

$$= \frac{(0.90)-(0.81)\ (0.65)}{\sqrt{1-(0.81)^2}\ \sqrt{1-(0.65)^2}}$$

$$= \frac{0.3735}{\sqrt{0.3439}\ \sqrt{0.5775}}$$

$$= \frac{0.3735}{0.5864 \times 0.7599} = \frac{0.3735}{0.4456} = 0.838$$

Example 14. **Given the following zero order correlation coefficient, find (i) partial correlation coefficient between x_2 and x_3 and (ii) multiple correlation taking x_1 as dependent on x_2 and x_3.**

$$r_{12} = 0.98,\ r_{13} = 0.44,\ r_{23} = 0.54$$

Solution. (i)

$$r_{23.1} = \frac{r_{23} - r_{21} \cdot r_{31}}{\sqrt{1-r_{21}^2}\ \sqrt{1-r_{31}^2}}$$

$$= \frac{0.54-(0.98)\ (0.44)}{\sqrt{1-(0.98)^2}\ \sqrt{1-(0.44)^2}}$$

$$= \frac{0.54-0.4312}{\sqrt{1-0.9604}\ \sqrt{1-0.1936}}$$

$$= \frac{0.1088}{\sqrt{0.0396}\ \sqrt{0.8064}} = \frac{0.1088}{0.1786} = 0.6091$$

(ii)

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1-r_{23}^2}}$$

$$= \sqrt{\frac{(0.98)^2 + (0.44)^2 - 2(0.98)\ (0.44)\ (0.54)}{1-(0.54)^2}}$$

$$= \sqrt{\frac{0.9604+0.1936-0.4656}{1-(0.2916)}} = \sqrt{\frac{0.6884}{0.7084}}$$

$$= \sqrt{0.9717} = 0.985.$$

Example 15. **Is it possible to get the following from a set of experimental data:**

(i) $r_{23} = 0.8,\ r_{31} = 0.5,\ r_{12} = 0.6$

(ii) $r_{23} = 0.7,\ r_{31} = -0.4,\ r_{12} = 0.6$

Solution. (i) In order to see whether there is any inconsistency, we should calculate $r_{12.3}$, If its value exceed one, there is inconsistency, otherwise not.

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\ \sqrt{1-r_{23}^2}}$$

$$= \frac{0.6-(0.5)\ (0.8)}{\sqrt{1-(0.5)^2}\ \sqrt{1-(0.8)^2}}$$

$$= \frac{0.20}{\sqrt{0.75}\ \sqrt{0.36}} = \frac{0.20}{0.52} = 0.384$$

Since, the value of $r_{12.3}$ is less than one, the data is consistent.

(ii)

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\ \sqrt{1-r_{23}^2}} = \frac{(0.6)-(-0.4)\ (0.7)}{\sqrt{1-(-0.4)^2}\ \sqrt{1-(0.7)^2}}$$

$$= \frac{0.6+0.28}{\sqrt{0.84}\ \sqrt{0.51}} = \frac{0.88}{0.655} = 1.344$$

Since, the value or $r_{12.3}$ is greater than 1 there is some inconsistency in the given data.

Example 16. **Test the consistency of the following data : $r_{12} = 0.8,\ r_{13} = 0.4,\ r_{23} = -0.56.$**

Solution. For testing whether the given computations are consistent or not, we compute the value of $r_{13.2}$. If $r_{13.2}$ comes out to be greater than one, the computations cannot regarded as consistent.

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\ \sqrt{1-r_{23}^2}}$$

$$= \frac{(0.8)-(0.4)\ (-0.56)}{\sqrt{1-(0.4)^2}\ \sqrt{1-(-0.56)^2}}$$

$$= \frac{0.8+0.224}{\sqrt{0.84}\ \sqrt{0.6864}} = \frac{1.024}{0.7593} = 1.349$$

Since, $r_{12.3}$ is greater than one, the given computations of r_{12}, r_{13} and r_{23} are not consistent.

Example 17. **If $r_{12} = 0.77$, $r_{13} = 0.72$ and $r_{23} = 0.52$, find the partial correlation coefficient $r_{12.3}$ and multiple correlation coefficient $R_{1.23}$.**

Solution. Given : $r_{12} = 0.77$, $r_{13} = 0.72$ and $r_{23} = 0.52$

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}}$$

$$= \frac{.77 - .72 \times .52}{\sqrt{1-(.72)^2}\sqrt{1-(.52)^2}}$$

$$= \frac{.77 - .37}{\sqrt{1-.5184}\sqrt{1-.2704}}$$

$$= \frac{.40}{\sqrt{.4816} \times \sqrt{.7296}} = \frac{.4}{.593} = 0.6745$$

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1-r_{23}^2}}$$

$$= \sqrt{\frac{(.77)^2 + (.72)^2 - 2(.77)(.72)(.52)}{1-(.52)^2}}$$

$$= \sqrt{\frac{.5929 + .5184 - .5766}{1-.2704}} = \sqrt{\frac{.5347}{.7296}} = 0.856.$$

Example 18. **If $r_{12} = 0.60$, $r_{13} = 0.70$, $r_{23} = 0.65$, find partial correlation between x_1 and x_2 and multiple correlation between x_1 dependent on x_2 and x_3.**

Solution. Given : $r_{12} = 0.60$, $r_{13} = 0.70$, $r_{23} = 0.65$

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

$$= \frac{0.6 - 0.7 \times 0.65}{\sqrt{1-(0.7)^2}\sqrt{1-(0.65)^2}}$$

$$= \frac{0.6 - 0.455}{\sqrt{0.51 \times 0.5775}} = \frac{0.145}{0.543} = 0.2670$$

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1-r_{23}^2}}$$

$$= \sqrt{\frac{(.6)^2 + (.7)^2 - 2(.6)(.7)(.65)}{1-(.65)^2}} = \sqrt{\frac{.304}{.5775}} = 0.726.$$

Example 19. **Following table shows the correlation matrix of three variables x_1 (Height), x_2 (Weight) and x_3 (Diameter of Chest) of 10 randomly selected players :**

	x_1	x_2	x_3
x_1	1.0000	0.8630	0.6480
x_2		1.0000	0.7090
x_3			1.0000

Calculate $r_{12.3}$ and $R_{1.23}$.

Solution. Given : $r_{12} = 0.863$, $r_{13} = 0.648$, $r_{23} = 0.709$

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

$$= \frac{.863 - .648 \times .709}{\sqrt{1-(.648)^2}\sqrt{1-(.709)^2}}$$

$$= \frac{.863 - .4594}{\sqrt{.580} \times \sqrt{.497}} = \frac{.4036}{\sqrt{.2883}} = \frac{.4036}{.537} = 0.752$$

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1-r_{23}^2}}$$

$$= \sqrt{\frac{(.863)^2 + (.648)^2 - 2(.863)(.648)(.709)}{1-(.709)^2}}$$

$$= \sqrt{\frac{.745 + .42 - .793}{.497}} = 0.865$$

EXERCISE – 1

1. In a trivariate distribution, it is found that
$$r_{12} = 0.41, r_{13} = 0.71, r_{23} = 0.5$$
Find the value of $r_{23.1}$ and $r_{13.2}$. [Ans. $r_{23.1} = 0.325, r_{13.2} = 0.639$]
2. If $r_{12} = 0.7$, $r_{13} = 0.61$ and $r_{23} = 0.4$, find the value of $r_{12.3}, r_{13.2}$ and $r_{23.1}$.
[Ans. $r_{12.3} = 0.629, r_{13.2} = 0.505, r_{23.1} = -0.048$]
3. Is it possible to have the following experimental data:
$$r_{12} = 0.6, r_{23} = 0.8, r_{31} = -0.5$$
[Ans. $r_{12.3} = 1.92$, Inconsistency]
4. In a trivariate distribution, $r_{23} = .2, r_{13} = .5, r_{12} = .6$. Compute $r_{12.3}$ and $R_{1.23}$.
[Ans. $r_{12.3} = 0.47, R_{1.23} = 0.714$]
5. Suppose a computer has found for a given set of values of $x_1, x_2, x_3 : r_{12} = 0.91, r_{13} = 0.33$ and $r_{23} = 0.81$. Explain whether these computations may be said to be free from errors.
[Ans. $r_{12.3} = 1.161$; Not free from errors]
6. The following zero order correlation coefficients are given:
$$r_{12} = 0.98, r_{13} = 0.44, r_{23} = 0.54$$

Calculate :

(i) the partial correlation coefficient between first (x_1) and third (x_3) variables; and

(ii) multiple correlation coefficient treating first variable (x_1) as dependent and second and third variable as independent. **[Ans.** $r_{13.2} = -0.53$, $R_{1.23} = 0.986$**]**

7. If $r_{12} = 0.9, r_{13} = 0.75, r_{23} = 0.7$, find the $R_{1.23}$. **[Ans.** $R_{1.23} = 0.916$**]**

8. Test the consistency of the data ::

$$r_{12} = 0.6, r_{13} = 0.5 \text{ and } r_{23} = 0.2.$$

Compute $r_{12.3}$, and $R_{1.23}$. **[Ans.** $r_{12.3} = 0.589$, Consistent**]**

9. Given the following values:

$$r_{12} = 0.6, r_{23} = r_{31} = 0.8,$$

find : $r_{23.1}$ and $R_{1.23}$. **[Ans.** $r_{23.1} = 0.667$, $R_{1.23} = 0.803$**]**

10. For a large group of students, x_1 = Score in Economics, x_2 = Score in Maths, x_3 = Score in Statistics, $r_{12} = 0.69, r_{13} = 0.45, r_{23} = 0.58$. Determine the coefficient of multiple correlation $R_{3.12}$. **[Ans.** $R_{3.12} = 0.98$**]**

11. The simple correlation coefficient between temperature (x_1), corp yield (x_2) and rainfall (x_3) are :

$r_{12} = 0.59, r_{13} = 0.46$ and $r_{23} = 0.77$. Calculate $r_{12.3}$ and $R_{1.23}$.

[Ans. $r_{12.3} = 0.416$, $R_{1.23} = .588$**]**

12. x_1, x_2 and x_3 are measured from their means with:

$N = 6, \Sigma x_1^2 = 90, \Sigma x_2^2 = 140, \Sigma x_3^2 = 4008$

$\Sigma x_1 x_2 = -100, \Sigma x_1 x_3 = -582, \Sigma x_2 x_3 = 720$

Calculate $r_{12.3}$ and $R_{1.23}$

[Ans. $r_{12} = -0.891, r_{13} = -.969, r_{23} = 0.961, r_{12.3} = 0.605, R_{1.23} = 0.97$**]**

(3) MULTIPLE REGRESSION

In multiple regression, we study three variables and we consider one variable as dependent variable and the other two as independent variables. Multiple regression analysis is used to estimate the most probable value of the dependent variable for given values of the independent variables.

Methods to obtain Multiple Regression Equations

Multiple regression equations can be worked out by two methods, which are as follows:

(1) Multiple Regression Equations using Normal Equations

(2) Multiple Regression Eqautions in terms of Simple Correlation Coefficients

Let us discuss them.

(1) Multiple Regression Equations using Normal Equations : This method is also called as **Least Square Method.** Under this method computation of regression equations is done by solving three normal equations. This method becomes clear by the following :

Multiple regression equation of X_1 on X_2 and X_3 is given by :

$$X_1 = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$$

Where, X_1 = Dependent variable, X_2 and X_3 = Independent variables.

$b_{12.3}$ and $b_{13.2}$ = Partial regression coefficients.

Using least square method, the values of constants $a_{1.23}$, $b_{12.3}$ and $b_{13.2}$ are obtained by solving the following three normal equations:

$$\Sigma X_1 = N . a_{1.23} + b_{12.3} \Sigma X_2 + b_{13.2} \Sigma X_3 \quad ...(1)$$

$$\Sigma X_1 X_2 = a_{1.23} \Sigma X_2 + b_{12.3} \Sigma X_2^2 + b_{13.2} \Sigma X_2 X_3 \quad ...(2)$$

$$\Sigma X_1 X_3 = a_{1.23} \Sigma X_3 + b_{12.3} \Sigma X_2 X_3 + b_{13.2} \Sigma X_3^2 \quad ...(3)$$

Similarly, the multiple regression equations of X_2 on X_1 and X_3 and X_3 ; on X_1 and X_2 and their normal equations can also be written as:

Multiple Regression Equation of X_2 on X_1 and X_3 is given by:

$$X_2 = a_{2.13} + b_{21.3} X_1 + b_{23.1} X_3$$

Three Normal Equations are :

$$\Sigma X_2 = N a_{2.13} + b_{21.3} \Sigma X_1 + b_{23.1} \Sigma X_3$$

$$\Sigma X_2 X_1 = a_{2.13} \Sigma X_1 + b_{21.3} \Sigma X_1^2 + b_{23.1} \Sigma X_3 X_1$$

$$\Sigma X_2 X_3 = a_{2.13} \Sigma X_3 + b_{21.3} \Sigma X_1 X_3 + b_{23.1} \Sigma X_3^2$$

Multiple Regression Equation of X_3 on X_1 and X_2 is given by:

$$X_3 = a_{3.12} + b_{31.2} X_1 + b_{32.1} X_2$$

Three Normal Equations are:

$$\Sigma X_3 = N a_{3.12} + b_{31.2} \Sigma X_1 + b_{32.1} \Sigma X_2$$

$$\Sigma X_3 X_1 = a_{3.12} \Sigma X_1 + b_{31.2} \Sigma X_1^2 + b_{32.1} \Sigma X_2 X_1$$

$$\Sigma X_3 X_2 = a_{3.12} \Sigma X_2 + b_{31.2} \Sigma X_1 X_2 + b_{32.1} \Sigma X_2^2$$

The following example illustrate the procedure of fitting multiple regression equations:

Example 1. **For the following set of data, calculate multiple regression equation of X_1 on X_2 and X_3 :**

X_1 :	4	6	7	9	13	15
X_2 :	15	12	8	6	4	3
X_3 :	30	24	20	14	10	4

Solution. The regression equation of X_1 on X_2 and X_3 is

$$X_1 = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$$

The three normal equations are :

$$\Sigma X_1 = N a_{1.23} + b_{12.3} \Sigma X_2 + b_{13.2} \Sigma X_3$$

$$\Sigma X_1 X_2 = a_{1.23} \Sigma X_2 + b_{12.3} \Sigma X_2^2 + b_{13.2} \Sigma X_3 X_2$$

$$\Sigma X_1 X_3 = a_{1.23} \Sigma X_3 + b_{12.3} \Sigma X_2 X_3 + b_{13.2} \Sigma X_3^2$$

X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	X_2^2	X_3^2
4	15	30	60	120	450	225	900
6	12	24	72	144	288	144	576
7	8	20	56	140	160	64	400
9	6	14	54	126	84	36	196
13	4	10	52	130	40	16	100
15	3	4	45	60	12	9	16
$\Sigma X_1 = 54$	$\Sigma X_2 = 48$	$\Sigma X_3 = 102$	$\Sigma X_1X_2 = 339$	$\Sigma X_1X_3 = 720$	$\Sigma X_2X_3 = 1034$	$\Sigma X_2^2 = 494$	$\Sigma X_3^2 = 2188$

Substituting the values in the normal equations :

$$54 = 6a_{1.23} + 48b_{12.3} + 102b_{13.2} \quad \text{...(i)}$$
$$339 = 48a_{1.23} + 494b_{12.3} + 1034b_{13.2} \quad \text{...(ii)}$$
$$720 = 102a_{1.33} + 1034b_{12.3} + 2188b_{13.2} \quad \text{...(iii)}$$

Multiplying (i) by 8, we get

$$432 = 48a_{1.23} + 384b_{12.3} + 816b_{13.2} \quad \text{...(iv)}$$

Subtracting (ii) from (iv), we get

$$-93 = 110b_{12.3} + 218b_{13.2} \quad \text{...(v)}$$

Multiplying (i) by 17, we get

$$918 = 102a_{1.23} + 816b_{12.3} + 1734b_{13.2} \quad \text{...(vi)}$$

Subtracting (iii) from (vi), we get

$$-198 = 218b_{12.3} + 454b_{13.2} \quad \text{...(vii)}$$

Multiplying (v) by 109, we obtain

$$-10137 = 11990b_{12.3} + 23762b_{13.2} \quad \text{...(viii)}$$

Multiplying (vii) by 55, we get

$$-10890 = 11990b_{12.3} + 24970b_{13.2} \quad \text{...(ix)}$$

Subtracting (viii) from (ix), we get

$$753 = -1208b_{13.2}$$

$$\therefore \quad b_{13.2} = -\frac{753}{1208} = -0.623$$

Substituting the value of $b_{13.2}$ in equation (v), we get

$$-93 = 110b_{12.3} + 218(-0.623)$$
$$135.814 - 93 = 110b_{12.3}$$
$$\therefore \quad b_{12.3} = \frac{42.814}{110} = 0.389$$

Substituting the values of $b_{12.3}$ and $b_{13.2}$ in equation (i), we get

$$6a_{1.23} + 48(0.389) + 102(-0.623) = 54$$
$$6a_{1.23} + 18.672 - 63.546 = 54$$
$$6a_{1.23} = 54 - 18.672 + 63.546$$
$$6a_{1.23} = 98.874$$
$$\therefore \quad a_{1.23} = \frac{98.874}{6} = 16.479$$

Hence, the required equation is

$$X_1 = 16.479 + 0.389X_2 - 0.623X_3$$

EXERCISE – 2

1. From the following data, find the least square regression of X_1, X_2 and X_3 and estimate the value of X_1 for given values of $X_2 = 16$ and $X_3 = 4$:

X_1 :	10	5	10	4	8
X_2 :	16	13	21	10	13
X_3 :	3	6	4	5	3

[Ans. $X_1 = 4.753 + 0.502X_2 - 1.115X_3$, 8.325]

2. Compute the values of b_0, b_1 and b_2 for the equation $Y = b_0 + b_1X_1 + b_2X_2$ from the following data:

Y :	3	5	6	8	12	14
X_1 :	16	10	7	4	3	2
X_2 :	90	72	54	42	30	12

[Ans. $Y = 16.1067 + .426X_1 - 0.221X_2$]

3. Obtain the parameters of the multiple linear regression model: $Y = \beta_1 + \beta_2X_2 + \beta_3X_3$ from the following data:

$$N = 6,\ \Sigma Y = 54,\ \Sigma X_2 = 48,\ \Sigma X_3 = 102$$
$$\Sigma YX_2 = 339,\ \Sigma YX_3 = 720,\ \Sigma X_2X_3 = 1034,\ \Sigma X_2^2 = 494,\ \Sigma X_3^2 = 2188$$

[Ans. $Y = 16.479 + 0.389X_2 - 0.623X_3$]

Short-Cut Method : When the size of the values of the variables are very large, then the above system of solving normal equations becomes a very tedious procedure. In such a case, in place of actual values, deviations from the means of the variables are used to simplify the computation procedure.

Multiple Regression Eqaution of X_1 on X_2 and X_3 in deviation form is given by :

$$X_1 - \overline{X}_1 = b_{12.3}(X_2 - \overline{X}_2) + b_{13.2}(X_3 - \overline{X}_3)$$

or $\quad x_1 = b_{12.3}x_2 + b_{13.2}x_3 \quad$ where $x_1 = X_1 - \overline{X}_1,\ x_2 = X_2 - \overline{X}_2,\ x_3 = X_3 - \overline{X}_3$

The values of the partial regression coefficients ($b_{12.3}$ and $b_{13.2}$) can be obtained by solving the following two normal equations :

$$\Sigma x_1x_2 = b_{12.3}\Sigma x_2^2 + b_{13.2}\Sigma x_3x_2$$
$$\Sigma x_1x_3 = b_{12.3}\Sigma x_2x_3 + b_{13.2}\Sigma x_3^2$$

Further solved, we have

$$b_{12.3} = \frac{(\Sigma x_1x_2)(\Sigma x_3^2) - (\Sigma x_1x_3)(\Sigma x_2x_3)}{(\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2x_3)^2}$$

$$b_{13.2} = \frac{(\Sigma x_1x_3)(\Sigma x_2^2) - (\Sigma x_1x_2)(\Sigma x_3x_2)}{(\Sigma x_3^2)(\Sigma x_2^2) - (\Sigma x_3x_2)^2}$$

Similarly, the multiple regression equation of X_2 on X_1 and X_3 ; and X_3 on X_1 and X_2 and their normal equations can also be written.

Multiple Regression Equation of X_2 on X_1 and X_3 in deviation form is given by:

$$X_2 - \overline{X}_2 = b_{21.3}(X_1 - \overline{X}_1) + b_{23.1}(X_3 - \overline{X}_3)$$

or

$$x_2 = b_{21.3}x_1 + b_{23.1}x_3$$

Two normal equations are :

$$\Sigma x_2x_1 = b_{21.3}\Sigma x_1^2 + b_{23.1}\Sigma x_1x_3$$
$$\Sigma x_2x_3 = b_{21.3}\Sigma x_1x_3 + b_{23.1}\Sigma x_3^2$$

Further solved, we have

$$b_{21.3} = \frac{(\Sigma x_2x_1)(\Sigma x_3^2) - (\Sigma x_2x_3)(\Sigma x_1x_3)}{(\Sigma x_1^2)(\Sigma x_3^2) - (\Sigma x_1x_3)^2}$$

$$b_{23.1}=\frac{(\Sigma x_2x_3)(\Sigma x_1^2)-(\Sigma x_2x_1)(\Sigma x_3x_1)}{(\Sigma x_3^2)(\Sigma x_1^2)-(\Sigma x_3x_1)^2}$$

Multiple Regression Equation of X_3 on X_1 and X_2 in deviation form is given by:

$$X_3-\overline{X}_3=b_{31.2}(X_1-\overline{X}_1)+b_{32.1}(X_2-\overline{X}_2)$$

$$x_3=b_{31.2}\,x_1+b_{32.1}x_2$$

or

Two normal equations are :

$$\Sigma x_1x_3=b_{31.2}\,\Sigma x_1^2+b_{23.1}\,\Sigma x_1x_2$$

$$\Sigma x_2x_3=b_{31.2}\,\Sigma x_1x_2+b_{32.1}\,\Sigma x_2^2$$

Further solved, we have

$$b_{31.2}=\frac{(\Sigma x_3x_1)(\Sigma x_2^2)-(\Sigma x_3x_2)(\Sigma x_1x_2)}{(\Sigma x_1^2)(\Sigma x_2^2)-(\Sigma x_1x_2)^2}$$

$$b_{32.1}=\frac{(\Sigma x_3x_2)(\Sigma x_1^2)-(\Sigma x_3x_1)(\Sigma x_2x_1)}{(\Sigma x_2^2)(\Sigma x_1^2)-(\Sigma x_2x_1)^2}$$

The following examples would clarify the method:

Example 1. **From the following data, find the least square regression of X_3 on X_1 and X_2 using actual mean method. Also estimate X_3 when $X_1=10$ and $X_2=6$.**

X_1 :	3	5	6	8	12	14
X_2 :	16	10	7	4	3	2
X_3 :	90	72	54	42	30	12

Solution.

X_1	$x_1=(X_1-\overline{X}_1)$	x_1^2	X_2	$x_2=(X_2-\overline{X}_2)$	x_2^2	X_3	$x_3=(X_3-\overline{X}_3)$	x_3^2	x_1x_2	x_1x_3	x_2x_3
3	−5	25	16	+9	81	90	+40	1600	−45	−200	+360
5	−3	9	10	+3	9	72	+22	484	−9	−66	+66
6	−2	4	7	0	0	54	+4	16	0	−8	0
8	0	0	4	−3	9	42	−8	64	0	0	+24
12	+4	16	3	−4	16	30	−20	400	−16	−80	+80
14	+6	36	2	−5	25	12	−38	1444	−30	−228	+190
$\Sigma X_1=48$	$\Sigma x_1=0$	$\Sigma x_1^2=90$	$\Sigma X_2=42$	$\Sigma x_2=0$	$\Sigma x_2^2=140$	$\Sigma X_3=300$	$\Sigma x_3=0$	$\Sigma x_3^2=4008$	$\Sigma x_1x_2=-100$	$\Sigma x_1x_3=-582$	$\Sigma x_2x_3=720$

$$\overline{X}_1=\frac{48}{6}=8,\ \overline{X}_2=\frac{42}{6}=7,\ \overline{X}_3=\frac{300}{6}=50,$$

Regression Equation of X_3 on X_1 and X_2 is:

$$X_3-\overline{X}_3=b_{31.2}(X_1-\overline{X}_1)+b_{32.1}(X_2-\overline{X}_2)$$

$$b_{31.2}=\frac{(\Sigma x_3x_1)(\Sigma x_2^2)-(\Sigma x_3x_2)(\Sigma x_1x_2)}{(\Sigma x_1^2)(\Sigma x_2^2)-(\Sigma x_1x_2)^2}$$

$$=\frac{(-582)(140)-(720)(-100)}{(90)(140)-(-100)^2}$$

$$=\frac{-81480+72000}{12600-10000}=\frac{-9480}{2600}=-3.646$$

$$b_{32.1}=\frac{(\Sigma x_3x_2)(\Sigma x_1^2)-(\Sigma x_3x_1)(\Sigma x_2x_1)}{(\Sigma x_2^2)(\Sigma x_1^2)-(\Sigma x_2x_1)^2}$$

$$=\frac{(720)(90)-(-582)(-100)}{(90)(140)-(-100)^2}$$

$$=\frac{64800-58200}{12600-10000}=\frac{6600}{2600}=2.538$$

Substituting the values in the above equations, we get

$$X_3-50=-3.646(X_1-8)+2.538(X_2-7)$$

$$X_3-50=-3.646X_1+29.168+2.538X_2-17.766$$

$$X_3=-3.646X_1+2.538X_2+61.402$$

When $X_1=10$ and $X_2=6$, So, $X_3=-3.646(10)+2.538(6)+61.402$

$$=-36.46+15.228+61.402=40.17 \text{ or } 40.$$

Example 3. **Given the following information (variables are measured from their respective means):**

$$\Sigma x_1x_2=720,\ \Sigma x_2x_3=-582,\ \Sigma x_1x_3=-100$$

$$\Sigma x_2^2=4008,\ \Sigma x_3^2=90,\ \Sigma x_1^2=140$$

$$\overline{X}_1=7,\ \overline{X}_2=50,\ \overline{X}_3=8$$

Find the multiple regression equation of X_1 on X_2 and X_3. Estimate X_1 when $X_2=10$ and $X_3=95$.

Solution. Regression Equation of X_1 on X_2 and X_3 is given by :

$$X_1-\overline{X}_1=b_{12.3}(X_2-\overline{X}_2)+b_{13.2}(X_3-\overline{X}_3)$$

$$b_{12.3}=\frac{(\Sigma x_1x_2)(\Sigma x_3^2)-(\Sigma x_1x_3)(\Sigma x_2x_3)}{(\Sigma x_2^2)(\Sigma x_3^2)-(\Sigma x_2x_3)^2}$$

$$=\frac{(720)(90)-(-100)(-582)}{(4008)(90)-(-582)^2}$$

$$=\frac{64800-58200}{360720-338724}=\frac{6600}{21996}=0.30$$

$$b_{13.2}=\frac{(\Sigma x_1x_3)(\Sigma x_2^2)-(\Sigma x_1x_2)(\Sigma x_3x_2)}{(\Sigma x_2^2)(\Sigma x_3^2)-(\Sigma x_2x_3)^2}$$

$$=\frac{(-100)(4008)-(720)(-582)}{(4008)(90)-(-582)^2}$$

$$=\frac{-400800+419040}{360720-338724}=\frac{18240}{21996}=0.829=0.83$$

We are given : $\overline{X}_1=7,\ \overline{X}_2=50,\ \overline{X}_3=8$

Substituting the values in the above equation, we get

$$X_1 - 7 = 0.30\,(X_2 - 50) + 0.83\,(X_3 - 8)$$

$$X_1 - 7 = 0.30X_2 - 15 + 0.83X_3 - 6.64$$

or

$\therefore X_1 = 0.30X_2 + 0.83X_3 - 14.64$ is the required equation

When $X_2 = 20$ and $X_3 = 30$

$$X_1 = 0.30(20) + 0.83(30) - 14.64 = 6 + 24.9 - 14.64 = 16.26$$

Example 4. **The following data for three variables X_1, X_2 and X_3 are given below :**

$\Sigma x_1x_2 = 218$, $\Sigma x_1x_3 = -198$, $\Sigma x_2x_3 = -93$

$\Sigma x_1^2 = 454$, $\Sigma x_2^2 = 110$, $\Sigma x_3^2 = 90$

x_1, x_2 and x_3 are measured from their means. Find the two partial regression coefficients ($b_{12.3}$ and $b_{13.2}$).

Solution.

$$b_{12.3} = \frac{(\Sigma x_1x_2)(\Sigma x_3^2) - (\Sigma x_1x_3)(\Sigma x_2x_3)}{(\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2x_3)^2}$$

$$= \frac{(218)(90) - (-198)(-93)}{(90)(110) - (-93)^2}$$

$$= \frac{19620 - 18414}{9900 - 8649}$$

$$= \frac{1206}{1251} = 0.964$$

$$b_{13.2} = \frac{(\Sigma x_1x_3)(\Sigma x_2^2) - (\Sigma x_1x_2)(\Sigma x_3x_2)}{(\Sigma x_3^2)(\Sigma x_2^2) - (\Sigma x_3x_2)^2}$$

$$= \frac{(-198)(110) - (218)(-93)}{(90)(110) - (-93)^2}$$

$$= \frac{-21780 + 20274}{9900 - 8649} = \frac{-1506}{1251} = -1.203$$

EXERCISE – 3

1. For the following set of data, find the multiple regression of X_1 on X_2 and X_3 using actual mean method. Also predict the value of X_1 when $X_2 = 5$ and $X_3 = 7$:

X_1 :	12	24	32	28
X_2 :	6	12	16	22
X_3 :	4	6	12	18

[**Ans.** $X_1 = 2.577 + 1.661\,X_2 + 0.0169\,X_3$, $X_1 = 11.0$]

2. From the data given below, find the multiple regression equation of X_1 on X_2 and X_3 using actual mean method :

X_1:	4	6	7	9	13	15
X_2 :	15	12	8	6	4	3
X_3 :	30	24	20	14	10	4

[**Ans.** $X_1 = 16.479 + 0.389\,X_2 + 0.623\,X_3$]

3. From the data given below, find the multiple linear regression of X_1 on X_2 and X_3 using actual mean method :

X_1 :	18	20	17	14	21
X_2 :	38	40	25	28	44
X_3 :	20	15	5	12	18

[**Ans.** $X_1 = 0.5\,X_2 - 0.36X_3 + 5.54$]

4. Given the following information (variables are measured from their respective means):

$\Sigma x_1x_2 = 1900$, $\Sigma x_1x_3 = -20$, $\Sigma x_2x_3 = -50$,

$\Sigma x_1^2 = 1350$, $\Sigma x_2^2 = 2800$, $\Sigma x_3^2 = 24$,

$\overline{X}_1 = 65$, $\overline{X}_2 = 55$, $\overline{X}_3 = 30$,

Obtain the partial regression coefficients ($b_{12.3}$ and $b_{13.2}$)

Also estimate the value of X_1 when $X_2 = 60$ and $X_3 = 25$.

[**Ans.** $b_{12.3} = 0.689$, $b_{13.2} = 0.603$, $X_1 = 65.43$]

5. Given the following information (variates are measured from their respective means) :

$\Sigma x_1^2 = 1350$, $\Sigma x_2^2 = 2800$, $\Sigma x_3^2 = 24$

$\Sigma x_1x_2 = 1900$, $\Sigma x_1x_3 = -20$, $\Sigma x_2x_3 = -50$

Determine the regression equation of X_1 on X_2 and X_3. [**Ans.** $X_1 = 0.689\,X_2 + .603\,X_3$]

(2) Multiple Regression Equations in terms of Simple Correlation Coefficients

When the values of $\overline{X}_1, \overline{X}_2$ and $\overline{X}_3, \sigma_1, \sigma_2$ and σ_3 and r_{12}, r_{13} and r_{23} are given, then the multiple regression equations are expressed in the following manner :

(1) Multiple Regression Equation of X_1 on X_2 and X_3

$$X_1 - \overline{X}_1 = b_{12.3}\,(X_2 - \overline{X}_2) + b_{13.2}\,(X_3 - \overline{X}_3)$$

or $x_1 = b_{12.3}\,x_2 + b_{13.2}\,x_3$ Where, $x_1 = X_1 - \overline{X}_1$, $x_2 = X_2 - \overline{X}_2$, $x_3 = X_3 - \overline{X}_3$

The values of partial regression coefficients $b_{12.3}$ and $b_{13.2}$ are determined by using the following formulae :

$$b_{12.3} = \left[\frac{\sigma_1}{\sigma_2}\right] \cdot \left[\frac{r_{12} - r_{13} \cdot r_{23}}{1 - r_{23}^2}\right]$$

$$b_{13.2} = \left[\frac{\sigma_1}{\sigma_3}\right] \cdot \left[\frac{r_{13} - r_{12} \cdot r_{32}}{1 - r_{32}^2}\right]$$

Multiple Regression Equation of X_1 on X_2 and X_3 can also be written as :

$$x_1 = \left[\frac{\sigma_1}{\sigma_2}\right]\left[\frac{r_{12} - r_{13} \cdot r_{23}}{1 - r_{23}^2}\right]x_2 + \left[\frac{\sigma_1}{\sigma_3}\right]\left[\frac{r_{13} - r_{12} \cdot r_{32}}{1 - r_{32}^2}\right]x_3$$

or

$$X_1 - \overline{X}_1 = \left[\frac{\sigma_1}{\sigma_2}\right]\left[\frac{r_{12} - r_{13} \cdot r_{23}}{1 - r_{23}^2}\right](X_2 - \overline{X}_2) + \left[\frac{\sigma_1}{\sigma_3}\right]\left[\frac{r_{13} - r_{12} \cdot r_{32}}{1 - r_{23}^2}\right](X_3 - \overline{X}_3)$$

(2) Multiple Regression Equation of X_2 on X_1 and X_3 :

$$X_2 - \overline{X}_2 = b_{21.3}\,(X_1 - \overline{X}_1) + b_{23.1}\,(X_3 - \overline{X}_3)$$

or

$$x_2 = b_{21.3}x_1 + b_{23.1}x_3$$

Where,

$$b_{21.3} = \left[\frac{\sigma_2}{\sigma_1}\right]\left[\frac{r_{21} - r_{23} \cdot r_{13}}{1 - r_{13}^2}\right]$$

$$b_{23.1} = \left[\frac{\sigma_2}{\sigma_3}\right]\left[\frac{r_{23} - r_{21} \cdot r_{31}}{1 - r_{31}^2}\right]$$

Multiple Regression Equation of X_2 on X_1 and X_3 can also be written :

$$x_2 = \left[\frac{\sigma_2}{\sigma_1}\right]\left[\frac{r_{21} - r_{23} \cdot r_{13}}{1 - r_{13}^2}\right]x_1 + \left[\frac{\sigma_2}{\sigma_3}\right]\left[\frac{r_{23} - r_{21} \cdot r_{31}}{1 - r_{31}^2}\right]x_3$$

or $$X_2 - \overline{X}_1 = \left[\frac{\sigma_2}{\sigma_1}\right]\left[\frac{r_{21} - r_{23} \cdot r_{13}}{1 - r_{13}^2}\right](X_1 - \overline{X}_1) + \left[\frac{\sigma_2}{\sigma_3}\right]\left[\frac{r_{23} - r_{21} \cdot r_{31}}{1 - r_{31}^2}\right](X_3 - \overline{X}_3)$$

(3) Multiple Regression Equation of X_3 on X_1 and X_2 :

$$X_3 - \overline{X}_3 = b_{31.2}(X_1 - \overline{X}_1) + b_{32.1}(X_2 - \overline{X}_2)$$

or $$x_3 = b_{31.2}\, x_1 + b_{32.1}\, x_2$$

Where $$b_{31.2} = \left[\frac{\sigma_3}{\sigma_1}\right]\left[\frac{r_{31} - r_{32} \cdot r_{12}}{1 - r_{12}^2}\right]$$

$$b_{32.1} = \left[\frac{\sigma_3}{\sigma_2}\right]\left[\frac{r_{32} - r_{31} \cdot r_{21}}{1 - r_{21}^2}\right]$$

Multiple regression on X_3 on X_1 and X_2 can also be written as :

$$x_3 = \left[\frac{\sigma_3}{\sigma_1}\right]\left[\frac{r_{31} - r_{32} \cdot r_{12}}{1 - r_{12}^2}\right]x_1 + \left[\frac{\sigma_3}{\sigma_2}\right]\left[\frac{r_{32} - r_{31} \cdot r_{21}}{1 - r_{21}^2}\right]x_2$$

or $$X_3 - \overline{X}_3 = \left[\frac{\sigma_3}{\sigma_1}\right]\left[\frac{r_{31} - r_{32} \cdot r_{12}}{1 - r_{12}^2}\right](X_1 - \overline{X}_1) + \left[\frac{\sigma_3}{\sigma_2}\right]\left[\frac{r_{32} - r_{31} \cdot r_{21}}{1 - r_{21}^2}\right](X_2 - \overline{X}_2)$$

Note : $r_{12} = r_{21}, r_{23} = r_{32}, r_{13} = r_{31}$.

The following examples would clarify the procedure :

Example 5. **A teacher in mathematics wishes to determine the relationship of marks in final examination to those in two tests given during the semester. Calling X_1, X_2 and X_3, the marks of a student on Ist, 2nd and final examination respectively, he made the following computations from a total of 120 students :**

$$\overline{X}_1 = 6.8 \quad \overline{X}_2 = 7.0 \quad \overline{X}_3 = 74$$

$$\sigma_1 = 1.0 \quad \sigma_2 = 0.80 \quad \sigma_3 = 9.0$$

$$r_{12} = 0.60 \quad r_{13} = 0.70 \quad r_{23} = 0.65$$

(i) Find the relevant regression equation.

(ii) Estimate the final marks of two students who secured respectively 9 and 7, 4 and 8 on the two tests.

Solution. The relevant least square regression equation will be X_3 on X_1 and X_2 which is given by :

$$X_3 - \overline{X}_3 = b_{31.2}(X_1 - \overline{X}_1) + b_{32.1}(X_2 - \overline{X}_2)$$

$$b_{31.2} = \frac{\sigma_3}{\sigma_1} \cdot \left[\frac{r_{31} - r_{32} \cdot r_{12}}{1 - r_{12}^2}\right]$$

$$= \frac{9}{1} \times \left[\frac{(.70) - (.65)(.60)}{1 - (.60)^2}\right]$$

$$= 9 \times \left[\frac{(.70) - (.39)}{1 - (.60)^2}\right] = 9 \times \left[\frac{.31}{.64}\right] = \frac{2.79}{.64} = 4.36$$

$$b_{32.1} = \frac{\sigma_3}{\sigma_2} \cdot \left[\frac{r_{32} - r_{31} \cdot r_{21}}{1 - r_{21}^2}\right]$$

$$= \frac{9}{.80} \times \left[\frac{(.65) - (.70)(.60)}{1 - (.60)^2}\right]$$

$$= \frac{9}{.80} \times \left[\frac{.65 - .42}{0.64}\right] = 4.04$$

Thus, the regression equation of X_3 on X_1 and X_2 is

$$X_3 - 74 = 4.36(X_1 - 6.8) + 4.04(X_2 - 7)$$

$\therefore$ $$X_3 = 16.07 + 4.36X_1 + 4.04X_2$$

Final marks of students who scored 9 and 7 marks :

When $X_1 = 9$ and $X_2 = 7$

$$X_3 = 16.07 + 4.36(9) + 4.04(7)$$

$$= 16.07 + 39.24 + 28.28 = 83.59 \text{ or } 84$$

Final marks of students who scored 4 and 8 marks

When $X_1 = 4$ and $X_2 = 8$

$$X_3 = 16.07 + 4.36(4) + 4.04(8)$$

$$= 16.07 + 17.44 + 32.32 = 65.8 \text{ or } 66$$

Example 6. **Given the following, determine the regression equations of :**

(i) x_1 on x_2 and x_3 and

(ii) x_2 on x_1 and x_3 when the variates are measured from their means :

$$r_{12} = 0{\cdot}8 \quad r_{13} = 0{\cdot}6 \quad r_{23} = 0{\cdot}5$$

$$\sigma_1 = 10, \sigma_2 = 8, \sigma_3 = 5$$

Solution. (i) The regression equation of x_1 on x_2 and x_3 when variates are measured from means is given by :

$$x_1 = b_{12.3}x_2 + b_{13.2}x_2 \quad \text{where, } x_1 = X_1 - \overline{X}_1, x_2 = X_2 - \overline{X}_2, x_3 = X_3 - \overline{X}_3$$

$$b_{12\cdot3}=\left[\frac{\sigma_1}{\sigma_2}\right]\times\left[\frac{r_{12}-r_{13}\cdot r_{23}}{1-r_{23}^2}\right]$$

$$=\left[\frac{10}{8}\right]\times\left[\frac{(0\cdot8)-(0\cdot6)(0\cdot5)}{1-(0\cdot5)^2}\right]=0\cdot833$$

$$b_{13\cdot2}=\left[\frac{\sigma_1}{\sigma_3}\right]\times\left[\frac{r_{13}-r_{12}\cdot r_{32}}{1-r_{32}^2}\right]$$

$$=\left[\frac{10}{5}\right]\times\left[\frac{(0\cdot6)-(0\cdot8)(0\cdot5)}{1-(0\cdot5)^2}\right]=0\cdot533$$

∴ The required regression equation is :

$$x_1=\cdot833\,x_2+\cdot533\,x_3$$

(ii) The regression equation of x_2 on x_1 and x_3 when variates are measured from means is given by :

$$x_2=b_{21\cdot3}\,x_1+b_{23\cdot1}\,x_3$$

$$b_{21\cdot3}=\left[\frac{\sigma_2}{\sigma_1}\right]\times\left[\frac{r_{21}-r_{23}\cdot r_{13}}{1-r_{13}^2}\right]$$

$$=\left[\frac{8}{10}\right]\times\left[\frac{(0\cdot8)-(\cdot5)(\cdot6)}{1-(\cdot6)^2}\right]=\cdot625$$

$$b_{23\cdot1}=\left[\frac{\sigma_2}{\sigma_3}\right]\times\left[\frac{r_{23}-r_{21}\cdot r_{31}}{1-r_{31}^2}\right]$$

$$=\left[\frac{8}{5}\right]\times\left[\frac{(\cdot5)-(0\cdot8)(0\cdot6)}{1-(\cdot6)^2}\right]=0\cdot05$$

∴ The required regression equation is :

$$x_2=\cdot625\,x_1+\cdot05\,x_3$$

Example 7. **A random sample of 15 students of Basic Statistics course when observed for weights (X_1), age (X_2) and height (X_3) offered the following information:**

$r_{12}=0\cdot8$, $r_{23}=0\cdot3$, $r_{13}=0\cdot5$, $S_1=8\cdot5$, $S_2=4\cdot5$, $S_3=2\cdot1$

$\overline{X}_1=70$ kg, $\overline{X}_2=22$ yrs and $\overline{X}_3=160$ cms.

Obtain :

(i) Multiple and partial correlation coefficients $R_{1\cdot23}$ and $r_{13\cdot2}$.

(ii) Multiple regression of X_1 on X_2 and X_3 and estimate the value of X_1 for $X_2=25$ yrs and $X_3=140$ cms.

Solution. (i)

$$R_{1\cdot23}=\sqrt{\frac{r_{12}^2+r_{13}^2-2r_{12}\cdot r_{13}\cdot r_{23}}{1-r_{23}^2}}$$

$$=\sqrt{\frac{(0\cdot8)^2+(0\cdot5)^2-2(0\cdot8)(0\cdot5)(0\cdot3)}{1-(0\cdot3)^2}}$$

$$=\sqrt{\frac{0\cdot64+0\cdot25-0\cdot24}{0\cdot91}}=\sqrt{\frac{0\cdot65}{0\cdot91}}=0\cdot8452$$

$$r_{12\cdot3}=\frac{r_{12}-r_{13}\cdot r_{23}}{\sqrt{1-r_{13}^2}\,\sqrt{1-r_{23}^2}}$$

$$=\frac{(0\cdot8)-(0\cdot5)(0\cdot3)}{\sqrt{1-(0\cdot5)^2}\,\sqrt{1-(0\cdot3)^2}}$$

$$=\frac{0\cdot8-0\cdot15}{\sqrt{\cdot75}\,\sqrt{0\cdot91}}=\frac{0\cdot65}{0\cdot8261}=0\cdot7868$$

(ii) Multiple Regression on X_1 on X_2 and X_3 is given by :

$$X_1-\overline{X}_1=b_{12\cdot3}(X_2-\overline{X}_2)+b_{13\cdot2}(X_3-\overline{X}_3)$$

$$b_{12\cdot3}=\left[\frac{S_1}{S_2}\right]\times\left[\frac{r_{12}-r_{13}\cdot r_{23}}{1-r_{23}^2}\right]$$

$$=\left[\frac{8\cdot5}{4\cdot5}\right]\times\left[\frac{0\cdot8-(0\cdot5)(0\cdot3)}{1-(\cdot3)^2}\right]=1\cdot349$$

$$b_{13\cdot2}=\left[\frac{S_1}{S_3}\right]\times\left[\frac{r_{13}-r_{12}\cdot r_{32}}{1-r_{32}^2}\right]$$

$$=\left[\frac{8\cdot5}{2\cdot1}\right]\left[\frac{(0\cdot5)-(0\cdot8)(0\cdot3)}{1-(0\cdot3)^2}\right]=1\cdot156$$

Substituting the values in the equation, we get

$$X_1-70=1\cdot349\,(X_2-22)+1\cdot156\,(X_3-160)$$

$$X_1-70=1\cdot349X_2-29\cdot678+1\cdot156X_3-184\cdot96$$

$$\therefore\quad X_1=1\cdot349X_2+1\cdot156X_3-144\cdot638$$

Estimation of X_1 for $X_2=25$ and $X_3=140$:

When $X_2=25$ and $X_3=140$, $X_1=1\cdot349(25)+1\cdot156(140)-144\cdot638$

$=33\cdot725+161\cdot84-144\cdot638=50\cdot927$

Example 8. **In a trivariate distribution :**

$\sigma_1=3$, $\sigma_2=4$, $\sigma_3=5$

$r_{23}=0\cdot4$, $r_{31}=0\cdot6$, $r_{12}=0\cdot7$

(i) Compute $r_{23\cdot1}$ and $R_{1\cdot23}$

(ii) Determine the regression equation of x_1 on x_2 and x_3 if the variates are measured from their means :

Solution. (i)

$$r_{23\cdot1}=\frac{r_{23}-r_{21}\cdot r_{31}}{\sqrt{1-r_{21}^2}\,\sqrt{1-r_{31}^2}}$$

$$= \frac{(0\cdot4)-(0\cdot7)(0\cdot6)}{\sqrt{1-(0\cdot7)^2}\sqrt{1-(0\cdot6)^2}}$$

$$= \frac{0\cdot4-(\cdot42)}{\sqrt{0\cdot51}\sqrt{0\cdot64}} = \frac{-0\cdot02}{0\cdot5713} = -0\cdot035$$

$$R_{1\cdot23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\cdot r_{13}\cdot r_{23}}{1-r_{23}^2}}$$

$$= \sqrt{\frac{(0\cdot7)^2 + (0\cdot6)^2 - 2(0\cdot7)(0\cdot6)(0\cdot4)}{1-(0\cdot4)^2}}$$

$$= \sqrt{\frac{0\cdot49+0\cdot36-0\cdot336}{0\cdot84}} = \sqrt{\frac{0\cdot514}{0\cdot84}} = 0\cdot782$$

(ii) The regression equation of x_1 on x_2 and x_3 when variates are measured from mean is given by :

$$x_1 = b_{12\cdot3}x_2 + b_{13\cdot2}x_3 \quad \text{where, } x_1 = X_1 - \overline{X}_1,\ x_2 = X_2 - \overline{X}_2,\ x_3 = X_3 - \overline{X}_3$$

$$b_{12\cdot3} = \left[\frac{\sigma_1}{\sigma_2}\right] \times \left[\frac{r_{12} - r_{13}\cdot r_{23}}{1-r_{23}^2}\right]$$

$$= \left[\frac{3}{4}\right]\left[\frac{(0\cdot7)-(0\cdot6)(0\cdot4)}{1-(0\cdot4)^2}\right]$$

$$= \frac{0\cdot75\times0\cdot46}{0\cdot84} = \frac{0\cdot345}{0\cdot84} = 0\cdot41$$

$$b_{13\cdot2} = \left[\frac{\sigma_1}{\sigma_3}\right] \times \left[\frac{r_{13} - r_{12}\cdot r_{32}}{1-r_{32}^2}\right]$$

$$= \left[\frac{3}{5}\right] \times \left[\frac{(0\cdot6)-(0\cdot7)(0\cdot4)}{1-(0\cdot4)^2}\right] = \frac{0\cdot6\times0\cdot32}{0\cdot84} = \frac{0\cdot192}{0\cdot84} = 0\cdot229$$

Thus, the required regression equation is :

$$x_1 = 0\cdot41x_2 + 0\cdot229x_3.$$

STANDARD ERROR OF ESTIMATE (OR RELIABILITY OF ESTIMATES) FOR MULTIPLE REGRESSION

The standard error of estimate measures the realiability of the estimates given by the multiple regression equation. It shows to what extent the estimated valued given by the regression equations are closer to the actual values.

For three regression equations, there are three standard error or estimates :

(1) Standard Error of Estimate of X_1 on X_2 and X_3 ($S_{1\cdot23}$)

(2) Standard Error of Estimate of X_2 on X_1 and X_3 ($S_{2\cdot13}$)

(3) Standard Error of Estimate of X_3 on X_1 and X_2 ($S_{3\cdot12}$)

The formulae for calculating the standard error of estimates are given as follows :

$$S_{1\cdot23} = \sigma_1 \cdot \sqrt{\frac{1-r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}\cdot r_{13}\cdot r_{23}}{1-r_{23}^2}}$$

$$S_{2\cdot13} = \sigma_2 \cdot \sqrt{\frac{1-r_{21}^2 - r_{23}^2 - r_{13}^2 + 2r_{21}\cdot r_{23}\cdot r_{13}}{1-r_{13}^2}}$$

$$S_{3\cdot12} = \sigma_3 \cdot \sqrt{\frac{1-r_{31}^2 - r_{32}^2 - r_{13}^2 + 2r_{31}\cdot r_{32}\cdot r_{12}}{1-r_{12}^2}}$$

Example 9. **If $r_{12} = 0\cdot8$, $r_{13} = 0\cdot5$, $r_{23} = 0\cdot3$ and $S_1 = 8\cdot5$, compute the standard error of estimate of X_1 on X_2 and X_3.**

Solution. Standard Error of Estimate of X_1 on X_2 and X_3 is given by :

$$S_{1\cdot23} = \sigma_1 \cdot \sqrt{\frac{1-r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}\cdot r_{13}\cdot r_{23}}{1-r_{23}^2}}$$

$$= 8\cdot5\sqrt{\frac{1-(0\cdot8)^2-(0\cdot5)^2-(0\cdot3)^2+2(0\cdot8)(0\cdot5)(0\cdot3)}{1-(0\cdot3)^2}} = 4\cdot543$$

Coefficient of Multiple Determination (R^2)

The coefficient of determination in multiple regression denoted by $R_{1\cdot23}^2$ is similar to the coefficient of determination r^2 in the simple linear regression. It represents the proportion (fraction) of the total variation in the dependent variable X_1 that has been explained by the independent variables (X_2 and X_3) in the multiple regression equation.

For example, if $R_{1\cdot23} = 0\cdot7252$, then $R_{1\cdot23}^2 = 0\cdot5259 = 0\cdot526$

The value of $R_{1\cdot23}^2 = 0\cdot526$ indicates that 52·6% variation in the dependent variable (X_1) are explained by the independent variables X_2 and X_3 in the multiple regression equation of X_1 on X_2 and X_3.

Example 10. **A random sample of 15 students of advanced course in statistics when observed for weight (X_1), age (X_2) and height (X_3) offered the following information :**

$r_{12} = 0\cdot8$, $r_{13} = 0\cdot5$, $r_{23} = 0\cdot3$

$S_1 = 8\cdot5$, $S_2 = 4\cdot5$ and $S_3 = 2\cdot1$

Find the following :

(a) Partial regression coefficient $b_{1\cdot23}$ and $b_{13\cdot2}$.

(b) Standard error of estimate $S_{1\cdot23}$.

(c) Correlation Coefficients $R_{1\cdot23}$ and $r_{12\cdot3}$.

(d) Multiple regression of X_1 on X_2 and X_3 when $\overline{X}_1 = 70$ kg, $\overline{X}_2 = 22$ years and $\overline{X}_3 = 150$ cm.

(e) Weight of a student (X_1) of 25 years of age and 140 cm in height.

Solution. Given : $r_{12} = 0\cdot8$, $r_{13} = 0\cdot5$, $r_{23} = 0\cdot3$

$$S_1 = 8\cdot5,\ S_2 = 4\cdot5,\ S_3 = 2\cdot1$$

(a)
$$b_{12\cdot3} = \left[\frac{S_1}{S_2}\right]\cdot\left[\frac{r_{12} - r_{13}\cdot r_{23}}{1 - r_{23}^2}\right]$$

$$= \left[\frac{8\cdot5}{4\cdot5}\right]\times\left[\frac{0\cdot8 - (0\cdot5)(0\cdot3)}{1 - (0\cdot3)^2}\right] = 1\cdot349$$

$$b_{13\cdot2} = \left[\frac{S_1}{S_3}\right]\times\left[\frac{r_{13} - r_{12}\cdot r_{32}}{1 - r_{32}^2}\right]$$

$$= \left[\frac{8\cdot5}{2\cdot1}\right]\times\left[\frac{(0\cdot5) - (0\cdot8)(0\cdot3)}{1 - (0\cdot3)^2}\right] = 1\cdot156$$

(b)
$$S_{1\cdot23} = S_1\sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}\cdot r_{13}\cdot r_{23}}{1 - r_{23}^2}}$$

$$= 8\cdot5\times\sqrt{\frac{1 - (0\cdot8)^2 - (0\cdot5)^2 - (0\cdot3)^2 + 2(0\cdot8)(0\cdot5)(0\cdot3)}{1 - (0\cdot3)^2}}$$

$$= 8\cdot5\times\sqrt{\frac{1 - \cdot64 - \cdot25 - 0\cdot09 + 0\cdot24}{0\cdot91}} = 4\cdot543$$

(c)
$$R_{1\cdot23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\cdot r_{13}\cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0\cdot8)^2 + (0\cdot5)^2 - 2(0\cdot8)(0\cdot5)(0\cdot3)}{1 - (0\cdot3)^2}}$$

$$= \sqrt{\frac{0\cdot64 + \cdot25 - 0\cdot24}{0\cdot91}} = \sqrt{\frac{0\cdot65}{0\cdot91}} = 0\cdot8452$$

$$r_{12\cdot3} = \frac{r_{12} - r_{13}\cdot r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

$$= \frac{(0\cdot8) - (0\cdot5).(0\cdot3)}{\sqrt{1 - (0\cdot5)^2}\sqrt{1 - (0\cdot3)^2}} = \frac{0\cdot65}{0\cdot8261} = 0\cdot7868$$

(d) Multiple Regression Equation of X_1 on X_2 and X_3 :

$$X_1 - \overline{X}_1 = b_{12\cdot3}(X_2 - \overline{X}_2) + b_{13\cdot2}(X_3 - \overline{X}_3)$$

Substituting the values, we have

$$X_1 - 70 = 1\cdot349(X_2 - 22) + 1\cdot156(X_3 - 150)$$

$$X_1 = -133\cdot078 + 1\cdot349X_2 + 1\cdot156X_3$$

(e) For $X_2 = 25$ and $X_3 = 140$,

$$X_1 = -133\cdot078 + 1\cdot348(25) + 1\cdot156(140)$$

$$= -133\cdot078 + 33\cdot725 + 161\cdot84 = 62\cdot487$$

Example 11. **Given the following data, determine the regression equation of x_1 on x_2 and x_3 if the variates are measured from their means :**

$$r_{12} = 0\cdot8,\quad r_{13} = 0\cdot6,\quad r_{23} = 0\cdot5$$
$$\sigma_1 = 10,\quad \sigma_2 = 8,\quad \sigma_3 = 15$$

Also find the standard error of the estimate of x_1 on x_2 and x_3.

Solution. The regression equation of x_1 on x_2 and x_3 is :

$$x_1 = b_{12\cdot3}x_2 + b_{13\cdot2}x_3 \text{ where, } x_1 = X_1 - \overline{X}_1, x_2 = X_2 - \overline{X}_2 \text{ and } x_3 = X_3 - \overline{X}_3$$

Here,
$$b_{12\cdot3} = \left[\frac{\sigma_1}{\sigma_2}\right]\cdot\left[\frac{r_{12} - r_{13}\cdot r_{23}}{1 - r_{23}^2}\right]$$

$$= \left[\frac{10}{8}\right]\times\left[\frac{0\cdot8 - (0\cdot6)(0\cdot5)}{1 - (0\cdot5)^2}\right]$$

$$= \left[\frac{10}{8}\right]\times\left[\frac{0\cdot8 - \cdot30}{1 - \cdot25}\right] = \frac{10}{8}\times\frac{0\cdot50}{0\cdot75} = 0\cdot833$$

$$b_{13\cdot2} = \frac{\sigma_1}{\sigma_3}\cdot\left[\frac{r_{13} - r_{12}\cdot r_{32}}{1 - r_{32}^2}\right]$$

$$= \frac{10}{5}\times\left[\frac{(0\cdot6) - (0\cdot8)(0\cdot5)}{1 - (0\cdot5)^2}\right]$$

$$= \frac{10}{5}\times\frac{0\cdot20}{0\cdot75} = \frac{2}{3\cdot75} = 0\cdot53$$

Thus, regression equation of x_1 on x_2 and x_3 is :

$$x_1 = 0\cdot833x_2 + 0\cdot53x_3$$

Standard Error of Estimate of X_1 on X_2 and X_3

$$S_{1\cdot23} = \sigma_1\sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}\cdot r_{13}\cdot r_{23}}{1 - r_{23}^2}}$$

$$= 10.\sqrt{\frac{1 - (0\cdot8)^2 - (0\cdot6)^2 - (0\cdot5)^2 - 2(0\cdot8)(0\cdot6)(0\cdot5)}{1 - (0\cdot5)^2}}$$

$$= 10.\sqrt{\frac{1 - \cdot64 - \cdot36 - \cdot25 + \cdot48}{0\cdot75}}$$

$$= 10.\sqrt{\frac{0\cdot23}{0\cdot75}} = 10\times\cdot5537 = 5\cdot537$$

Example 11A. **The following values have been obtained from the measurement of three variables x_1, x_2 and x_3 :**

$$\overline{X}_1 = 6\cdot8 \quad \overline{X}_2 = 7\cdot0 \quad \overline{X}_3 = 7\cdot4$$
$$S_1 = 1\cdot0 \quad S_2 = 0\cdot80 \quad S_3 = 0\cdot90$$
$$r_{12} = 0\cdot60 \quad r_{13} = 0\cdot70 \quad r_{23} = 0\cdot65$$

(i) Obtain regression eqaution of X_1 on X_2 and X_3.

(ii) Estimate the value of X_1 for $X_2 = 10$ and $X_3 = 9$.

(iii) Find the coefficient of multiple determination $R^2_{1\cdot23}$ from r_{12} and $r_{13\cdot2}$.

Solution. The regression equation of X_1 on X_2 and X_3 is given by :

$$X_1 - \overline{X}_1 = b_{12\cdot3}(X_2 - \overline{X}_2) + b_{13\cdot2}(X_3 - \overline{X}_3) \quad \text{...(i)}$$

where,

$$b_{12\cdot3} = \frac{s_1}{s_2}\left[\frac{r_{12} - r_{13}\cdot r_{23}}{1 - r_{23}^2}\right] = \frac{1}{0\cdot80}\left[\frac{0\cdot60 - 0\cdot70 \times 0\cdot65}{1-(0\cdot65)^2}\right]$$

or

$$b_{12\cdot3} = (1\cdot25)\left[\frac{0\cdot60 - 0\cdot455}{0\cdot578}\right] = 0\cdot313$$

$$b_{13\cdot2} = \frac{s_1}{s_2}\cdot\left[\frac{r_{13} - r_{12}\cdot r_{32}}{1 - r_{32}^2}\right] = \frac{1}{0\cdot90}\left[\frac{0\cdot70 - 0\cdot60 \times 0\cdot65}{1-(0\cdot65)^2}\right]$$

$$= (1\cdot111)\left[\frac{0\cdot70 - 0\cdot39}{0\cdot578}\right] = 0\cdot595$$

Substituting the values in equation (i), we have,

$$X_1 - 6\cdot8 = 0\cdot313(X_2 - 7\cdot0) + 0\cdot595(X_3 - 7\cdot4)$$

or

$$X_1 = 0\cdot206 + 0\cdot313 X_2 + 0\cdot595 X_3$$

(ii) Substituting for $X_2 = 10$ and $X_3 = 9$ in the above regression and solving for x_1.

$$X_1 = 0\cdot206 + 0\cdot313(10) + 0\cdot595(9) = 8\cdot691$$

(iii) Multiple and partial correlation coefficients are related as :

$$R^2_{12\cdot3} = 1 - (1 - r_{12}^2)(1 - r_{13\cdot2}^2)$$

$$r_{13\cdot2} = \frac{r_{13} - r_{12}\cdot r_{32}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{32}^2}} = \frac{0\cdot70 - 0\cdot60 \times 0\cdot65}{\sqrt{1-(0\cdot60)^2}\sqrt{1-(0\cdot65)^2}}$$

$$= \frac{0\cdot70 - 0\cdot39}{0\cdot8 \times 0\cdot760} = 0\cdot509$$

or,

$$r^2_{13\cdot2} = 0\cdot259$$

Substituting the values of r_{12}^2 and $r_{13\cdot2}^2$ for $R^2_{1\cdot23}$ we have

$$R^2_{1\cdot23} = 1 - (1 - 0\cdot36)(1 - 0\cdot259) = 0\cdot526$$

EXERCISE – 4

1. The following constants are obtained from measurements of length in mm (x_1), volume in c.c. (x_2) and weight in gm (x_3) of 300 eggs :

$\overline{X}_1 = 55\cdot95$ $\quad S_1 = 2\cdot26$ $\quad r_{12} = 0\cdot578$

$\overline{X}_2 = 51\cdot48$ $\quad S_2 = 4\cdot39$ $\quad r_{13} = 0\cdot581$

$\overline{X}_3 = 56\cdot03$ $\quad S_3 = 4\cdot41$ $\quad r_{23} = 0\cdot974$

Obtain the linear regression equation of egg weight on egg length and egg volume. Hence estimate the weight of an egg whose length is 58 mm and volume is 52·5 c.c.

[**Ans.** $X_3 = 3\cdot54 + 0\cdot052X_1 + 0\cdot963X_2$, $X_3 = 57\cdot11$ gms.]

2. In a trivarlate distribution :

$\sigma_1 = 2\cdot7,$ $\quad \sigma_2 = 2\cdot4,$ $\quad \sigma_3 = 2\cdot7$

$r_{12} = 0\cdot28,$ $\quad r_{23} = 0\cdot49,$ $\quad r_{31} = 0\cdot51$

Determine the regression equation of x_3 on x_1 and x_2 if the variates are measured from their means.

[**Ans.** $x_3 = 0\cdot405x_1 + 0\cdot424x_2$]

3. Given the following data :

$\overline{X}_1 = 6,$ $\quad \overline{X}_2 = 7,$ $\quad \overline{X}_3 = 8$

$\sigma_1 = 1,$ $\quad \sigma_2 = 2,$ $\quad \sigma_3 = 3$

$r_{12} = 0\cdot6,$ $\quad r_{13} = 0\cdot7,$ $\quad r_{23} = 0\cdot8$

Obtain the linear regression equation of X_3 on X_1 and X_3. Hence estimate X_3 when $X_1 = 4$ and $X_2 = 5$.

[**Ans.** $x_3 = -4\cdot41 + 1\cdot03x_1 + 0\cdot89x_2$, $x_3 = 4\cdot16$]

4. The following results were obtained in the analysis of a trivariate distribution ;

$$S_1 = 3,\ S_2 = S_3 = 5,\ r_{12} = 0\cdot7,\ r_{23} = r_{31} = 0\cdot6$$

Find (i) partial correlation coefficient $r_{12\cdot3}$ (ii) Multiple correlation coefficient $R_{1\cdot23}$ and (iii) the partial regression coefficients ($b_{12\cdot3}$ and $b_{13\cdot2}$).

[**Ans.** $r_{12\cdot3} = 0\cdot531$, $R_{1\cdot23} = 0\cdot735$, $b_{12\cdot3} = 0\cdot319$, $b_{13\cdot2} = 0\cdot169$]

5. If $r_{12} = 0\cdot926$, $r_{13} = 0\cdot891$, $r_{23} = 0\cdot955$ and $S_1 = 1\cdot51$, compute the standard error of estiamte of X_1 on X_2 and X_3 ($S_{1\cdot23}$).

[**Ans.** $S_{1\cdot23} = 0\cdot5702$]

6. A random sample of 50 students of M. Com. when observed for weight (x_1), age (x_2), and height (x_3) offered the following information :

$r_{12} = 0\cdot7,$ $\quad r_{13} = 0\cdot8,$ $\quad r_{23} = 0\cdot5$

$S_1 = 5\cdot6,$ $\quad S_2 = 4\cdot5,$ $\quad S_3 = 3\cdot5$

Obtain the following :

(a) Partial regression coefficient $b_{12\cdot3}$ and $b_{13\cdot2}$.

(b) Standard error of estimate $S_{1\cdot23}$.

(c) Coefficient of multiple correlation ($R_{1\cdot23}$).

(d) Coefficient of partial correlation ($r_{12\cdot3}$).

[**Ans.** $b_{12\cdot3} = 0\cdot496$, $b_{13\cdot2} = 0\cdot96$, $S_{1\cdot23} = 2\cdot74$, $R_{1\cdot23} = 0\cdot87$, $r_{12\cdot3} = 0\cdot516$]

7. Given the following data, determine the regression equation of x_1 on x_2 and x_3 if the variates measured from their means :

$r_{12} = 0\cdot8,$ $\quad r_{13} = 0\cdot6,$ $\quad r_{23} = 0\cdot5$

$\sigma_1 = 10,$ $\quad \sigma_2 = 8,$ $\quad \sigma_3 = 5$

Also find the standard error of estimate of x_1 on x_2 and x_3.

[**Ans.** $x_1 = 0\cdot833x_2 + 0\cdot53x_2$, $S_{1\cdot23} = 5\cdot537$]

8. Given the following data, calculate the estimated value of X_1 when $X_2 = 20$ and $X_3 = 25$.

$\overline{X}_1 = 30$ $\quad S_1 = 5$ $\quad r_{12} = -0\cdot4$

$\overline{X}_2 = 35$ $\quad S_2 = 10$ $\quad r_{13} = -0\cdot5$

$\overline{X}_3 = 40$ $\quad S_3 = 15$ $\quad r_{23} = 0\cdot6$

Also calculate the standard error of estimate of X_1 on X_2 and X_3.

[**Ans.** $X_3 = 38\cdot225 - 0\cdot075X_2 - 0\cdot14X_3$, $X_1 = 33$, $S_{1\cdot23} = 1\cdot1$]

9. Given the following data :

$\overline{X}_1 = 55$ $\quad S_1 = 5$ $\quad r_{12} = 0{\cdot}57$

$\overline{X}_2 = 51$ $\quad S_2 = 7$ $\quad r_{13} = 0{\cdot}58$

$\overline{X}_3 = 56$ $\quad S_3 = 9$ $\quad r_{23} = 0{\cdot}97$

Calculate :

(i) Multiple Regression of X_3 on X_1 and X_2.

(ii) Multiple Correlation coefficient $R_{1{\cdot}23}$

[**Ans.** $X_3 = 0{\cdot}08X_1 + 1{\cdot}21X_2 - 10{\cdot}11$, $R_{1{\cdot}23} = 0{\cdot}97$]

10. From the following data,

$r_{12} = 0{\cdot}8, r_{13} = 0{\cdot}5, r_{23} = 0{\cdot}3$

$S_1 = 8{\cdot}5, S_2 = 4{\cdot}5$ and $S_3 = 2{\cdot}1$

(i) Obtain the regression equation of X_1 on X_2 and X_3 with $\overline{X}_1 = 70$ kg, $\overline{X}_2 = 22$ years, and $\overline{X}_3 = 150$ cm.

(ii) Estimate the value of X_1 on $X_2 = 25$ years and $X_3 = 140$ cm, and

(iii) Find the coefficient of multiple determination $R^2_{1{\cdot}23}$ from r_{12} and $r_{13{\cdot}2}$. What does R^2 indicate ?

[**Ans.** $X_1 = 1{\cdot}349X_2 + 1{\cdot}156X_3 - 133{\cdot}078$, $X_1 = 62{\cdot}487$ $R^2_{1{\cdot}23} = 0{\cdot}7443$, R^2 indicates the 74·43% variation in X_1 are explained by the multiple regression equation]

MISCELLANEOUS SOLVED EXAMPLE

Example 12. **In a trivariate distribution :**

$$\overline{X}_1 = 28{\cdot}02,\ \overline{X}_2 = 4{\cdot}91,\ \overline{X}_3 = 594,\ S_1 = 4{\cdot}4,\ S_2 = 1{\cdot}1,\ S_3 = 80$$

$$r_{12} = 0{\cdot}80,\ r_{23} = -0{\cdot}56,\ r_{31} = -0{\cdot}40$$

(i) Find the correlation coefficient $r_{23{\cdot}1}$ and $R_{1{\cdot}23}$.

(ii) Estimate the value of X_1 when $X_2 = 6{\cdot}0$ and $X_3 = 650$.

Solution.

(i)
$$r_{23{\cdot}1} = \frac{r_{23} - r_{21} \cdot r_{31}}{\sqrt{1 - r_{21}^2}\sqrt{1 - r_{31}^2}}$$

Substituting the values, we get

$$r_{23{\cdot}1} = \frac{-0{\cdot}56 - (0{\cdot}80)\,.\,(-0{\cdot}40)}{\sqrt{1-(0{\cdot}80)^2}\,\sqrt{1-(-0{\cdot}40)^2}}$$

$$= \frac{-0{\cdot}56 + {\cdot}32}{\sqrt{1 - {\cdot}64}\,\sqrt{1 - {\cdot}16}} = \frac{-0{\cdot}24}{0{\cdot}6 \times 0{\cdot}916} = -0{\cdot}436$$

$$R_{1{\cdot}23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 + 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

Substituting values, we get

$$R_{1{\cdot}23} = \sqrt{\frac{(0{\cdot}80)^2 + (-0{\cdot}40)^2 - 2\,(0{\cdot}80)\,(-0{\cdot}40)\,(-0{\cdot}56)}{1-(-{\cdot}56)^2}}$$

$$= \sqrt{\frac{0{\cdot}64 + {\cdot}16 - {\cdot}3584}{1 - {\cdot}3136}} = \sqrt{\frac{0{\cdot}4416}{0{\cdot}6864}} = 0{\cdot}802$$

(ii) The linear regression equation of X_1 on X_2 and X_3 is :

$$X_1 - \overline{X}_1 = b_{12{\cdot}3}\,(X_2 - \overline{X}_2) + b_{13{\cdot}2}\,(X_3 - \overline{X}_3)$$

$$b_{12{\cdot}3} = \frac{S_1}{S_2} \times \left[\frac{r_{12} - r_{13} \cdot r_{23}}{1 - r_{23}^2}\right]$$

$$= \frac{4{\cdot}4}{1{\cdot}1} \cdot \left[\frac{0{\cdot}80 - 0{\cdot}224}{0{\cdot}6864}\right] = \frac{4{\cdot}4}{1{\cdot}1} \cdot \left[\frac{0{\cdot}576}{0{\cdot}6864}\right] = 3{\cdot}357$$

$$b_{13{\cdot}2} = \frac{S_1}{S_3} \cdot \left[\frac{r_{13} - r_{12} \cdot r_{32}}{1 - r_{32}^2}\right] = \frac{4{\cdot}4}{80} \cdot \left[\frac{-0{\cdot}40 - (0{\cdot}80)\,(-0{\cdot}56)}{1-(-{\cdot}56)^2}\right]$$

$$= \frac{4{\cdot}4}{80} \cdot \left[\frac{-0{\cdot}40 + 0{\cdot}448}{0{\cdot}6884}\right] = \frac{4{\cdot}4}{80} \cdot \left[\frac{0{\cdot}048}{0{\cdot}6864}\right] = 0{\cdot}0038 = 0{\cdot}004$$

Substituting the values in the equation, we get

$$X_1 - 28{\cdot}02 = 3{\cdot}357\,(X_2 - 4{\cdot}91) + {\cdot}004\,(X_3 - 594)$$

or
$$X_1 - 28{\cdot}02 = 3{\cdot}357\,X_2 - 16{\cdot}4828 + {\cdot}004\,X_3 - 2{\cdot}376$$

$$X_1 = 3{\cdot}357\,X_2 + {\cdot}004\,X_3 + 9{\cdot}1612$$

(iv) Estimation of X_1 for X_2 and X_3 :

When $X_2 = 6{\cdot}00, X_3 = 650$,
$$X_1 = 3{\cdot}357\,(6{\cdot}00) + {\cdot}004\,(650) + 9{\cdot}1612$$
$$= 20{\cdot}142 + 2{\cdot}6 + 9{\cdot}1612$$
$$= 31{\cdot}9032$$

IMPORTANT FORMULAE

Multiple Correlation Coefficients :

$$R_{1{\cdot}23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$R_{2{\cdot}13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{21} \cdot r_{23} \cdot r_{13}}{1 - r_{13}^2}}$$

$$R_{3{\cdot}12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{31} \cdot r_{32} \cdot r_{12}}{1 - r_{12}^2}}$$

Partial Correlation Coefficients :

$$r_{12{\cdot}3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}}$$

$$r_{23\cdot1} = \frac{r_{23} - r_{21} \cdot r_{31}}{\sqrt{1-r_{21}^2}\,\sqrt{1-r_{31}^2}}$$

$$r_{13\cdot2} = \frac{r_{13} - r_{12} \cdot r_{32}}{\sqrt{1-r_{12}^2}\,\sqrt{1-r_{32}^2}}$$

Relationship between Simple, Partial and Multiple Correlation Coefficients

$$1 - R_{1\cdot23}^2 = (1 - r_{12}^2)(1 - r_{13\cdot2}^2)$$

$$1 - R_{2\cdot13}^2 = (1 - r_{21}^2)(1 - r_{23\cdot1}^2) \quad \text{and}$$

$$1 - R_{3\cdot12}^2 = (1 - r_{21}^2)(1 - r_{32\cdot1}^2)$$

Multiple Regression of X_1 on X_2 and X_3 :

$$X_1 = a_{1\cdot23} + b_{12\cdot3} X_2 + b_{13\cdot2} X_3$$

Where,

$$b_{12\cdot3} = \left[\frac{\sigma_1}{\sigma_2}\right] \times \left[\frac{r_{12} - r_{13} \cdot r_{23}}{1 - r_{23}^2}\right]$$

Standard Error of Estimate :

$$S_{1\cdot23} = \sigma_1 \cdot \sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

QUESTIONS

1. Distinguish between partial correlation and multiple correlation.
2. Write down the expression for $r_{12\cdot3}$ and $R_{1\cdot23}$ in terms of r_{12}, r_{13} and r_{23}. Also state the limits within which $r_{12\cdot3}$ and $R_{1\cdot23}$ must lie.
3. Explain the concept of multiple regression and discuss its utility in business.
4. How will you fit a multiple regression equation of X_1 on X_2 and X_3 ?
5. Write the normal equations in case of multiple linear regression of X_1 on X_2 and X_3.
6. Write a short note on "Standard Error of Estiamte" for multiple regression.
7. Define simple, partial and multiple correlation coefficients and find their relationship with one another.
8. Write the equations/formulae to calculate the followings :
   (i) Regression equation for X_2 on X_1 and X_3.
   (ii) Partial regression coefficients ($b_{12\cdot3}$ and $b_{13\cdot2}$).
   (iii) Standard error of estimate of X_1 on X_2 and X_3 ($S_{1\cdot23}$)
   (iv) Multiple correlation coefficient ($R_{1\cdot23}$)
   (v) Partial correlation coefficient ($r_{12\cdot3}$)
9. How will you interpret the value of R^2 in a multiple regression equation ?

❀❀❀

2

Sampling & Sampling Distribution

INTRODUCTION

In all the spheres of life (such as Economic, Social and Business) the need for statistical investigation and data analysis is rising day by day. There are two methods of collection of statistical data : (i) **Census Method**, and (ii) **Sample Method.** Under census method, information relating to the entire field of investigation or units of population is collected; whereas under sample method, rather than collecting information about all the units of population, information relating to only selected units is collected. Before we make a detailed study of both the methods, we will explain some basic concepts related to them.

SOME BASIC CONCEPTS

(1) Universe or Population : In statistics, universe or population means an aggregate of items about which we obtain information. A universe or population means the entire field under investigation about which knowledge is sought. For example, if we want to collect information about the average monthly expenditure of all the 2,000 students of a college, then the entire aggregate of 2,000 students will be termed as Universe or Population. A population can be of two kinds (i) Finite and (ii) Infinite. In a finite population, number of items is definite such as, number of students or teachers in a college. On the other hand, an infinite population has infinite number of items *e.g.*, number of stars in the sky, number of water drops in an ocean, number of leaves on a tree or number of hairs on the head.

(2) Sample : A part of population is called sample. In other words, selected or sorted units from the population is known as a sample. In fact, a sample is that part of the population which we select for the purpose of investigation. For example, if an investigator selects 200 students from 2000 students of a college who represent all of them, then these 200 students will be termed as a sample. Thus, sample means some units selected out of a population which represent it.

CENSUS AND SAMPLE METHODS

There are two methods to collect statistical data :

(1) Census Method

(2) Sample Method

(1) Census Method

Census method is that method in which information or data is collected from each and every unit of the population relating to the problem under investigation and conclusions are drawn on their basis. This method is also called as **Complete Enumeration Method.** For example, suppose some information (like Monthly Expenditure, Average Height, Average Weight, etc.) is to be collected regarding 2000 students of a college. For that purpose if we collect data by inquiring each and every student of the college then this method will be called as Census method. In this example, the whole college *i.e.*, all 2000 students will be considered as a population and every student as an individual will be called the unit of the population. Population in India is conducted after every ten years by using census method.

Merits and Demerits of Census Method

Merits

(i) **Reliable and Accurate Data :** Data obtained by census method have more reliability and accuracy because in this method data are collected by contacting each and every unit of the universe.

(ii) **Extensive Information :** This method gives detailed information about each unit of the universe. For example, Indian population census does not only provide the knowledge about the number of persons but also information about their age, occupation, income, education, marital status, etc.

(iii) **Suitability :** This method is more suitable for the population with limited scope and diverse characteristics. Use of this method is also appropriate where intensive study is desired.

Demerits

(i) **More Expensive :** Census method is an expensive one. More money is needed for it as information is collected from each unit of the population. This is why this method is used by Government mostly for very important issues like Census, etc.

(ii) **More Time :** This method involves much time for data collection because data are collected from each and every unit of the population. This results in delay in making statistical inferences.

(iii) **More Labour :** This method of data collection also involves very much labour. For this the enumerators in a large number are required.

(iv) **Not Suitable for Specific Problems :** This method is not suitable relating to certain specific problems and infinite population. For example, if the population is infinite or items of the population are perishable or very complex type, then the census method is not suitable.

(2) Sampling Method

Sampling method is that method in which data is collected from the sample of items selected from population and conclusions are drawn from them. For example, if a study is to be made regarding the monthly expenditure of 2000 students of a college, then instead of collecting information from each student of the college, if we collect information by selecting some students like 100, then this will be called Sampling Method. On the basis of sampling method, it is possible to study the monthly expenditure of all the students of the college. Sampling method has three main stages (i) to select a sample (ii) to collect information from it and (ii) to make inferences regarding the population.

Importance of Sampling Method

In modern times sampling method is an important and popular method of statistical inquiry. Besides economic and business world, this method is widely used in daily life. For example, a housewife comes to know of the coating of the whole lot of rice by observing two-three grains only. A doctor tests the blood of a patient by examining one or two drops of blood only. In the same way, we learn about the quality of a commodity while buying the items of daily use like wheat, rice, pulses, etc. by observing the sample or specimen. In factories, statistical quality controller inspects the quality of items by examining a few units produced. A teacher gets the knowledge about the efficacy of his teaching by putting questions to a few students. In reality, there is scarcely any area left where sampling method is not used.

Merits and Demerits of Sampling Method

Merits

(i) **Saving of Time and Money :** Sampling method is less expensive. It saves money and labour because only a few units of the population are studied.

(ii) **Saving of Time :** In sampling method, data can be collected more quickly as these are obtained from some items of the universe. Thus much time is saved.

(iii) **Intensive Study :** As number of items is less in sampling method, they can be intensively studied.

(iv) **Organisational Convenience :** In this method, research work can be organised and executed more conveniently. More skilled and competent investigators can be appointed.

(v) **More Reliable Results :** If sample is selected in such a manner as it represents totally the universe, then the results derived from it will be more accurate and reliable.

(vi) **More Scientific :** Sampling method is more scientific because data can be inquired with other samples.

(vii) **Only Method :** In some fields where inquiry by census method is impossible, then in such situation, sampling method alone is more appropriate. If the population is infinite or too widespread or of perishable nature, then sampling method is used in such cases.

Demerits

(i) **Less Accurate :** Sampling method has less accuracy because rather than making inquiry about each unit of the universe, partial inquiry or inquiry relating to some selected units only is made.

(ii) **Wrong Conclusions :** If method of selecting a sample is not unbiased or proper caution has not been taken, then results are definitely misleading.

(iii) **Less Reliable :** Compared to census method, there is more likelihood of the bias of the investigator, which makes the results less reliable.

(iv) **Need of Specified Knowledge :** This is a complex method as specialised knowledge is required to select a sample.

(v) **Not Suitable :** If all units of a population are different from one another, then sampling method will not prove to be much useful.

Difference between Census and Sample Method

The main difference between the census method and the sampling method are as follows :

(i) **Scope :** In census method, all items relating to a universe are investigated whereas in sampling method only a few items are inquired.

(ii) **Cost :** Census method is expensive from the point of view of time, money and labour whereas Sampling method economises on them.

(iii) **Field of Investigation :** Census method isused in investigations with limited field whereas sampling method is used for investigations with large field.

(iv) **Homogeneity :** Census method is useful where units of the population are heterogeneous whereas sampling method proves more useful where population units are homogenous.

(v) **Type of Universe :** In such fields where study of each and every unit of the universe is necessary, census method is more appropriate. On the contrary, when population is infinite or vast or liable to be destroyed as a result of complete enumeration, then sampling method is considered to be more appropriate.

SAMPLING METHODS

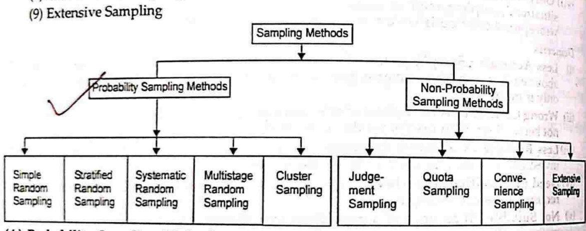
The method of selecting a sample out of a given population is called sampling. In other words, sampling denotes the selection of a part of the aggregate statistical material with a view of obtaining information about the whole. Now a days, there are various methods of selecting sample from a population in accordance with various needs.

(A) Probability Sampling Methods :

(1) Simple Random Sampling
(2) Stratified Random Sampling
(3) Systematic Random Sampling
(4) Multistage Random Sampling
(5) Cluster Sampling

(B) Non-Probability Sampling Methods :

(6) Judgement Sampling
(7) Quota Sampling
(8) Convenience Sampling
(9) Extensive Sampling



(A) Probability Sampling Methods

Probability sampling methods are such methods of selecting a sample from the population in which all units of the universe are given equal chances of being included in the sample.

There are various variants of probability sampling methods, which are given below :

(1) **Simple Random Sampling :** Simple random sampling is that method in which each item of the universe has an equal chance of being selected in the sample. Which item will included in the sample and which not, such decision is not made by an investigator on his will but selection of the units is left on chance. According to random sampling, there are two methods of selecting a random sample:

(i) **Lottery Method :** In this method, each unit of the population is named or numbered which is marked on separate piece of paper. Such chits are folded and put into some urn or bag. Thereafter as many chits are made selected by some person as many units are to be included in a sample.

(ii) **Tables of Random Numbers :** Some experts has constructed random number tables. These tables help in selection of a sample. Of all such various tables, Tippett's Tables

are most famous and are in use. Tippett has constructed a four-digit table of 10,400 numbers by using numbers as many as 41,600. In this method, first of all, all the items of a population are written serially. Thereafter by making use of Tippett's tables, in accordance with the size of the sample, numbers are selected. The selection of a sample with the help of Tippett's table can be made clear by an example :

An Extract of Tippett's Table

2952	6641	3992	9792	7979	5911
3170	5524	4167	9525	1545	1396
7203	4356	1300	2693	2370	7483
3408	2762	3563	6107	6913	7691
0560	5246	1112	9025	6008	8127

For example, suppose 12 units are to be chosen out of 5000 units. With Tippett's table, to decide such units, firstly 5000 units will be serially ordered from 1 to 5000 and then from Tippett's table, 12 numbers will be chosen from the beginning which are less than 5000. These 12 numbers are follows :

2952	4167	4356	2370
3992	1545	1300	3408
3170	1396	2693	2762

The items of such serial numbers will be included in the sample. If units of the population are less than 100, then 4 digit random numbers will be made compact into two digit numbers, and then such two digit numbers will be selected. Like as to select 6 units out of 60 units, the units with serial numbers 29, 39, 31, 41, 15 and 13 will be included in the sample.

Merits

(i) There is no possibility of personal prejudice in this method. In other words, this method is free from personal bias.
(ii) Under this method, every unit of the universe gets the equal chance of being selected.
(iii) The use of this method saves time, money and labour.

Demerits

(i) If sample size is small, then sample is not adequately represented.
(ii) If universe is very small, then this method is not suitable.
(iii) If some items of the universe are so important that their inclusion in the sample is very essential, then this method will not be appropriate.
(iv) This method will not be appropriate when population has units with diverse characteristics.

(2) **Stratified Random Sampling :** This method is used when units of the universe are heterogeneous rather than homogeneous. Under this method, first of all units of the population are divided into different *strata* in accordance with their characteristics. Thereafter by using random sampling, sample items are selected from each stratum. For example, if 150 students are to be selected out of 1500 students of a college, then firstly the college students will be divided into three groups on the basis of Arts, Commerce and Science. Suppose there are 500, 700, 300 students respectively in three faculties and 10% sample is to be taken, then on the basis of random sampling 50, 70 and 30 students respectively will be selected by using random sampling. Thus, this method assumes equal representation to each class or group and all the units of the universe get equal chance of being selected in the sample.

Merits

(i) There is more likelihood of representation of units in this method.

(ii) Comparative study on the basis of facts at different strata is possible under this method.

(iii) This method has more accuracy.

Demerits

(i) This method has limited scope because this method can be adopted only when the population and its different strata are known.

(ii) There can be the possibility of prejudice if the population is not properly stratified.

(iii) If the population is too small in size, it is difficult to stratify it.

(3) Systematic Random Sampling: In this method, all the items of the universe are systematically arranged and numbered and then sample units are selected at equal intervals. For example, if 5 out of 50 students are to be selected for a sample, then 50 students would be numbered and systematically arranged. One item of the first 10 would be selected at random. Subsequently, every 10th item from the selected number will be selected to frame a sample. If the first selected number is 5th item, then the subsequent numbers would be 15th, 25th, 35th and 45th.

Merits

(i) It is a simple method. Samples can be easily obtained by it.

(ii) This method involves very little time in sample selection and results are almost accurate.

Demerits

(i) In this method, each unit does not stand the equal chances of being selected because only the first unit is selected on random sampling basis.

(ii) If all the units are different in characteristics, then results will not be appropriate.

(4) Multistage Random Sampling : When sampling procedure passes through many stages, then it is known as multi-stage sampling. In this method, firstly the entire universe or population is divided into stages or substages. From the each stage some units are selected on random sampling basis. Thereafter these units are subdivided and on the basis of random sampling again some sub-units are selected. Thus, this goes on with sub-division further and selection on. For example, for the purpose of a study regarding Adult Education in Haryana State, first some districts will be selected on random basis. Thereafter out of the selected districts, some tehsils and out of tehsils, some villages or towns may be thus selected, further out of the villages or towns, some neighbourhood, or wards and out of the wards, some households will be selected from whom the inquiry will be made concerning the problem at hand.

Merits

(i) This is the best method of studying a universe or population on regional basis.

(ii) This method is suitable for those problems where decisions on the basis of sample alone can not be taken.

Demerits

(i) This method requires many tests to correctly estimate the level of accuracy which involves a lot of time and labour.

(ii) In this method, level of estimated accuracy level is predecided which does not seem logical.

(5) Cluster Sampling : In this method, simply the universe is divided into many groups called cluster and out of which a few clusters are selected on random basis and then the clusters are complete enumerated. This method is usually applied in industries like as in pharmaceutical

industry, a machine produces medicines tablets in the batches of hundred each, then for quality inspection, a few randomly selected batches are examined.

(B) Non-Probability Sampling Methods

Non-probability sampling methods are those methods in which selection of units is made on the basis of convenience or judgement of the investigator rather than on the basis of probability or chance. In such methods, selection of units is made in accordance with the specific objectives and convenience of the investigator.

(6) Judgement Sampling : Under this method, the selection of the sample items depends exclusively on the judgement of the investigator. In other words, the investigator exercises his judgement in the choice and includes those items in the sample which he thinks are most typical of the universe with regard to the characteristics under study. For example, if a sample of 20 students is to be selected from a class of 80 students for analysing the spending habits of the 10 students, the investigator would select 20 students, who in his opinion are representative of the class.

Merits

(i) This method is less expensive.

(ii) This method is very simple and easy.

(iii) This method is useful in those fields where almost similar units exist or some units are too important to be left out of the sample.

Demerits

(i) There is greater chance of the investigator's own prejudice in this method.

(ii) This method is not very accurate and reliable.

(7) Quota Sampling : In this method, the investigators are assigned definite quotas according to some criteria. They are instructed to obtain the required number to fill in each quota. The investigators select the individuals (*i.e.,* sample items) to collect information on their personal judgements within the quotas. When all or a part of the whole quota is not available or approachable, the quota is completed by supplementing new responds. Quota sampling is a type of judgement sampling.

Merits

(i) In this method, there is greater chance of important units being included.

(ii) Statistical inquiry is more organised in this method on account of the units of the quotas being fixed.

Demerits

(i) Possibility of prejudice shall remain.

(ii) There is greater likelihood of sampling error in this method.

(8) Convenience Sampling : In this type of non-probability sampling, the choice of the sample is left completely to the convenience of the investigator. The investigator obtain a sample according to his convenience. For example, a book publisher selects some teachers conveniently on the basis of the list of the teachers from the college prospectus and gets feedback from them regarding his publication. This method is less expensive and more simple but is unscientific and unreliable. This method results in more dependence on the enumerators. This method is appropriate for sample selection where the universe or population is not clearly defined or list of the units is not available or sample units are not clear in themselves.

(9) Extensive Sampling : In this method, sample size is taken almost as big as the population itself like 90% the section of the population. Only those units are left out for which data collection is very difficult or almost impossible. Due to very large sample size, the method has greater level of accuracy. Intensive study of the problem becomes possible but this method involves heavy resources at disposal.

SAMPLING AND NON-SAMPLING ERRORS

The choice of a sample though may be made with utmost care, involves certain errors which may be classified into two types : (i) **Sampling Errors**, and (2) **Non-Sampling Errors.** These errors may occur in the collection, processing and analysis of data.

(1) Sampling Errors

Sampling errors are those which arise due to the method of sampling. Sampling errors arise primarly due to the following reasons:

(1) Faulty selection of the sampling method.
(2) Substituting one sample for the sample due to the difficulties in collecting the sample.
(3) Faulty demarcation of sampling units.
(4) Variability of the population which has different characteristics.

(2) Non-Sampling Errors

Non-sampling errors are those which creep in due to human factors which always varies from one investigator to another. These errors arise due to any of the following factors :

(1) Faulty planning.
(2) Faulty selection of the sample units.
(3) Lack of trained and experienced staff which collect the data.
(4) Negligence and non-response on the part of the respondent.
(5) Errors in compilation.
(6) Errors due to wrong statistical measures.
(7) Framing of a wrong questionnaire.
(8) Incomplete investigation of the sample survey.

Basic Concepts of Sampling

Sampling Distribution: The purpose of selecting and studying a sample from the population is to estimate or make inference about some population characteristics. In this process, the knowledge of the sampling distribution is of vital importance.

Some Important Terms:

The following terms are widely used in the study of the sampling distribution:

(1) Parameters: Any statistical measures computed from the population data is known as parameter. Thus, population mean, population standard deviation, population variance, population proportion, etc., are all parameters. Parameters are denoted by the Greek letters such as μ, σ^2, σ and P.

(2) Statistics: Any statistical measure computed from sample data is known as statistic. Thus, sample mean, sample standard deviation, sample variance, sample proportion, etc., are all statistics. Statistics are denoted by Roman letters such as $\overline{X}$, s, s^2 and p.

(3) Sampling with and without replacement: Sampling is a procedure of selecting a sample from the population. Sampling may be done with or without replacement. Sampling where each unit of a population may be chosen more than once is called sampling with replacement. If each unit cannot be chosen more than once, it is called sampling without replacement. In case of

sampling with replacement, the total number of possible samples each of size n drawn from a population of size N is N^n. But if the sampling is without replacement, the total number of possible samples will be $N_{c_n} = m$ (say).

Sampling Distribution of a Statistic

Sampling distribution constitutes the theoretical basis of statistical inference and is of considerable importance in business decision making. Sampling distribution of a statistic is the frequency distribution which is formed with various values of a statistic computed from different samples of the same size drawn from the same population. Suppose we draw all possible samples of size n from the population (N) with or without replacement. For each possible sample drawn from the population, we compute a statistic such as mean, median, standard deviation, variance, etc. The set of all possible values of a statistic is then classified and grouped into a frequency distribution (or probability distribution). The distribution so obtained is called the sampling distribution of a statistic. We could have various sampling distribution depending upon the nature of the statistic we have computed. If, for instance, the particular statistic computed is the sample mean, the distribution is called sampling distribution of mean. If, we compute variance of each sample, then it is called the sampling distribution of variance. Similarly, we could have sampling distributions of proportion, median, standard deviation, etc.

An Important Property of Sampling Distribution

An important property of the sampling distribution of a statistic is that if random samples of large size ($n > 30$) are taken from a population which may be normally distributed or not, then the sampling distribution of the statistic will approach a normal distribution.

Standard Error of a Statistic

The standard deviation of the sampling distribution of a statistic is known as the standard error of a statistic. As there are various types of sampling distributions, we could have various types of standard errors depending on the nature of sampling distribution. The standard deviation of the sampling distribution of means is called the standard error of the means. In sampling theory, instead of using the term standard deviation for measuring variation, we use a new term called standard error of mean. The standard error of mean measure the extent to which the sample mean differ from the population mean. Thus, **the basic difference between the standard deviation and standard error of mean is that the former measures the extent to which the individual items differ from the central value and the latter measures the extent to which individual sample mean differ from the population mean.** Like the standard error of the means, we could have standard error of the median, standard deviation, proportion, variance, etc.

Utility of Standard Error : The standard error is used in a large number of problems which are discussed as follows :

(1) Reliability of a Sample : The standard error gives an idea about the reliability and precision of a sample. That is, it indicates how much the estimated value differs from the observed values. The greater the standard error, the greater is the deviation between the estimated and observed values and lesser is the reliability of a sample. The smaller the standard error, the smaller is the deviation between the estimated and observed values and greater is the reliability of a sample.

(2) Tests of Significance : The standard error is also used to test the significance of the various results obtained from small and large samples. In case of large sample, if the difference between the observed and the expected value is greater than 1.96 standard error, then we reject the hypothesis at 5% and conclude that sample differs widely from the population. But if the difference between the observed and the expected value is greater than 2.58 S.E. (Standard error), then we reject the null hypothesis at 1% and conclude that the sample differs widely from the population.

(3) To determine the confidence limits of the unknown population mean : The standard error enables us in determining the confidence limits within which a population parameter is expected to lie with a certain degree of confidence. The confidence limits of the unknown population mean μ are given by.

Large Sample	Small Sample
95% confidence limits for μ	95% confidence limit for μ
$\bar{x} - 1.96$ S.E. and $\bar{x} + 1.96$ S.E.	$\bar{x} \pm t_{.05}$ S.E.
99% confidence limits for μ	99% confidence limits for μ
$\bar{x} - 2.58$ S.E. and $\bar{x} + 2.58$ S.E.	$\bar{x} \pm t_{01}$ S.E.

SAMPLING DISTRIBUTION OF MEANS

It is an important sampling distribution widely used in the sampling theory. Draw all possible samples of size n with or without replacement from population of size N with mean μ and variance σ^2. For each possible sample drawn from the population, we compute the mean $\bar{x}$ of each sample. The mean will vary from sample to sample. The set of all possible means obtained from different samples is called the sampling distribution of means.

Properties : The following are the important properties of the sampling distribution of means:

(i) The mean of the sampling distribution of means is equal to the population mean (μ). Symbolically,

$$\mu_{\bar{x}} = \mu \quad \text{or} \quad E(\bar{x}) = \mu$$

This property can be proved as follows :

Let $x_1, x_2, ..., x_n$ represent a random sample (with replacement) of size n from a finite population of size N having its mean μ and variance σ^2, then

$$\bar{x} = \frac{x_1 + x_2 + ... x_n}{n}$$

$$E(\bar{x}) = E\left[\frac{\Sigma x}{n}\right] = E\left[\frac{x_1 + x_2 + ... + x_n}{n}\right]$$

$$= \frac{1}{n}\{E(x_1) + E(x_2) + ... + E(x_n)\}$$

$$= \frac{1}{n}\{\mu + \mu + \mu + ... + \mu\} = \frac{1}{n} \cdot n\,\mu = \mu$$

Thus, the mean of the sampling distribution of means is equal to the population mean.

(2) The standard error of the sampling distribution of means is obtained as :

$$\text{S.E.}_{\bar{x}} \text{ or } \sigma_{\bar{x}} = \frac{\text{S.D. of Population}}{\sqrt{\text{Size of the sample}}} = \frac{\sigma}{\sqrt{n}}$$

This property can be proved as follows :

$$\text{Var}(\bar{x}) = \text{Var}\left(\frac{\Sigma x}{n}\right) = \text{Var}\left(\frac{x_1 + x_2 + ... x_n}{n}\right)$$

$$= \frac{1}{n^2}[\text{Var}(x_1) + \text{Var}(x_2) + ... \text{Var}(x_n)]$$

$$= \frac{1}{n^2}[\sigma^2 + \sigma^2 + ... \sigma^2]$$

$$= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

where, σ^2 is the population variance, x is the sample.

Because $n > 1$, obviously, $\frac{\sigma^2}{n} < \sigma^2 \Rightarrow V(\bar{x}) <$ Population variance.

$$\therefore \quad \text{S.E.}_{\bar{x}} = \sigma_{\bar{x}} = \sqrt{\text{Var}(\bar{x})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

This formula holds only when sampling is with replacement.

Note : When the population is finite and the samples are drawn without replacement, then $\text{S.E.}_{\bar{x}}$ is obtained as :

$$\text{S.E.}_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

(3) The sampling distribution of means is approximately a normal distribution with mean μ and variance σ^2, provided the sample is large ($n > 30$).

(4) The following formula is used to find the probability of the sampling distribution of means.

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

Let us illustrate the concept of sampling distribution of means by the following example :

Example 1. **Consider a population consisting of three values : 2, 5 and 8. Draw all possible samples of size 2 with replacement from the population. Construct sampling distribution of means. Also find the mean and standard error of the distribution.**

Solution. The population consists of three values. The total number of possible samples of size 2 drawn with replacement are $N^n = 3^2 = 9$. All possible random samples and their sample mean are shown is the following table :

Sample No.	Sample Values	Sample Mean $\bar{x}$
1.	(2, 2)	$\frac{1}{2}(2+2) = 2$
2.	(5, 2)	$\frac{1}{2}(5+2) = 3.5$
3.	(8, 2)	$\frac{1}{2}(8+2) = 5$
4.	(2, 5)	$\frac{1}{2}(2+5) = 3.5$
5.	(5, 5)	$\frac{1}{2}(5+5) = 5.0$
6.	(8, 5)	$\frac{1}{2}(8+5) = 6.5$

7.	(2, 8)	$\frac{1}{2}(2+8)=5.0$
8.	(5, 8)	$\frac{1}{2}(5+8)=6.5$
9.	(8, 8)	$\frac{1}{2}(8+8)=8.0$

On the basis of the means ($\bar{x}$) of all the 6 possible samples, the sampling distribution of means is given below :

Sample Means ($\bar{x}$)	f	$f\bar{x}$	$d=\bar{x}-\mu_{\bar{x}}$	d^2	fd^2
2	1	2	− 3	9	9
3.5	2	7	− 1.5	2.25	4.50
5.0	3	15	0	0	0
6.5	2	13	1.5	2.25	4.50
8.0	1	8	+ 3	9.0	9.0
	$\Sigma f=9$	$\Sigma f\bar{x}=45$			$\Sigma fd^2=27$

Mean of the Sampling Distribution of Means

$$\mu_{\bar{x}}=\frac{\Sigma f\bar{x}}{\Sigma f}=\frac{45}{9}=5$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x})=\frac{\Sigma f(\bar{x}-\mu_{\bar{x}})^2}{\Sigma f}=\frac{\Sigma fd^2}{\Sigma f}=\frac{27}{9}=3$$

Hence, $\text{S.E}_{\bar{x}}=\sigma_{\bar{x}}=\sqrt{3}=1.732$

Aliter : The sampling distribution of means can also be written in terms of probability as :

Sample Means ($\bar{x}$)	2	3.5	5.0	6.5	8.0
Probability (p)	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Since 3.5 occurs twice, its probability of occurence is $\frac{2}{9}$, 5 occurs thrice, its probability of occurrence is $\frac{3}{9}$ and 6.5 occurrence is $\frac{2}{9}$. Each of the other sample mean occurs only once with probability $\frac{1}{9}$.

Mean of the Sampling Distribution of Means

$$E(\bar{x})=\Sigma p\bar{x}=2\times\frac{1}{9}+3.5\times\frac{2}{9}+5\times\frac{3}{9}+6.5\times\frac{2}{9}\times 8.0\times\frac{1}{9}$$

$$=\frac{1}{9}\cdot[2+7+15+13+8]=\frac{45}{9}=5$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x})=E(\bar{x}^2)-[E(\bar{x})]^2$$

$$E(\bar{x}^2)=2^2\times\frac{1}{9}+3.5^2\times\frac{2}{9}+5^2\times\frac{3}{9}+6.5^2\times\frac{2}{9}+8^2\times\frac{1}{9}$$

$$=\frac{1}{9}\cdot[4+25+75+84.5+64]$$

$$=\frac{1}{9}[252.5]=28.055$$

$$\text{Var}(\bar{x})=E(\bar{x}^2)-[E(\bar{x})^2]=28.055-25=3.055=3$$

Hence, $\text{S.E.}_x=\sqrt{\text{Var}(\bar{x})}=\sqrt{3}=1.732$

Example 2. **Construct a sampling distribution of the sample means from the following population :**

Population Unit :	**1**	**2**	**3**	**4**
Observation :	**22**	**24**	**26**	**28**

when random sample of size 2 are taken from it without replacement. Also find the mean and standard error of the distribution.

Solution. The population consists of four values (22, 24, 26, 28). The total number of possible sample of size 2 drawn without replacement are ${}^4C_2=6$. All the possible random samples and their sample means are shown in the table given below :

Sample No.	Sample Values	Sample Mean $\bar{x}$
1.	(22, 24)	$\frac{1}{2}(22+24)=23$
2.	(22, 26)	$\frac{1}{2}(22+26)=24$
3.	(22, 28)	$\frac{1}{2}(22+28)=25$
4.	(24, 26)	$\frac{1}{2}(24+26)=25$
5.	(24, 28)	$\frac{1}{2}(24+28)=26$
6.	(26, 28)	$\frac{1}{2}(26+28)=27$

On the basis of the means ($\bar{x}$) of all the 6 samples without replacement, the sampling distribution of mean is given below:

Sampling Distribution of Means without Replacement

Sample Means ($\bar{x}$)	f	$f\bar{x}$	$d=\bar{x}-\mu_{\bar{x}}$	d^2	fd^2
23	1	23	− 2	4	4
24	1	24	− 1	1	1
25	2	50	0	0	0
26	1	26	1	1	1
27	1	27	2	4	4
	$\Sigma f=6$	$\Sigma f\bar{x}=150$			$\Sigma fd^2=10$

Mean of the Sampling Distribution of Means

$$\mu_{\bar{x}} = \frac{\Sigma f\bar{x}}{\Sigma f} = \frac{150}{6} = 25$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x}) = \frac{\Sigma f(\bar{x}-\mu_x)^2}{\Sigma f} = \frac{\Sigma fd^2}{\Sigma f} = \frac{10}{6} = \frac{5}{3}$$

Hence, $\quad S.E._{\bar{x}} = \sigma_x = \sqrt{\text{Var}\,\bar{x}} = \sqrt{\frac{5}{3}} = 1.29$

Aliter : The sampling distribution of means can also be written in terms of probability as below :

Sample Means ($\bar{x}$)	23	24	25	26	27
Probability (p)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Since, 25 occurs twice, its probability of occurence is $\frac{2}{6}$. Each of the other sample means occurs only once with probability $\frac{1}{6}$.

Mean of the Sampling Distribution of Means

$$E(\bar{x}) = \Sigma p\bar{x} = \frac{1}{6}\times 23 + \frac{1}{6}\times 24 + \frac{2}{6}\times 25 + \frac{1}{6}\times 26 + \frac{1}{6}\times 27$$

$$= \frac{1}{6}\cdot[23+24+50+26+27] = \frac{150}{6} = 25$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})]^2$$

$$E(\bar{x}^2) = 23^2\times\frac{1}{6} + 24^2\times\frac{1}{6} + 25^2\times\frac{2}{6} + 26^2\times\frac{1}{6} + 27^2\times\frac{1}{6}$$

$$= \frac{1}{6}\cdot[529+576+1250+676+729]$$

$$= \frac{3760}{6} = 626.17$$

$$\text{Var}(\bar{x}) = E(\bar{x}^2) = [E(\bar{x})]^2 = 626.17 - 625 = 1.67$$

Hence, $\quad S.E._{\bar{x}} = \sigma_{\bar{x}}\ \sqrt{\text{Var}(\bar{x})} = \sqrt{1.67} = 1.29$

Example 3. **A population consists of four elements : 3, 7, 11, 15. Consider all possible samples of size two which can be drawn with replacement from this population. Find (i) the population mean μ (ii) the population variance σ^2 (iii) the mean of the sampling distribution of means (iv) standard error (or S.D.) of the sampling distribution of means. Verify (iii) and (iv) by using (i) and (ii) and by use of suitable formula.**

Solution. (i) μ = population mean $= \frac{\Sigma X}{N} = \frac{3+7+11+15}{4} = \frac{36}{4} = 9$

(ii) σ^2 = population variance $= \frac{\Sigma(X-\mu)^2}{N} = \frac{(-6)^2+(-2)^2+(2)^2+(6)^2}{4} = \frac{80}{4} = 20$

$\therefore \quad \sigma = S.D. = \sqrt{20}$

(iii) All possible random samples of size two with replacement is $N^n = 4^2 = 16$ and their sample means are shown in the following table :

Sample No.	Sample Values	Sample Mean $\bar{x}$	Sample No.	Sample Values	Sample Mean $\bar{x}$
1.	(3, 3)	3	9	(11, 3)	7
2.	(3, 7)	5	10	(11, 7)	9
3.	(3, 11)	7	11	(11, 11)	11
4.	(3, 15)	9	12	(11, 15)	13
5.	(7, 3)	5	13	(15, 3)	9
6.	(7, 7)	7	14	(15, 7)	11
7.	(7, 11)	9	15	(15, 11)	13
8.	(7, 15)	11	16	(15, 15)	15

On the basis of the mean ($\bar{x}$) of all the 16 samples with replacement, the sampling distribution of $\bar{x}$ can be written as :

Sample Means ($\bar{x}$)	Frequency (f)	$f\bar{x}$	$d = \bar{x} - \mu_{\bar{x}}$	d^2	fd^2
3	1	3	− 6	36	36
5	2	10	− 4	16	32
7	3	21	− 2	4	12
9	4	36	0	0	0
11	3	33	+ 2	4	12
13	2	26	+ 4	16	32
15	1	15	+ 6	36	36
	$\Sigma f = 16$	$\Sigma f\bar{x} = 144$			$\Sigma fd^2 = 160$

Mean of the Sampling Distribution of Means

$$\mu_{\bar{x}} = \frac{\Sigma f\bar{x}}{\Sigma f} = \frac{144}{16} = 9$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x}) = \frac{\Sigma f(\bar{x}-\mu_{\bar{x}})^2}{\Sigma f} = \frac{160}{16} = 10$$

Hence $\quad S.E._{\bar{x}} = \sigma_{\bar{x}} = \sqrt{\text{Var}(\bar{x})} = \sqrt{10}$ = 3.16

Using the formula, $\mu_{\bar{x}} = \mu$ and $V(\bar{x}) = \frac{\sigma^2}{n}$, we get the mean of the sampling distribution of means $= \mu_{\bar{x}} = \mu = 9$ and variance of the sampling distribution of means $= \frac{\sigma^2}{n} = \frac{20}{2} = \frac{\sigma^2}{2} = 10.$

Hence, the results of (iii) and (iv) are verified by using the results of (i) and (ii).

Example 4. A population consists of the following elements :

2, 4, 5, 8, 11

Find:

(a) How many different samples of size 3 are possible when sampling is done without replacement.

(b) List all of the possible different samples.

(c) Compute the mean of each of the samples given in part (b).

(d) Find the sampling distribution of sample mean $\overline{X}$.

(e) If all the elements are euqally likely, compute the value of the population mean μ.

Solution. The population consists of five elements (2, 4, 5, 8, 11).

(a) The total number of possible samples of size 3 drawn without replacement are $5_{c_3} = 10$.

(b) All the possible different samples and their sample means are shown in the following table.

Sample No.	Sample Values	Sample Mean $\bar{x}$
1.	(2, 4, 5)	$\frac{1}{3}(2+4+5) = 3.67$
2.	(2, 4, 8)	$\frac{1}{3}(2+4+8) = 4.67$
3.	(2, 4, 11)	$\frac{1}{3}(2+4+11) = 5.67$
4.	(2, 5, 8)	$\frac{1}{3}(2+5+8) = 5.0$
5.	(2, 5, 11)	$\frac{1}{3}(2+5+11) = 6.0$
6.	(2, 8, 11)	$\frac{1}{3}(2+8+11) = 7.0$
7.	(4, 5, 8)	$\frac{1}{3}(4+5+8) = 5.67$
8.	(4, 5, 11)	$\frac{1}{3}(4+5+11) = 6.67$
9.	(4, 8, 11)	$\frac{1}{3}(4+8+11) = 7.67$
10.	(5, 8, 11)	$\frac{1}{3}(5+8+11) = 8.0$

In the above table, we have 10 possible samples of size 3 without replacement. Since, 5.67 occurs twice, its probability of occurrence is $\frac{2}{10}$. Each of the other sample means occur only once with probability $\frac{1}{10}$.

Sampling distribution of means $\bar{x}$ (*i.e.*, the probability distribution of sample mean $\bar{x}$) is given below :

Sampling Distribution of $\bar{x}$

Sample Mean $\bar{x}$	3.67	4.67	5	5.67	6	6.67	6	7.67	8.0
Probability (p)	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

(e) Population consists of the values (2, 4, 5, 8, 11). Since, each value occurs equally likely, the probability of occurence of each value is $\frac{1}{5}$. Hence,

Population Values (X) :	2	4	5	8	11
Probability (p) :	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

$$\text{Population Mean} = \mu = 2\times\frac{1}{5} + 4\times\frac{1}{5} + 5\times\frac{1}{5} + 8\times\frac{1}{5} + 11\times\frac{1}{5}$$

$$= \frac{1}{5}\cdot[2+4+5+8+11] = \frac{30}{5} = 6$$

LAW OF LARGE NUMBERS AND CENTRAL LIMIT THEOREM

Law of Large Numbers and the Central Limit Theorem both serve the basis for the development of sampling distribution of a statistic.

Law of Large Numbers : The law of large numbers states that as the sample size increases, the sample mean would be closer and closer to the population mean. It does not gurantee that if the sample size is increased sufficiently, the sample mean will be equal to the population mean. There are two implications of the law of large numbers (i) the difference between sample mean and population mean can be reduced by increasing the sample size, and (ii) variation from one sample mean to another sample mean (of the same size) also decreases as the size of the sample increases.

Central Limit Theorem : It is widely used in the field of estimation and inference. This theorem states that if we select random sample of large size n from any population with mean μ and standard deviation σ and compute the mean of each sample, then the sampling distribution of mean $\bar{x}$ approaches normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. This is true even if the population itself is not normal. The utility of this theorem is that it requires virtualy no conditions on the distribution pattern of the population.

QUESTIONS

1. Distinguish between population and sample. Discuss the relative merits and demerits of census and sampling Methods.
2. Explain the various methods of sampling. Also discuss their relative merits and demerits.
3. Explain the simple random sampling technique and state when it is used.
4. (a) What is a random sample ? Discuss the various methods of drawing a random sample.

   (b) Distinguish between sampling and non-sampling errors.

5. What is meant by sampling distribution of a statistic? Also, define standard error of a statistic.

OR

Discuss briefly the concept of sampling distribution of an estimator.

6. Distinguish between :
   (i) Population and Sample (ii) Parameters and Statistics
   (iii) Sampling with and without replacement.
7. Write a note on "Sampling Distribution of Means."
8. A population consists of five numbers (2, 3, 6, 8, 11). Draw all possible random samples of size 2 which can be drawn with replacement from this population. Construct a sampling distribution of means. Also, find the mean and standard error of the distribution.
9. A population consists of four numbers (3, 7, 11, 15). Consider all possible samples of size 2 which can be drawn without replacement from this population. Construct the sampling distribution of means. Also find the mean and standard error of the distribution.
10. A population consists of the following five elements :

    3, 5, 9, 11, 17

    Find :
    (a) How many different samples of size 3 are possible when sampling is done without replacement ?
    (b) List all of the possible different samples.
    (c) Compute the sample mean for each of the samples given in part (b).
    (d) Find the sampling distribution of the sample mean $\bar{x}$. Use a probability histogram to graph the sampling distribution of $\bar{x}$.
    (e) If all five population values are equally likely, compute the value of the population mean, μ.
11. A population consists of numbers : 2, 3, 6, 8, 11. (i) Enumerate all possible samples of size 2 which can be drawn from this population (without replacement) (ii) Calculate the mean of the sampling distribution of means and show that the mean of the sampling distribution of means is equal to the population mean. (iii) Calculate the variance of the sampling distribution of means and show that it is less than that population variance.
12. (a) Show that the mean of the sampling distribution of means is equal to the population mean *i.e.*, $E(\bar{x}) = \mu$.

    (b) Derive the variance of the sampling distribution of the sample mean. Is it more than population variance ?
13. Give statements of Law of Large Numbers and Central Limit Theorem. Discuss their significance in sampling theory.

❀❀❀

3

Tests of Hypothesis – Large Sample Tests

INTRODUCTION

The main objective of the sampling theory is the study of the **Tests of Hypothesis or Tests of Significance.** In many circumstances, we are to make decisions about the population on the basis of only sample information. For example, on the basis of sample data, (i) a quality control manager is to determine whether a process is working properly, (ii) a drug chemist is to decide whether a new drug is really effective in curing a disease, (iii) a statistician has to decide whether a given coin is unbiased, etc. Such decisions are called statistical decisions (a simply decisions). The theory of testing of hypothesis employs various statistical techniques to arrive at such decisions on the basis of the sample study.

BASIC CONCEPTS OF HYPOTHESIS TESTING

The following basic concepts are used in the study of tests of hypothesis:

(1) Hypothesis (or Statistical Hypothesis) : In attempting to arrive at decisions about the population on the basis of sample information, it is necessary to make assumptions about the population parameters involved. Such an assumption (or statement) is called a statistical hypothesis which may or may not be true.

There are two types of hypothesis:

(a) Null Hypothesis **and** **(b) Alternative Hypothesis.**

(a) Null Hypothesis : In tests of hypothesis we always begin with an assumption or hypothesis (*i.e.,* assumed value of a population parameter). This is called Null Hypothesis. The null hypothesis asserts that there is no (significant) difference between the sample statistic and the population parameter and whatever the observed difference is there, is merely due to fluctuations in sampling from the same population. Null hypothesis is usually denoted by the symbol H_0. **R.A. Fisher** defined null hypothesis as "the hypothesis which is tested for possible rejection under the assumption that it is true". In other words, the hypothesis (regarding some characteristic of population) which is to be verified with the help of a random sample or the hypothesis which is under test is called null hypothesis. For example, if we want to test the hypothesis that the mean of the population to be taken as μ_0, then the null hypothesis (H_0) is $\mu = \mu_0$.

(b) Alternative Hypothesis : Any hypothesis different from the null hypothesis (H_0) is called an alternative hypothesis and is denoted by the symbol H_1. The two hypothesis H_0 and H_1 are such that if one is accepted, the other is rejected and vice versa. For example, if we want to test whether the population mean μ has a specified value μ_0, then (i) Null Hypothesis is $H_0 : \mu = \mu_0$ and

(ii) Alternative Hypothesis may be (a) $H_1 : \mu \neq \mu_0$ (*i.e.*, $\mu > \mu_0$ or $\mu < \mu_0$), or (b) $H_1 : \mu > \mu_0$ or (c) $H_1 : \mu < \mu_0$. Thus, there can be more than one alternative hypothesis.

(2) Type I and Type II Errors : In the process of hypothesis testing we usually come across same sort of errors, called **errors in hypothesis testing** which are grouped in two types as:

(i) Type I Errors and (ii) Type II Errors.

(i) Type I Errors : Type I errors are made when we reject the null hypothesis H_0 though it is true. In other words, when H_0 is rejected despite its being true, then it is called Type I errors. The probability of making a type I error is denoted by $p(E_1) = \alpha$ and the probability of making a correct decision is then $1 - \alpha$ *i.e.*, $p = 1 - \alpha$.

(ii) Type II Errors : Type II errors are made when we accept the null hypothesis though it is false. In other words, when H_0 is accepted despite its being false, then it s called Type II errors. The probability of making a type II error is denoted by β. Thus, $p(E_2) = \beta$.

The following table illustrates Type I and Type II errors.

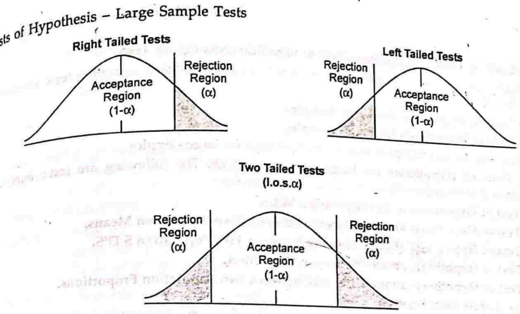
	To Accept H_0	**To Reject H_0**
H_0 is True	Correct Decision $p = 1 - \alpha$	Type I Error, $p = \alpha$
H_0 is False	Type II Error $p = \beta$	Correct Decision $p = 1 - \beta$

While testing hypothesis, attempts are made to minimise both the types of errors, although it is not at all possible to reduce them both at the same time.

(3) Level of Significance : This refers to the degree of significance with which we accept or reject a particular hypothesis. Since 100% accuracy is not possible in taking a decision over the acceptance or rejection of a hypothesis, we have to take the decision at a particular level of confidence which would speak of the probability of one being correct or wrong in accepting or rejecting a hypothesis. In most of the cases of hypothesis testing, such a confidence is fixed at 5% level, which implies that our decision would be correct to the extent of 95%. For a greater precise, however, such a confidence may be fixed at 1% level which would imply that the decision would be correct to the extent of 99%. This level is usually denoted by the symbol, α (alpha) which represents the probability of committing the type I error (*i.e.* rejecting a null hypothesis which is true). The level of confidence (or significance), is always fixed in advance before applying the test procedures. It is important to note that if no level of significance is given, then we always take $\alpha = 0.05$.

(4) Critical Region or Rejection Region : The critical region or rejection region is the region of the standard normal curve corresponding to a pre-determined level of significance. The region under the normal curve which is not covered by the rejection region is known as Acceptance Region. Thus, the statistic which leads to the rejection of null hypothesis H_0 gives us a region known as Rejection Region or Critical Region. While those which lead to acceptance of H_0 give us a region called as Acceptance Region.

(5) One Tailed Test and Two Tailed Test : A test of any statistical hypothesis where the alternative hypothesis is expressed by the symbol (<) or the symbol (>) is called a one tailed test since the entire critical region lies in one tail of the distribution of the test statistic. The critical region for all alternative hypothesis containing the symbol (>) lies entirely on the right tail of the distribution while the critical region for an alternative hypothesis containing a less than (<) symbol lies entirely in the left tail. The symbol indicates the direction where the critical region lies. A test of any statistical hypothesis where the alternative is written with a symbol '≠' is called a two-tailed



test, since the critical region is split into two equal parts, one in each tail of the distribution of the test statistic. The following figures illustrate one tailed and two tailed tests :

(6) Critical Value : The critical values of the standard normal variate (Z) for both the two-tailed and one tailed tests at different level of significance are very often required in hypothesis testing. The following table gives critical values for both one tailed and two tailed tests at various level of significance.

Level of Significance (α)	$\alpha = 0{\cdot}10$	$\alpha = 0{\cdot}05$	$\alpha = {\cdot}01$	$\alpha = {\cdot}005$
Critical values of Z (for one tailed test)	– 1·28 or + 1·28	– 1·645 or + 1·645	– 2·33 or + 2·33	– 2·58 or + 2·58
Critical values of Z (for two tailed test)	– 1·645 or + 1·645	– 1·96 or + 1·96	– 2·58 or + 2·58	– 2·81 or + 2·81

Note : For other level of significance α, the critical value of Z can be found from the table **"Area under the Normal Curve."**

Procedure of Testing a Hypothesis:

Testing of a hypothesis passes through the following steps :

(1) **Set up a null hypothesis :** It is denoted H_0. Null hypothesis assumes that difference between any values to be compared is not significant.

(2) **Set up a suitable level of significance :** A suitable level of significance is determined to test the null hypothesis. In practice, 5% significance level is used.

(3) **Set up a suitable test of statistic :** A number of test statistics like Z, t, χ^2, F etc. may be applied to test the null hypothesis. It is decided only on the basis of available information.

(4) **Doing necessary calculation :** After selecting appropriate statistic, computations relating to the test statistic are made and values are worked out.

(5) **Making Decisions :** In the process of hypothesis testing, results are interpreted at the final stage. For this purpose, we compare the computed value of a test statistic with the table value at a pre-determined level of significance. If computed value is greater that the table value at 5% or 1% level of significance, then null hypothesis is rejected. In such a situation, the sample does not represent population.

Applications of Tests of Hypothesis/Test of Significance/Sampling Tests

The applications of tests of hypothesis or tests of significance or sampling tests are studied under the following heads :

(A) Tests of Hypothesis for Large Samples

(B) Tests of Hypothesis for Small Samples

In this chapter we shall discuss tests of hypothesis for large samples.

(A) Tests of Hypothesis for Large Samples ($n \geq 30$): The following are some important applications of tests of hypothesis in case of large samples :

(1) Test of Hypothesis about Population Mean

(2) Test of Hypothesis about difference between Two Population Means.

(3) Test of Hypothesis about difference between Two Population S.D'S.

(4) Test of Hypothesis about Population Proportion.

(5) Test of Hypothesis about difference between two Population Proportions.

Let us discuss them briefly.

(1) Test of Hypothesis about Population Mean μ : The test of hypothesis concerning population mean μ in case of large sample requires the use of normal distribution. Let $\overline{X}$ be the mean of a large random sample of size n drawn from a normal population with mean μ and standard deviation σ. To test the hypothesis that population mean μ has a specified value, the appropriate test statistic to be used is :

$$Z = \frac{\overline{X} - \mu}{S.E_{\overline{X}}}$$

Where, $\overline{X}$ = sample mean; μ = population mean, $S.E_{\overline{X}}$ = Standard Error of Mean.

Procedure : The following steps are taken for test of hypothesis about population mean:

(i) Set up the null hypothesis $H_0 : \mu = \mu_0$ *i.e.*, there is no difference between the sample mean and population mean. Alternative Hypothesis : $H_1 : \mu \neq \mu_0$ (Two tailed test).
or $H_1 : \mu > \mu_0$ or $\mu < \mu_0$ (One tailed test)

(ii) Compute the $S.E_{\overline{X}}$ by using the following formula :

(a) $S.E_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ When population S.D. σ is known.

(b) $S.E_{\overline{X}} = \frac{s}{\sqrt{n}}$ When sample S.D. is known.

(iii) Substituting the value of $\overline{X}$, μ and $S.E_{\overline{X}}$ in the above stated Z-statistic.

(iv) Select the desired level of significance α if not given and corresponding to that level of significance (l.o.s.), we find the critical value of Z_α from the table **"Areas under the normal curve".**

(v) The computed value of Z is compared with the critical value of Z. If the computed value of $Z = |Z| <$ critical value of Z at a level of significance α, then we accept the null hypothesis H_0 and if the computed value of $Z = |Z| >$ critical value of Z at a level of significance α, then we reject the null hypothesis and accept the alternative hypothesis H_1.

Note :

1. If the population S.D. σ is not known, then sample S.D. (s) is used for large samples.
2. The critical value of Z_α (for large samples) corresponding to various level of significance are given below :

Critical Value (Z_α)	Level of Significance (α)	
	1%	5%
Two Tailed Test	$\lvert Z \rvert = 2 \cdot 58$	$\lvert Z \rvert = 1 \cdot 96$
One Tailed Test	$\lvert Z \rvert = 2 \cdot 33$	$\lvert Z \rvert = 1 \cdot 64$

For other level of significance α, the critical values Z can be found from the table **"Area under the normal curve"** given at the end of the book.

Note 3. When no reference to the level of significance is made given, then we always take $\alpha = .05$ *i.e.*, 5% level of significance.

The following examples illustrate the procedure for test of hypothesis about population mean :

Example 1. The mean height of a random sample of 100 students is 64" and standard deviation is 3". Test the statement that the mean height of population is 67" at 5% level of significance.

Solution. We are given : $n = 100, \overline{X} = 64, s = 3, \mu = 67$

Null Hypothesis :

$$H_0 : \mu = 67$$

Alternative hypothesis : $H_1 : u \neq 67$ (⇒ Two tailed test)

$$S.E._{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{3}{\sqrt{100}} = \frac{3}{10} = 0 \cdot 3 \quad \text{(For large samples, } \sigma = s\text{)}$$

Now, we compute Z-statistic as :

$$|Z| = \frac{|\overline{X} - \mu|}{S.E._{\overline{X}}} = \frac{64 - 67}{0 \cdot 3} = 10$$

At 5% level of significance, the critical value of Z for two tailed test = 1·96.

Since, the calculated value of | Z | is more than the critical value of Z at 5% level of significance, we reject the null hypothesis and conclude that the population mean cannot be equal 67.

Example 2. A random of 400 male students is found to have a mean height of 171·38 cm. Can it be reasonably regarded as a sample from a large population with mean height 171·17 cm and standard deviation 3·30 cm ? Use $\alpha = \cdot 05$.

Solution. We are given : $n = 400, \overline{X} = 171 \cdot 38, \mu = 171 \cdot 17, \sigma = 3 \cdot 30$

Null Hypothesis $H_0 : \mu = 171 \cdot 17$ *i.e.*, the sample has been drawn from population with $\mu = 171 \cdot 17$ and $\sigma = 3 \cdot 30$.

Alternative Hypothesis $H_1 : \mu \neq 171 \cdot 17$ (⇒ Two tailed test)

$$S.E._{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3 \cdot 30}{\sqrt{400}} = 0.165$$

$$|Z| = \frac{\overline{X} - \mu}{S.E._{\overline{X}}} = \frac{171 \cdot 38 - 171 \cdot 17}{0 \cdot 165} = 1 \cdot 273$$

At 5% level of significance, the critical value of Z for two tailed test = 1·96.

Since, the calculated value of | Z | is less than the critical value of Z at 5% level of significance, we accept the null hypothesis and conclude that the population mean is equal to 171·17 *i.e.*, the sample has been drawn from the population with mean 171·17.

Example 3. **A stenographer claims that she can take dictation at the rate of 120 words per minute. Can we reject her claim on the basis of 100 trials in which she demonstrates a mean of 116 words with a standard deviation of 15 words. (Use 5% l.o.s.).**

Solution. We are given : $n = 100$, $\overline{X} = 116$, $s = 15$, $\mu = 120$

Null hypothesis $H_0 : \mu = 120$ (*i.e.*, the claim is accepted)

Alternative hypothesis $H_1 : \mu < 120$ (*i.e.*, the claim is rejected)

$$S.E._{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1.5$$

$$|Z| = \frac{|\overline{X} - \mu|}{S.E._{\overline{X}}} = \frac{|116 - 120|}{1\cdot5} = \frac{4}{1\cdot5} = 2\cdot6$$

At 5% l.o.s., the critical value of Z for one tailed test = 1·645.

Since, the calculated value of $Z > 1\cdot645$, we reject H_0 and conclude that the claim of the stenographer is rejected.

Aliter : This question can also be solved by using two tailed test. Let us have; $H_0 : \mu = 120$ and $H_1 : \mu \neq 120$, it is a two tailed test.

And $H_1 : \mu \neq 120$, it is a two tailed test.

$$|Z| = 2.6$$

At 5% l.o.s. $Z_{.05} = 1.96$ (for two tailed test)

Since $|Z| > 1.96$, we reject H_0 and conclude that the claim of the stenographer is rejected.

Example 4. **An educator claims that the average IQ of government college students is no more than 110. To test this claim, a random sample of 150 students was taken and given relevant tests. Their average IQ score come to 111.2 with a standard deviation of 7·2. At level of significance of 0.01, test if the claim of the educator is justified.**

Solution. We are given: $n = 150$, $\overline{X} = 111.2$, $s = 7.2$, $\mu = 110$

Null hypothesis $H_0 : \mu \leq 110$

Alternative hypothesis : $\mu > 110$ (Right tailed test)

$$S.E_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{7\cdot2}{\sqrt{150}} = \frac{7\cdot2}{12\cdot25} = 0\cdot59.$$

$$|Z| = \frac{|\overline{X} - \mu|}{S.E_{\overline{X}}} = \frac{111\cdot2 - 110}{0\cdot59} = \frac{1\cdot2}{\cdot59} = 2\cdot03$$

At 1% level of significance, the critical value of Z for one tailed test 2·33. Since, the calculated value of $|Z|$ < critical value of Z at 1%, we accept H_0. This means that the claim of the educator is justified.

EXERCISE – 1

1. A random sample of 100 students gave a mean weight of 58 kg with S.D. of 4 kg. Test the hypothesis that the mean weight of the population is 60 kg. **[Ans.** $|Z| = 5$, H_0 is rejected**]**
2. A sample of 100 units is found to have 5 lbs as mean. Could it be regarded as a simple random sample from a large population whose mean is 5.64 lbs and $\sigma = 1\cdot5$ lbs. Use $\alpha = \cdot05$. **[Ans.** $|Z| = 4\cdot26$, H_0 is rejected**]**
3. The manufacturer of a particular make of a small car claim that on an average the car is driven 2000 kms per month. A random sample of 100 owners of the car are asked to keep a record of kilometers they drive their cars. On the basis of these sample records, it was found that an average the car was driven 2200 kms per month with a standard deviation of 600 kms. Do the sample data support the hypothesis that the average distance the car is driven has increased ? Use $\alpha = \cdot05$. **[Ans.** $Z = 3\cdot33$, H_0 is rejected**]**
4. A company claims that life of its product is 1600 hours. A sample of 100 units was tested and mean life of its products was found to be 1570 hours with a standard deviation of 120 hours. Test the claim of the company at 5% level of significance ? **[Ans.** $|Z| = 2\cdot5$, H_0 is rejected**]**
5. A weighing machine without any display was used by an average of 320 persons a day with a standard deviation of 50 parsons. When an attractive display was used on the machine, the average for 100 days increased by 15 persons. Can we say that the display did not keep much ? Use a level of significance of 0·05. **[Ans.** $Z = 3$, rejected H_0**]**

(2) Test of Hypothesis about difference two Population Means : Let $\overline{X}_1$ and $\overline{X}_2$ be the sample means of two independent random samples of large sizes n_1 and n_2 drawn from two populations having means μ_1 and μ_2 and standard deviations σ_1 and σ_2. To test whether the two population means are equal or not, the appropriate test statistic to be used is :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S.E._{\overline{X}_1 - \overline{X}_2}}$$

Where, $\overline{X}_1$ = Mean of 1st Sample, $\overline{X}_2$ = Mean of 2nd Sample $S.E._{\overline{X}_1 - \overline{X}_2}$ = Standard error of the difference of two means.

Procedure : The following steps are taken for test of hypothesis about difference between two populations means :

(i) Set up the null hypothesis $H_0 : \mu_1 = \mu_2 = 0$ *i.e.*, $\mu_1 = \mu_2$ there is no difference between the two population means.

Alterantive Hypothesis : $H_1 : \mu_1 \neq \mu_2$ (Two tailed test)

or $H_1 : \mu_1 > \mu_2$ or $\mu_1 < \mu_2$ (One tailed test)

(ii) Compute the $S.E._{\overline{X}_1 - \overline{X}_2}$ by using the following formulae :

(a) $S.E._{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ When Population S.D.s σ_1 and σ_2 are given

(b) $S.E._{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ When S.D. s_1 and s_2 of the two samples are given.

(c) $S.E_{\overline{X}_1 - \overline{X}_2} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ When two random samples have been drawn form the same population with S.D. σ.

(d) $S.E_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} - 2r \cdot \frac{s_1 \times s_2}{n_1 \times n_2}}$ When two samples are correlated.

(iii) Substituting the values of $\overline{X}_1, \overline{X}_2, \mu_1, \mu_2$ and $S.E_{\overline{X}_1 - \overline{X}_2}$ in the above stated Z-statistics.

The other steps such as (i) Level of significance, (ii) Critical Value of Z_α; (iii) Decision making for testing hypothesis of the difference between two means are the same as those given in test of hypothesis about population mean μ.

Example 5. **A random sample of 1000 workers from South India show that their mean wages are Rs. 47 per week with a standard deviation of Rs. 28. A random sample of 1500 workers from North India gives a mean wages of Rs. 49 per week with a standard deviation of Rs. 40. Is there any significant difference between the mean level of wages in two places ?**

Solution. We are given : $n_1 = 1000, \overline{X}_1 = 47, s_1 = 28$

$n_2 = 1500, \overline{X}_2 = 49, s_2 = 40$

Null Hypothesis $H_0 : \mu_1 = \mu_2$ *i.e.*, there is no significant difference between two mean wages.

Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$ (⇒ Two tailed test)

$$S.E_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(28)^2}{1000} + \frac{(40)^2}{1500}} = 1 \cdot 36$$

$$|Z| = \frac{|\overline{X}_1 - \overline{X}_2|}{S.E_{\overline{X}_1 - \overline{X}_2}} = \frac{|47 - 49|}{1 \cdot 36} = \frac{2}{1 \cdot 36} = 1 \cdot 47$$

At 5% level of significance, the critical value of Z for two tailed test $= \pm 1 \cdot 96$.

Since, the calculated value of $|Z|$ is less than the critical value of Z, we accept the null hypothesis and concluded that there is no significant difference between the two mean level of wages.

Example 6. **The mean yield of wheat from Patiala District was 210 kgs with a standard deviation 10 kgs per acre from a sample of 100 plots. In another district Ludhiana, the mean yield was 220 kgs with standard deviation 12 kgs from a sample of 150 plots. Assuming that the standard deviation of the yield in the entire state was 11 kgs, test whether there is any significant difference between the mean yield of crops in the two districts.**

Solution. We are given : $n_1 = 100, \overline{X}_1 = 210, s_1 = 10$

$n_2 = 150, \overline{X}_2 = 220, s_2 = 12$

S.D. of population $= \sigma = 11$

Null hypothesis : $H_0 : \mu_1 = \mu_2$ *i.e.*, there is no significant difference between the mean yield of crops in two districts.

Alternative hypothesis : $H_1 : \mu_1 \neq \mu_2$ (⇒ Two tailed test)

$$S.E_{\overline{X}_1 - \overline{X}_2} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{(11)^2\left(\frac{1}{100} + \frac{1}{150}\right)}$$

$$= \sqrt{\frac{121}{100} + \frac{121}{150}} = \sqrt{1 \cdot 21 + 0 \cdot 807} = \sqrt{2 \cdot 017} = 1 \cdot 42$$

$$|Z| = \frac{|\overline{X}_1 - \overline{X}_2|}{S.E_{\overline{X}_1 - \overline{X}_2}} = \frac{|210 - 220|}{1 \cdot 42} = 7 \cdot 04$$

At 5% level of significance, the critical value of Z for two tailed test $= 1 \cdot 96$.

Since, the calculated value of $|Z| = 7 \cdot 05 >$ critical value of Z, we reject H_0 in favour of H_1. It means that there is a significant difference between the mean yield of crops in two districts.

Example 7. **Following information is available in respect of two brands of bulbs (Price same):**

	Brand A	Brand B
Mean Life (Hrs.)	1300	1248
S.D. (Hrs.)	82	93
Sample Size	100	100

Which brand should be preferred at 5 percent level of significance ?

Solution. We are given : $n_1 = 100, \overline{X}_1 = 1300, s_1 = 82$

$n_2 = 100, \overline{X}_2 = 1248, s_2 = 93$

Null hypothesis $H_0 : \mu_1 = \mu_2$ *i.e.*, there is no significant difference in the mean life of the two brands of bulb.

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ (⇒ Two tailed test)

$$S.E_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(82)^2}{100} + \frac{(93)^2}{100}}$$

$$= \sqrt{\frac{6724}{100} + \frac{8649}{100}} = \sqrt{67 \cdot 24 + 86 \cdot 49} = 12 \cdot 399$$

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{S.E_{\overline{X}_1 - \overline{X}_2}} = \frac{1300 - 1248}{12 \cdot 399} = 4 \cdot 19$$

At 5% level, the critical value of Z for two tailed test $= 1 \cdot 96$.

Since, the calculated value of Z > the critical value of Z, we reject the hypothesis and conclude that there is a significant difference in the mean life of the two brands of bulbs. We should prefer to buy the bulbs of branch A since its average life is more.

Example 8. **A sample of heights of 6400 soldiers has a mean of 67·85 inches and a standard deviation of 2·56 inches while a sample of heights of 1600 sailors has a mean of 68·55 inches and a standard deviation of 2·52 inches. Do the data indicate that the sailors are on the average taller than the soldiers ?**

Solution. We are given : $n_1 = 6400, \bar{X}_1 = 67{\cdot}85, s_1 = 2{\cdot}56$
$n_2 = 1600, \bar{X}_2 = 68{\cdot}55, s_2 = 2{\cdot}52$

Null hypothesis $H_0 : \mu_1 = \mu_2$ *i.e.,* there is no significant difference in the mean height of soldiers and sailors.

Alternative hypothesis $H_1 : \mu_2 > \mu_1$ or $\mu_1 < \mu_2$ (⇒ one tailed test)

$$S.E_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(2{\cdot}56)^2}{6400} + \frac{(2{\cdot}52)^2}{1600}} = 0{\cdot}071$$

$$|Z| = \frac{|\bar{X}_1 - \bar{X}_2|}{S.E_{\bar{X}_1 - \bar{X}_2}} = \frac{|67{\cdot}85 - 68{\cdot}55|}{0{\cdot}071} = \frac{|-0{\cdot}7|}{0{\cdot}071} = 9{\cdot}859$$

At 5% level, the critical value of Z for one tailed test = 1·645.

Since, the calculated value |Z| > critical value of Z at 5% level, we reject H_0 in favour of H_1. It means that the data indicate that the sailors are on an average taller than the soldiers.

Example 9. **The means of two large sample of sizes 1000 and 2000 are 168·75 cms and 170 cms respectively. Can the samples be regarded as drawn from a population with same mean and S.D. 6·25 cms.**

Solution. We are given : $n_1 = 1000, \bar{X}_1 = 168{\cdot}75$
$n_2 = 2000, \bar{X}_2 = 170$
$\sigma = 6{\cdot}25$

Null Hypothesis : $H_0 : \mu_1 = \mu_2$ *i.e.,* both the samples are drawn from the population with mean and S.D. 6·25.

Alternative Hypothesis : $H_1 : \mu_1 \neq \mu_2$ (⇒ Two tailed test)

$$S.E_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= \sqrt{(6{\cdot}25)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}$$

$$= \sqrt{0{\cdot}03906 + 0{\cdot}01953} = 0{\cdot}242$$

$$|Z| = \frac{|\bar{X}_1 - \bar{X}_2|}{S.E_{\bar{X}_1 - \bar{X}_2}} = \frac{|168{\cdot}72 - 170|}{0{\cdot}242} = \frac{1{\cdot}28}{0{\cdot}242} = 5{\cdot}28$$

At 5% level, the critical value of Z for two tailed test = 1·96.

Since the calculated value of |Z| = 5·28 > critical value of Z = 1·96, we reject H_0 and conclude that both the samples are not drawn from the population with same mean and S.D. 6·25.

Example 10. **Two different samples from two districts yielded the following results :**

District A :	$\bar{X}_1 = 648,$	$s_1^2 = 120,$	$n_1 = 100$
District B :	$\bar{X}_2 = 495,$	$s_2^2 = 140,$	$n_2 = 90$

Test at 0·05 level of significance that $\mu_1 - \mu_2 > 150$.

Solution. We are given : $\bar{X}_1 = 648, \quad s_1^2 = 120, \quad n_1 = 100$
$\bar{X}_2 = 495, \quad s_2^2 = 140, \quad n_2 = 90$

Null hypothesis $H_0 : \mu_1 - \mu_2 = 150$.

Alternative hypothesis $H_1 : \mu_1 - \mu_2 > 150$ (⇒ Right tailed test)

$$S.E_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= \sqrt{\frac{120}{100} + \frac{140}{90}} = \sqrt{1{\cdot}20 + 1{\cdot}55}$$

$$= \sqrt{2{\cdot}75} = 1{\cdot}658$$

$$|Z| = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S.E_{\bar{X}_1 - \bar{X}_2}}$$

$$= \frac{(648 - 495) - (150)}{1{\cdot}658} = \frac{153 - 150}{1{\cdot}658}$$

$$= \frac{3}{1{\cdot}658} = 1{\cdot}809$$

At 5% level of significance, the critical value Z for one tailed test = 1·645.

Since, the calculated value of |Z| > the critical value of Z, we reject H_0 in favour of H_1. It means that the difference of the two means is greater than 150. *i.e.,* $\mu_1 - \mu_2 > 150$.

Example 10A. **In an intelligence test administrated to 60 fathers and them 100 children, the following results were obtained :**

Father's mean score 114; standard deviation 13

Son's mean score 110; standard deviation 11

Assuming the coefficient of correlation between them is + 0·75, calculate the standard error of the two means and state whether the difference is significant ?

Solution. $H_0 : \mu_1 = \mu_2$ *i.e.,* there is no significant difference in the mean score.

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ (⇒ Two tailed test)

We are given : $n_1 = 60, \quad \bar{X}_1 = 114; \quad s_1 = 13$
$n_2 = 100, \quad \bar{X}_2 = 110; \quad s_2 = 11$
$r = +{\cdot}75$

$$S.E_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} - 2r \cdot \frac{s_1 \times s_2}{n_1 \times n_2}}$$

$$= \sqrt{\frac{(13)^2}{60} + \frac{(11)^2}{100} - 2 \times 0{\cdot}75 \times \frac{13 \times 11}{60 \times 100}}$$

$= 2$

$$|Z| = \frac{|\bar{X}_1 - \bar{X}_2|}{S.E._{\bar{X}_1 - \bar{X}_2}} = \frac{|114 - 100|}{2}$$

$= 7$

At 5% level of significance, the critical value of $Z = 1{\cdot}96$ for two tailed test.
Since the calculated value of Z is greater than the critical value of Z, we reject H_0 in favour of H_1. It means that there is significant difference in the mean scores of fathers and their sons.

EXERCISE – 2

1. The following information relates to wages of workers of two factories A and B. Test whether there is any significant difference between their mean wages. Use $\alpha = .05$.

	Factory A	Factory B
Mean Wages (Rs.) :	100	105
Standard Deviation :	16	24
No. of Workers :	800	1600

[Ans. $Z = 6{\cdot}061$, H_0 is rejected]

2. A researcher claims that American 18 year old females are, on an average, taller than the British 18 year old females. To test this claim, a random sample of 50 American females and 50 British females was taken and their measurments are summaried as follows:

	American	British
Average height (in inches)	$\bar{X}_1 = 65.2$	$\bar{X}_2 = 64.5$
Standard deviation	$s_1 = 2.5$	$s_2 = 2.8$

Test the hypothesis that American females are taller than their British Counterparts at $\alpha = 0.05$

Hint : Use one tailed test. **[Ans.** $|Z| = 1.32$, Accept H_0]

3. A man buys 200 electric bulbs of each of 'Philips' and 'HMT'. He finds that Philips bulbs has a mean life of 2560 hours and a S.D. of 80 hours and HMT bulbs has a mean life of 2650 hours with a S.D. of 75 hours. Is there is a significant difference in the mean life of these two kinds of bulbs ? **[Ans.** $|Z| = 11{\cdot}612$, H_0 is rejected]

4. Two different samples from two districts yielded the following results :
District A : $\bar{X}_1 = 13{,}000$, $s_1 = 1300$, $n_1 = 100$
District B : $\bar{X}_2 = 13{,}900$, $s_2 = 1400$, $n_2 = 200$
Test at 0·05 level of significance that $\mu_1 - \mu_2 > 500$. **[Ans.** $Z = 2{\cdot}448$, H_0 is rejected]

5. 490 boys and 450 girls appeared at an examination in commerce. The mean and standard deviation of marks of boys are 54·3 and 17·5 respectively, whereas those of girls are 50·6 and 18·0. Is there a significant difference in marks of boys and girls at 1% level ?
[Ans. $Z = 3{\cdot}18$, H_0 is rejected]

6. 100 students of a college were put to tests in statistics and Accountancy respectively and the following results were obtained :
Mean marks in Statistics = 45; S. D. = 7
Mean marks in Accountancy = 43 ; S. D. = 6
between marks in the two subjects = + 0·75.
Calculate the standard error of the difference of the two means and state whether the difference is significant. **[Ans.** $S.E._{\bar{X}_1 - \bar{X}_2} = 0{\cdot}92$, $Z = 2{\cdot}17$, Reject H_0]

7. A test in statistics was conducted for a class of 70 boys and 60 girls. The test provides the following information :
Boys : $n_1 = 70$, $\bar{X}_1 = 70$, $\Sigma(X_1 - \bar{X}_1)^2 = 7500$
Girls : $n_2 = 60$, $\bar{X}_2 = 65$, $\Sigma(X_2 - \bar{X}_2)^2 = 7800$
Test whether there is a significant difference between the performance of boys and girls at 5% level of significance. **[Ans.** $|Z| = 2.60$, reject H_0]

8. Intelligence test of two groups of boys and girls gave the following results :

	Mean	S.D.	N
Girls :	61	2	84
Boys :	60	4	100

Is there a significance difference in the mean score obtained by boys and girls.
Use $\alpha = .05$ **[Ans.** $Z = 2.12$, reject H_0]

(3) Test of Hypothesis about difference between two population standard deviations : Let s_1 and s_2 be the standard deviation of two independent random samples of sizes n_1 and n_2 from two populations wtih standard deviations σ_1 and σ_2 respectively. To test whether the two population S.D.'s are equal or not, one appropriate test statistic is to be used as

$$Z = \frac{(s_1 - s_2) - (\sigma_1 - \sigma_2)}{S.E._{s_1 - s_2}}$$

Where, s_1 = S.D. of the Ist sample, s_2 = S.D. of the 2nd sample, $SE_{s_1 - s_2}$ = Standard error of the difference between two S.D's.

Procedure : The following steps are taken for testing hypothesis about difference between two population standard directions :

(1) Set up the null hypothesis $H_0 : \sigma_1 = \sigma_2$ *i.e.* there is no difference between the two population S.Ds.

Alternative hypothesis : $H_1 = \sigma_1 \neq \sigma_2$

(2) Compute the $S.E._{s_1 - s_2}$ by using the formulae.

(a) $SE_{s_1 - s_2} = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$, when Population S.Ds of σ_1 and σ_2 are given.

(b) $SE_{s_1 - s_2} = \sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}$, when S.D s_1 and s_2 of the two samples are given.

(3) Substituting the values of s_1, s_2 and $SE_{s_1 - s_2}$ in the above stated Z-statistic.

The other stages such as (i) correct significance, (ii) critical value of Z, (iii) Decision including for testing hypothesis of the difference between two SD's are the same as given in the testing hypothesis about population mean μ .

Example 11. **The mean yield of two sets of plots and their variability are as given below, Examine whether the difference in the variability in yeelds is significant at 5% level of significance.**

	Set of 40 plots	Set of 60 plots
Mean yield per plot	1258 lbs	1243 lbs
S.D. per plot	34	28

Solution. We are given : $n_1 = 40, n_2 = 60, \bar{x}_1 = 1258$ lbs, $\bar{x}_2 = 1243$ lbs, $s_1 = 34$ lbs, $s_2 = 28$ lbs.

Null Hypothesis $H_0 : \sigma_1 = \sigma_2$ *i.e.,* there is no significant difference in the variability in the yields between two sets at plots.

Alternative Hypothesis. $H_1 : \sigma_1 \neq \sigma_2$ (Two tailed test)

Level of significance $\alpha = 0.05$.

Test statistic : Under H_0 the test staristics, for large samples is

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_1}}} \sim N(0,1)$$

$$Z = \frac{34 - 28}{\sqrt{\frac{(34)^2}{80} + \frac{(28)^2}{120}}} = \frac{6}{\sqrt{\frac{1156}{80} + \frac{784}{120}}}$$

$$= \frac{6}{\sqrt{14.45 + 6.53}} = \frac{6}{\sqrt{20.98}} = \frac{6}{4.58} = 1.31$$

Since, $Z < 1.96$, it is not significant at 5% level of significance Hence, we may conclude that there is no siginficant difference in the variability in yields.

EXERCISE – 3

1. Random samples drawn from two countries gave the following data relating to the heights of adult males :

	Country A	Country B
Mean height in inches	67.42	67.25
Standard deviation in inches:	2.58	2.50
Number in samples	1000	1200

(i) Is the difference between the means significant ?

(ii) Is the difference between the standard deviations significant ?

[Ans. (*i*) $|Z| = 1.56$. Not significant. (*ii*) $|Z| = 1.03$. Not significant]

2. The standard deviation of the height of B.A. (Hons.) students of a college is 4.0". Two samples are taken. The standard deviation of the height of 100 B. Com. Hons. students is

3.5" and B.A. Econ. Hons. students is 4.5". Test the significance of the difference of standard deviations of the samples.

[Ans. $H_0 : \sigma_1 = \sigma_2$; $H_1 : \sigma_1 \neq \sigma_2$; $|Z| = 2.89$; Significant.]

3. Intelligence test given to two groups of boys and girls gave the follwoing results :

Girls : Mean Marks = 78, S.D. = 12, $N = 80$

Boys : Mean Marks = 75, S.D. = 15, $N = 120$

(i) Is the difference is the mean scores significant?

(ii) Is the difference between standard deviation significant ?

[Ans. (i) Accept H_0, (ii) Reject H_0]

(4) Test of Hypothesis about Population Proportion : A random sample of size n ($n \geq 30$) has a sample proportion p of members possessing a certain attribute (*i.e.,* proportion of success). To test the hypothesis that the population proportion P has a specified value, the appropriate test statistic to be used as :

$$Z = \frac{p - P}{SE_p}$$

Where, p = Sample proportion, P = Population proportion.

SE_p = Standard error of proportion.

Procedure : The following steps are taken for test of hypothesis about population proportion :

(i) Set up the null hypothesis $H_0 : P = P_0$ *i.e.,* there is no difference between the sample proportion and population proportion.

Alternative hypothesis $H_1 : P \neq P_0$ (⇒ Two tailed test)

or $H_1 : P > P_0$ or $H_1 : P < P_0$ (One tailed test)

(ii) Compute the SE_p by using the following formuale :

(a) $SE_p = \sqrt{\frac{PQ}{n}}$ and where P is population proportion and n is the sample size.

(b) $SE_p = \sqrt{\frac{pq}{n}}$ and $q = 1 - p$ where P is not known.

(iii) Substituting the values of p, P and SE_p in the above stated Z-statistic.

The other steps such as (i) Level of significance, (ii) Critical value of Z_α, (iii) Decision making for testing hypothesis of the population proportion P are the same as those given in test of hypothesis about the population mean μ.

Example 12. **A coin is tossed 100 times under identical conditions independently yielding 30 heads and 70 tails. Test at 1% level of significance whether or not the coin is unbiased.**

Solution: **Here, the sample proportion p = Proportion to heads in the sample $= \frac{30}{100} = 0{\cdot}30$**

Also the population proportion $P = \frac{1}{2} = 0{\cdot}50 \Rightarrow Q = 1 - 0{\cdot}50 = 0{\cdot}50$

Null hypothesis $H_0: P = 0.5$ (That is, coin is unbiased)

Alternative hypothesis $H_1: P \neq \frac{1}{2}$ ($\Rightarrow$ Two tailed test)

$$SE_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{.50 \times .50}{100}} = 0.05$$

$$|Z| = \frac{|p-P|}{SE_p} = \frac{|0.30 - .50|}{.05} = \frac{0.20}{.05} = 4$$

At 1% level of significance, the critical value of Z for two tailed test = 2·58.
Since, the calculated value of $|Z| >$ the critical value of Z, we reject the null hypothesis and hence conclude that the coin is biased.

Aliter : This question can also be solved on the basis of number of successes as follows :

Given : $n = 100$, $P = P(H) = \frac{1}{2} = 0.5; Q = 1 - P = 1 - 0.5 = 0.5$

$H_0 = nP = 50$, $H_1 = nP \neq 50$

$\alpha = .01$, $Z = 2.58$ (Critical value), $np = 30$

$$S.E_{np} = \sqrt{nPQ} = \sqrt{100 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{25} = 5$$

$$|Z| = \frac{|np - nP|}{S.E_{np}} = \frac{|30 - 50|}{5} = \frac{20}{5} = 4$$

Since, the calculated value of $Z >$ the critical value of Z, we reject the null hypothesis and hence conclude that the coin is biased.

Example 13. **In 324 throws of six-faced dice, odd points appeared 180 times. Would you say that the die is fair ? Use $\alpha = 0.05$.**

Solution. Here the sample proportion $p = \frac{180}{324} = 0.555$

Also the population proportion $P = \frac{1}{2} = 0.50 \Rightarrow Q = 1 - 0.5 = 0.50$

Null hypothesis $H_0: P = 0.5$

Alternative hypothesis $H_1: P \neq 0.5$ ($\Rightarrow$ Two tailed test)

$$S.E_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{.50 \times .50}{324}} = \sqrt{\frac{0.25}{324}} = 0.027$$

Using Z-statistic, we have

$$|Z| = \frac{p-P}{S.E_p} = \frac{0.555 - .50}{.027} = \frac{.055}{.027} = 2.03$$

At 5% level of significance, the critical value of Z for two tailed test = 1·96.
Since, the calculated value of $Z >$ the critical value of Z, we reject the null hypothesis and conclude that the die is not fair.

Aliter : This question can also be solved on the basis of number of successes as follows :

Given : $n = 324$, $P = P$ (odd points) $= \frac{1}{2} = 0.5; Q = 1 - P = 1 - 0.5 = 0.5$

$H_0: nP = 162$,

$H_1: nP \neq \frac{1}{2}, \alpha = .05$

At $\alpha = .05$, $Z = 1.96$ (Critical value), $np = 180$

$$S.E_{np} = \sqrt{nPQ} = \sqrt{324 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{81} = 9$$

$$|Z| = \frac{|np - nP|}{S.E_{np}} = \frac{|180 - 162|}{9} = \frac{18}{9} = 2$$

Since, the concluded value of $Z >$ the critical value of Z, we reject the null hypothesis and hence conclude that the die is not fair.

Example 14. **In a sample of 500 persons from a village in Haryana, 280 are found to be rice eaters and the rest wheat eaters. Can we assume that both the food articles are equally popular ?**

Solution. Here the sample proportion $= p = \frac{280}{500} = 0.56$

(*i.e.*, Proportion of rice eaters in the sample)

Also the population proportion $= P = \frac{1}{2} = 0.50 \Rightarrow 1 - 0.50 = 0.50$

Null hypothesis $H_0: P = \frac{1}{2}$ *i.e.*, both the food articles are equally popular

Alternative hypothesis $H_1: P \neq \frac{1}{2}$ ($\Rightarrow$ Two tailed test)

$$S.E_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{.50 \times .50}{500}} = 0.022$$

Using Z-statistic, we have

$$|Z| = \frac{p-P}{S.E_p} = \frac{0.56 - 0.50}{.022} = \frac{0.06}{.022} = 2.727$$

At 5% level of significance, the critical value of Z for two tailed test = 1·96.
Since the concluded value of $Z >$ the critical value of Z, we reject H_0 in favour of H_1 and hence both the food articles are not equally popular.

Aliter : This question can also be solved on the basis of number of successes as follows :

Given : $n = 500$, $P = P$ (rice eater) $= \frac{1}{2} = 0.5; Q = 1 - P = 1 - 0.5 = 0.5$

$H_0: nP = 250$,

$H_1: nP \neq 250$, $\alpha = .05$

At $\alpha = .05$, $Z = 1.96$ (Critical value), $np = 280$

$$S.E_{np} = \sqrt{nPQ} = \sqrt{500 \times \frac{1}{2} \times \frac{1}{2}} = 11 \cdot 18$$

$$|Z| = \frac{|np - nP|}{S.E_{np}} = \frac{|280 - 250|}{11 \cdot 18} = 2 \cdot 68$$

Since, the calculated value of Z > the critical value of Z, we reject the null hypothesis and hence conclude that both the food articles are not equally popular.

Example 15. **In a hospital, 480 female and 520 male babies were born in a week. Do these figures confirm the hypothesis that males and females are born in equal number ?**

Solution. Here, the sample proportion $= p = \frac{480}{1000} = 0 \cdot 48$

Also, the population proportion $= P = \frac{1}{2} = 0 \cdot 50$

Null hypothesis $H_0 : P = \frac{1}{2}$ *i.e.*, male and female are born in equal number.

Alternative hypothesis $H_1 : P \neq \frac{1}{2}$ ($\Rightarrow$ Two tailed test)

$$S.E_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{0 \cdot 50 \times \cdot 50}{1000}} = 0 \cdot 0158$$

Using Z-statistic, we have

$$|Z| = \frac{|p - P|}{S.E_p} = \frac{|0 \cdot 48 - 0 \cdot 50|}{0 \cdot 0158} = \frac{\cdot 02}{\cdot 0158} = 1 \cdot 26$$

At 5% level of significance, the critical value of Z for two tailed test $= 1 \cdot 96$.

Since, the calculated value of Z < critical value of Z, we accept H_0. This means that the figures support the hypothesis that males and females are born in equal number.

Aliter : This question can also be solved on the basis of number of successes as follows :

Given : $n = 1000$, $P = P$ (Female) $= \frac{1}{2} = 0 \cdot 5; Q = 1 - P = 1 - 0 \cdot 5 = 0 \cdot 5$

$H_0 : nP = 500$, $H_1 : nP \neq 500$, $\alpha = \cdot 05$

At $\alpha = \cdot 05$, $Z = 1 \cdot 96$ (Critical value), $np = 480$

$$S.E_{np} = \sqrt{nPQ} = \sqrt{1000 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{250} = 15 \cdot 81$$

$$|Z| = \frac{|np - nP|}{S.E_{np}} = \frac{|480 - 500|}{15 \cdot 81} = 1 \cdot 26$$

Since, the calculated value of Z < the critical value of Z, we accept the null hypothesis and hence conclude that males and females are born in equal numbers.

Example 16. **A wholesaler in apples claims that only 4% of the apples supplied by him are defective. A random sample of 600 apples contained 36 defective apples. Test the claim of the wholesaler at 5% level of significance.**

Solution. Here, the sample proportion $= p =$ proportion of defective apples $= \frac{36}{600} = 0 \cdot 06$.

Also, the population proportion $= P = 4\% = \cdot 04 \Rightarrow Q = 1 - \cdot 04 = \cdot 96$

Null hypothesis $H_0 : P = 4\%$ or $\cdot 04$

Alternative hypothesis $H_1 : P \neq 4\%$ or $\cdot 04$

It is a case of two tailed test

$$S.E_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{0 \cdot 04 \times 0 \cdot 96}{600}} = 0 \cdot 008$$

Using Z-statistic, we have

$$|Z| = \frac{|p - P|}{S.E_p} = \frac{|0 \cdot 06 - 0 \cdot 04|}{\cdot 008} = \frac{0 \cdot 02}{\cdot 008} = 2 \cdot 5$$

At 5% level of significance, the critical value of Z for two tailed test $= 1 \cdot 96$.

Since, the calculated value of Z > the critical value of Z, we reject H_0 in favour of H_1. It means that the claim of the wholesaler cannot be accepted.

Example 17. **A manufacturer claimed that 95% of the equipment which are supplied to a factory conform to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at a level of significance (a) ·05 and (b) 0·01.**

Solution. Here, the sample proportion

$p =$ Proportions of pieces conforming to specifications para

$$= \frac{200 - 18}{200} = \frac{182}{200} = 0 \cdot 91.$$

Also population proportion $= P = 95\%$ or $0 \cdot 95 \Rightarrow Q = 1 - 0 \cdot 95 = 0 \cdot 05$

Null hypothesis $H_0 : P \geq 0 \cdot 95$ (Claim is justified)

Alternative hypothesis $H_1 : P < 0 \cdot 95$ (Claim is not justified)

It is a case of left tailed test.

$$S.E_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{0 \cdot 95 \times 0 \cdot 05}{200}} = 0 \cdot 0154$$

Using Z-statistic, we have

$$|Z| = \frac{|p - P|}{S.E_p} = \frac{|0 \cdot 91 - 0 \cdot 95|}{\cdot 0154} = 2 \cdot 6$$

At 5% level, the critical value of Z for left tailed test $= 1 \cdot 645$

At 1% level, the critical value of Z for left tailed test $= 2 \cdot 33$

Since, the calculated value of Z > the critical value of Z for one tailed test both at 5% and 1%, we reject H_0 in favour of H_1. It means that the manufacture's claim is rejected.

Example 18. **In a big city, 450 men out of a sample of 850 were found to be smokers. Does this information support the view that the majority of men in the city are smokers ? Assume $\alpha = \cdot 01$.**

Solution. Here, sample proportion $= p =$ proportion of smokers $= \frac{450}{850} = 0{\cdot}53$

And, population proportion $= P = \frac{1}{2} = 0{\cdot}50$

Also $Q = 1 - P = 1 - 0{\cdot}5 = 0{\cdot}5$

Null hypothesis $H_0 : P = 0{\cdot}50$ (That is, smokers and non-smokers are equal in numbers)

Alternative hypothesis $H_1 : P > 0{\cdot}50$ ($\Rightarrow$ Right tailed test)

$$S.E_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{0{\cdot}50 \times 0{\cdot}50}{850}} = 0{\cdot}0171$$

Using Z-statistic, we have

$$|Z| = \frac{p - P}{S.E_p} = \frac{0{\cdot}53 - 0{\cdot}50}{{\cdot}00171} = \frac{{\cdot}03}{0{\cdot}0171} = 1{\cdot}754$$

At 1% level of significance, the critical value of Z for right tailed test $= 2{\cdot}33$.

Since, the calculated value of $|Z| <$ the critical value of Z, we accept H_0 and conclude that majority of mean in city are not smokers.

Example 19. **A manufacturer claims that a shipment of finished nails contains less than 2% defective nails. A random sample of 400 nails when examined for defective items, is found to be containing 16 defectives. Test the claim at $\alpha = 0{\cdot}05$ level of significance.**

Solution. Here, the sample proportion $= p =$ proportion of defective nail $= \frac{16}{400} = 0{\cdot}04$

And, population proportion $= P = 2\%$ or $0{\cdot}02 \Rightarrow Q = 1 - P = 1 - 0{\cdot}02 = 0{\cdot}98$

Null hypothesis $H_0 : P < 0{\cdot}02$ *i.e.*, his claim is accepted.

Alternative hypothesis $H_1 : P \geq 0{\cdot}02$ ($\Rightarrow$ right tailed test)

$$S.E_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{0{\cdot}02 \times 0{\cdot}98}{400}} = 0{\cdot}007$$

Using Z-statistic, we have

$$|Z| = \frac{p - P}{S.E_p} = \frac{0{\cdot}04 - 0{\cdot}02}{0{\cdot}007} = 2{\cdot}857$$

At 5% level of significance, the critical value of Z for one tailed test $= 1{\cdot}645$.

Since, the calculated value of Z > the critical value of Z for one tailed test, we reject the null hypothesis and conclude that the manufacturer's claim is not acceptable.

EXERCISE – 4

1. A coin is tossed 10,000 times and head turns up 5195 times. Would you consider the coin unbiased ? **[Ans.** $Z = 3{\cdot}9, H_0$ is rejected]
2. A die is thrown 49152 times and out of these 25145 yielded either 4 or 5 or 6. Is this consistent with the hypothesis that the die is unbiased. **[Ans.** $Z = 5{\cdot}37, H_0$ is rejected]

3. A die was thrown 9000 times and out of these 3220 yielded a 3 or 4. Is this consistent with the hypothesis that the die was unbiased ? **[Ans.** $Z = 5, H_0$ is rejected]
4. A sales clerk in the departmental store claims that 60% of the shoppers entering the store leave without making a purchase. A random sample of 50 shoppers showed that 35 of them left without buying any thing. Are these sample results consistent with the claim of the sale clerk ? Use a 5% level of significance. **[Ans.** $Z = 1{\cdot}44, H_0$ is accepted]
5. In a sample of 400 parts manufactured by a company, the number of defective parts were found to be 30. The company, however claimed that only 5% of their product is defective. Test at 5% level of significance whether the claim of the company is tenable. **[Ans.** $Z = 2{\cdot}29$, reject H_0 *i.e.*, claim is not tenable]
6. A manufacturer claims that at least 90% of his goods supplied conform to specifications. A sample of 100 pieces has shown that 20 were faulty. Test his claim at 5% level of significance. **[Ans.** $|Z| = 3{\cdot}33, H_0$ is rejected]
7. In a big city, 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers [State the hypothesis clearly]. **[Ans.** $|Z| = 2{\cdot}04, H_0$ is rejected]
8. A medicine is claimed to be successful in 90% cases. In a sample of 200 patients 160 were cured. Will you accept the claim at 1% level of significance ? **[Ans.** $Z = 4{\cdot}71$, reject H_0]

(5) Test of Hypothesis about the difference between two population Proportions : Let p_1 and p_2 be the sample proportions obtained in large sample of sizes n_1 and n_2 drawn from respective populations having proportions P_1 and P_2. To test the hypothesis that there is no difference between the two population proportions, the appropriate test statistic to be used is :

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{S.E_{p_1 - p_2}}$$

Where, p_1 = 1st sample proportion, p_2 = 2nd sample proportion

$S.E_{p_1 - p_2}$ = Standard error of the difference of two proportions.

Procedure : The following steps are taken for test of hypothesis about the difference between two population proportions :

(i) Set up the null hypothesis $H_0 : P_1 = P_2$ *i.e.*, there is no difference betweeen two population proportions.

Alternative hypothesis : $H_1 : P_1 \neq P_2$ (Two tailed test)

or $H_1 : P_1 > P_2$ or $P_1 < P_2$ (One tailed test)

(ii) Compute $S.E_{p_1 - p_2}$ by using any one of the following :

(a) When the population proportions P_1 and P_2 are known, then

$$S.E_{p_1 - p_2} = \sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}}$$

Where, n_1 and n_2 are the sizes of the two sampes.

(b) When the population proportion P_1 and P_2 are not known but sample proportions p_1 and p_2 are known.

$$S.E_{p_1 - p_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \text{Where } p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} \text{ so that } q = 1 - p$$

(P = pooled estimate)

(iii) Substituting the values of p_1, p_2, P_1, P_2 and $S.E._{p_1-p_2}$ in the above stated Z-statistic.

The other steps such as (i) level of significance, (ii) critical value of Z_α, (iii) Decision making for testing the hypothesis of the difference between two proportions are the same as those given in test of hypothesis.

Example 20. **In a certain district A, 450 persons were considered regular consumers of tea out of a sample of 1000 persons. In another district B, 400 persons were regular consumers of tea out of a sample of 800 persons. Do these data indicate a significant difference between the two districts so far as drinking habit is concerned ? (Use 5% level)**

Solution. Let P_1 and P_2 be the population proportions of persons who are regular consumers of tea in the districts A and B respectively.

We set up null hypothesis $H_0 : P_1 = P_2$ *i.e.*, there is no significant difference in tea drinking habits in two districts.

Alternative hypothesis $H_1 : P_1 \neq P_2$ (⇒ Two tailed test)

$$n_1 = 1000,\ p_1 = \frac{450}{1000} = 0{\cdot}45,\ n_2 = 800,\ p_2 = \frac{400}{800} = 0{\cdot}5$$

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{450 + 400}{800 + 1000} = \frac{850}{1800} = 0{\cdot}47$$

and $q = 1 - p = 1 - 0{\cdot}47 = 0{\cdot}53$

$$S.E._{p_1-p_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{(0{\cdot}47)\,(0{\cdot}53)\left(\frac{1}{1000} + \frac{1}{800}\right)}$$

$$= \sqrt{0{\cdot}00056} = 0{\cdot}0237$$

Now, using Z-statistic we have

$$|Z| = \frac{p_1 - p_2}{S.E._{p_1-p_2}}$$

$$= \frac{0{\cdot}45 - 0{\cdot}50}{0{\cdot}0237} = \frac{0{\cdot}05}{0{\cdot}0237} = 2{\cdot}1097$$

At 5% level, the critical value Z for two tailed test $= 1{\cdot}96$.

Since, the calculated value of Z > the critical value of Z, we reject the H_0 and conclude that there is significant difference between the two districts so far as tea-drinking habit is concerned.

Example 21. **A machine produced 20 defective articles in a batch of 400. After overhauling it produced 10 defectives in a batch of 300. Has the machine improved ?**

Solution. Let P_1 and P_2 be the proportions of defective articles in the population of articles manufactured by the machine before the after overhauling, respectively.

We set up the null hypothesis $H_0 : P_1 \leq P_2$.

Alternative hypothesis $H_1 : P_2 < P_1$ or $P_1 > P_2$ (⇒ One tailed test)

We have, $n_1 = 200,\ p_1 = \frac{20}{400} = 0{\cdot}05,\ n_2 = 300,\ p_2 = \frac{10}{300} = {\cdot}0333$

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{20 + 10}{400 + 300} = \frac{3}{70} = \frac{3}{70},\ q = 1 - p = \frac{67}{70}$$

$$S.E._{p_1-p_2} = \sqrt{\frac{3}{70} \times \frac{67}{70} \cdot \left(\frac{1}{400} + \frac{1}{300}\right)} = \sqrt{\frac{201}{4900} \times \frac{7}{1200}}$$

$$= \sqrt{\frac{201}{840000}} = \sqrt{0{\cdot}0002392} = 0{\cdot}0155$$

Now, using Z-statistic, we have

$$|Z| = \frac{p_1 - p_2}{S.E._{p_1-p_2}} = \frac{0{\cdot}05 - 0{\cdot}033}{0{\cdot}0155} = \frac{0{\cdot}017}{0{\cdot}0155} = 1{\cdot}096$$

At 5% level of significance, the critical value of Z for one tailed test $= 1{\cdot}645$.

Since the calculated value of $|Z| = 1{\cdot}096 <$ the critical value of Z, we accept H_0 at 5% level of significance and conclude that machine has not improved.

Example 22. **Before an increase in excise duty on tea, 400 persons out of a sample of 500 persons were found to be tea-drinkers. After an increase in excise duty, 400 persons were known to be tea drinkers in a sample of 600 persons. Do you think that there has been a significant decrese in the consumption of tea after the increase in excise duty ?**

Solution. Let P_1 and P_2 be the proportions of tea drinkers in the population of persons before and after the increase in excise duty.

We set up the null hypothesis $H_0 : P_1 = P_2$.

Alternative hypothesis $H_1 : P_2 < P_1$ or $P_1 > P_2$ (⇒ One tailed test)

Now, $n_1 = 500,\ p_1 = \frac{400}{500} = 0{\cdot}8,\ n_2 = 600,\ p_2 = \frac{400}{600} = 0{\cdot}667$

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{400 + 400}{500 + 600} = \frac{800}{1100} = \frac{8}{11}$$

And $q = 1 - p = 1 - \frac{8}{11} = \frac{3}{11}$

$$S.E._{P_1-P_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{8}{11} \times \frac{3}{11}\left(\frac{1}{500} + \frac{1}{600}\right)} = 0{\cdot}027$$

Using Z-statistic, we have

$$|Z| = \frac{p_1 - p_2}{S.E._{P_1-P_2}} = \frac{0{\cdot}8 - 0{\cdot}667}{0{\cdot}027} = 4{\cdot}93$$

At 5% level of significance, the critical value of Z for one tailed test $= 1{\cdot}645$.

Since, the calculated value of $|Z| = 4{\cdot}93 >$ critical value of $Z = 1{\cdot}645$, we reject the null hypothesis and conclude that there is a significant decrease in the consumption of tea after the increase in excise only.

Example 23. **A sample survey of tax payers belonging to Business class and Professional class yielded the following results :**

	Business Class	**Professional Class**
Sample Size	$n_1 = 400$	$n_2 = 420$
Defaulters in tax payment	$x_1 = 80$	$x_2 = 65$

Test the hypothesis that the defaulters rate is the same for the two classes of tax payers. (Use $\alpha = \cdot 05$ level of significance).

Solution. Let P_1 and P_2 be the proportions of tax-defaulters in business and professional classes, respectively.

We set up the null hypothesis $H_0 : P_1 = P_2$ *i.e.*, there is no significant difference in the defaulters rate for two classes of tax payers.

Alternative hypothesis $H_1 : P_1 \neq P_2$ (⇒ Two tailed test)

We have $n_1 = 400, p_1 = \frac{80}{400} = 0 \cdot 20, n_2 = 420, p_2 = \frac{64}{420} = 0 \cdot 15$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{80 + 65}{400 + 420} = \frac{145}{820} = 0 \cdot 177$$

and $q = 1 - p = 1 - 0 \cdot 177 = 0 \cdot 823$

$$\text{S.E}_{p_1 - p_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0 \cdot 177 \times 0 \cdot 823\left(\frac{1}{400} + \frac{1}{420}\right)} = 0 \cdot 0266$$

Using Z-statistic, we have

$$|Z| = \frac{p_1 - p_2}{\text{S.E}_{p_1 - p_2}} = \frac{0 \cdot 20 - 0 \cdot 15}{0 \cdot 0266} = \frac{\cdot 05}{\cdot 0266} = 1 \cdot 87$$

At 5% level, the critical value of Z for two tailed test $= 1 \cdot 96$.

Since, the calculated value of Z < the critical value of Z, we accept H_0 and conclude that there is no significant difference in the defaulters rate for two classes of tax payers.

Example 24. **A market researcher engaged by a particular company believes that the proportion of households using company's products in city A exceeds this proportion in city B by $0 \cdot 05$. The reseacher conducts survey of two cities and finds the following results :**

City A	Sample Size	No. of households using company's products
A	$n_1 = 160$	120
B	$n_2 = 150$	100

Use $0 \cdot 05$ level of significance and test the researcher's claim.

Solution. Let P_1 and P_2 the proportions of households using company's products in city A and city B.

We set up the null hypothesis $H_0 : P_1 - P_2 = 0 \cdot 05$

Alternative hypothesis : $P_1 - P_2 > 0 \cdot 05$ (⇒ One tailed test)

Now, $n_1 = 160, p_1 = \frac{120}{160} = 0 \cdot 75, n_2 = 150, p_2 = \frac{100}{150} = 0 \cdot 67$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 100}{160 + 150} = \frac{220}{310} = 0 \cdot 71$$

and $q = 1 - p = 1 - 0 \cdot 71 = 0 \cdot 29$

$$\text{S.E}_{P_1 - P_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0 \cdot 71 \times 0 \cdot 29\left(\frac{1}{160} + \frac{1}{150}\right)} = 0 \cdot 0515$$

Now, using Z-statistic, we have

$$|Z| = \frac{(p_1 - p_2) - (P_1 - P_2)}{\text{S.E}_{P_1 - P_2}} = \frac{(0 \cdot 75 - 0 \cdot 67) - (\cdot 05)}{0 \cdot 0515} = \frac{\cdot 03}{0 \cdot 0515} = 0 \cdot 582$$

At 5% level, the critical value of Z for one tailed test $= 1 \cdot 645$.

Since, $|Z| = 0 \cdot 58 < 1 \cdot 645$, we accept H_0 and conclude that the proportion of households using the company's products in city A exceeds to that of city B by $0 \cdot 05$.

EXERCISE – 5

1. In a sample of 600 students of a certain college, 400 are found to use dot pens. In another college from a sample of 900 students, 450 were found to use dot pens. Test whether the two colleges are significantly different with respect to the habit of using dot pens at 5% level of significance. **[Ans. $|Z| = 6 \cdot 51$ H_0 is rejected]**
2. A machine produced 20 defective articles in a batch of 500. After overhauling it produced 3 defective articles in a batch of 100. Has the machine improved ? **[Ans. $|Z| = \cdot 476$, H_0 is accepted]**
3. Before an increase in excise duty on tobacco, 400 people out of 500 were found to be smokers. After an increase in the excise duty, 400 persons out of 600 were known to be smokers. Do you think that there has been a significant decrease in the proportions? **[Ans. $|Z| = 4 \cdot 81$, H_0 is rejected]**
4. In a sample of 600 men from a city, 450 are found to be smokers. In a sample of 900 from another city, 450 are found to be smokers. Do the data indicate that two cities differs significantly in their smoking habits ? **[Ans. $|Z| = 9 \cdot 7$, H_0 is rejected]**
5. From a random sample of 200 students from the Kurukshetra University, 90 were found to be copying and from a similar sample of 160 students from the M.D.U. Rohtak, 64 were found to be copying. Do these figures suppose the hypothesis that there is no significant difference between the two universities so far as the proportion of coyping students is concerned ? **[Ans. $Z = 0 \cdot 95$, H_0 is accepted]**
6. A sample survey of tax payers belonging to business class and professional class yielded the following results :

	Business Class	Professional Class
Sample Size	$n_1 = 200$	$n_2 = 160$
Defaulters in tax payment	$x_1 = 90$	$x_2 = 64$

Test that the defaulters rate is the same for the two classes of tax payers. **[Ans. $|Z| = 0 \cdot 95$, H_0 is accepted]**

7. 500 units from a factory are inspected and 12 are found to be defective, similarly, 800 units from another factory are inspected and 12 are found to be defective can it be concluded at 5% level of significance that productions in second factory is better than in first factory. **[Ans. $Z = 1 \cdot 184$, H_0 is accepted]**
8. A survey of television audience in a big city revealed that a particular hight programme was liked by 50 out of 200 males and 80 out of 250 females. Test the hypothesis at $\alpha = \cdot 05$ level

of significance whether that is a great difference of opinion about the proramme between the males and females.

[Ans. $|Z| = 1\cdot627$, H_0 is accepted]

9. In a random sample of 2,100 persons from Panjab 1260 persons are found to be honest. In another random sample of 4000 persons from Haryana 2360 persons are found to be honest. Do the data indicate at 1 percent level of significance that (a) the two cities are different with respect to honesty ? (b) the persons in Panjab are more honest as compared to Haryana.

[Ans. (a) $H_0 : P_1 = P_2;\ H_1 : P_1 \neq P_2,\ |Z| = 0\cdot75$, Accept H_0
(b) $H_0 : P_1 \leq P_2;\ H_1 : P_1 > P_2,\ |Z| = 0\cdot75$, Accept H_0]

10. An advertising agency wants to find out if there is any difference in the degree of loyalty for a given brand of cereal between men and women. A randon sample of 200 men and 200 women was taken and it was determined that 58% of women and 65% of men showed brand loyalty. At $\alpha = 0.05$, test the null hypothesis that there is no significant difference between the population of men and women who are brand loyal.

[Ans. 2 = 1.47, Accept H_0]

11. Company is considering two different TV advertisements for promotion of a new product. Management believes that advertisement A is more effective than advertisement B. Two test market X and Y with virtually identical consumer characteristics were selected to test this belief. A is used in X and B is used in Y. In market X, out of 60 customers who saw the advertisement, 18 bought the product, while for market B, the respective figures were 100 and 22. Do the results indicate that advartisement A is more effective than advertisement B? Test at 5% level of significance.

(**Hint :** Use one tailed test) **[Ans.** Z = 1.133, Accept H_0]

MISCELLANEOUS SOLVED EXAMPLES

Example 25. **A man buys 50 electric bulbs of each of 'Philips' and 'HMT'. He finds that Philips bulbs has a mean life of 1500 hours and a S.D. of 60 hours and HMT bulbs has a mean life of 1512 hours with a S.D. of 80 hours. Is there a significant difference in the mean life of these two kinds of bulbs ?**

Solution. Let us take the hypothesis that there is no significant difference in mean life of two makes of bulbs *i.e.,*

Null hypothesis $H_0 : \mu_1 = \mu_2$ And Altenative hypothesis $H_1 : \mu_1 \neq \mu_2$

($\Rightarrow$ Two tailed test)

Given : $n_1 = 500, \overline{X}_1 = 1500, s_1 = 60$

$n_2 = 50, \overline{X}_2 = 1512, s_2 = 80$

$$S.E._{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(60)^2}{50} + \frac{(80)^2}{50}}$$

$$= \sqrt{\frac{3600}{50} + \frac{6400}{50}} = \sqrt{72 + 128}$$

$$= \sqrt{200} = 14\cdot14$$

$$|Z| = \frac{|\overline{X}_1 - \overline{X}_2|}{S.E._{\overline{X}_1 - \overline{X}_2}} = \frac{|1500 - 1512|}{14\cdot14} = 0\cdot848$$

At 5% level of significance, the critical value of Z for two tailed test = 1·96.
Since. the calculated value of Z is less than 1·96 at 5% level of significance, we accept the null hypothesis and conclude that the difference in mean life of two kinds of bulbs is not significant.

Example 26. **A sample of 400 persons was taken from a large population. The mean weight and standard deviation of these persons was found to be 68 kg. and 6 kg. Can it reasonably be said that in the population mean weight will be 67 kg ?**

Solution. Let us take the hypothesis that the population mean is equal to 67 *i.e.,*

$H_0 : \mu = 67$ and $H_1 : \mu \neq 67$

Given : $n = 400, \overline{X} = 68, \mu = 67, s = 6$

$$S.E._{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{6}{\sqrt{400}} = \frac{6}{20} = 0\cdot3$$

$$|Z| = \frac{\overline{X} - \mu}{S.E._{\overline{X}}} = \frac{68 - 67}{0\cdot3} = \frac{1}{\cdot3} = 3\cdot33$$

At 5% level of significance, the critical value of Z for two tailed test = 1·96.
Since, the calculated value of Z is more than 1·96 at 5% level of significanc, we reject the null hypothesis and conclude that population mean is not equal to 67.

Example 27. **The means of two large samples of size 400 and 900 are 75 and 75·75 respectively. Test the equality of the means of two populations each with S.D. of 2.**

Solution. Let us take the null hypothesis that there is no difference between the two population means *i.e.,*

$H_0 : \mu_1 = \mu_2$ and $H_1 : \mu \neq \mu_2$ ($\Rightarrow$ Two tailed test)

Given : $n_1 = 400, n_2 = 900, \overline{X}_1 = 75, \overline{X}_2 = 75\cdot75, \sigma = 2$

$$S.E._{(\overline{X}_1 - \overline{X}_2)} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 2\sqrt{\frac{1}{400} + \frac{1}{900}}$$

$$= 2\cdot\sqrt{\frac{9+4}{3600}} = 2\cdot\sqrt{\frac{13}{3600}} = 0\cdot12$$

$$|Z| = \frac{|\overline{X}_1 - \overline{X}_2|}{S.E._{\overline{X}_1 - \overline{X}_2}} = \frac{|75 - 75\cdot75|}{0\cdot12} = 6\cdot25$$

At 5% level significance, the critical value of Z for two tailed test = 1·96.
Since, the calculated value of $|Z|$ is greater than the critical value of Z, we reject the hypothesis and conclude that the means differ significantly.

Example 28. **Intelligence test given to two groups of boys and girls gave the following results :**

	Mean Marks	S.D.	Number of Students
Girls	78	12	80
Boys	75	14	120

Is the difference in the mean scores significant ?

Solution. We are given :

$n_1 = 80, \quad \bar{X}_1 = 78, \quad s_1 = 12$

$n_2 = 120, \quad \bar{X}_2 = 75, \quad s_2 = 14$

Let us take the null hypothesis that the mean of two populations are equal *i.e.,*

$H_0 : \mu_1 = \mu_2$

Alternative hypothesis : $H_1 : \mu_1 \neq \mu_2$ (⇒Two tailed test)

$$S.E_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(12)^2}{80} + \frac{(14)^2}{120}}$$

$$= \sqrt{1 \cdot 8 + 1 \cdot 63} = 1 \cdot 852$$

$$|Z| = \frac{|\bar{X}_1 - \bar{X}_2|}{S.E_{\bar{X}_1 - \bar{X}_2}} = \frac{78 - 75}{1 \cdot 852} = 1 \cdot 62$$

At 5% level of significance, the critical value of Z for two tailed test $= 1 \cdot 96$.

Since, the calculated vlaue of $|Z| = 1 \cdot 62 <$ critical value of Z, we accept the null hypothesis and conclude that there is no significant difference between the means score of boys and girls.

Example 29. An automobile manufacturer asserts that the seat belts of his firms are 90 percent effective. A consumer group tests the seat belt on 50 cars and find it effective on 37 of item. Test the correctness of the manufacturer assertion at 5% level of significance.

Solution. Here, the sample proportion $= p = \frac{37}{50} = 0 \cdot 74$

Also, the population proportion $= P = 90\%$ or $0 \cdot 90 \Rightarrow Q = 1 - \cdot 90 = 0 \cdot 10$

Null hypothesis $H_0 : P = 90\%$ or $0 \cdot 90$ (Claim is justified)

Alternative hypothesis $H_1 : P < 0 \cdot 90$ (Claim is not justified)

It is a case of left tailed test

$$S.E_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{0 \cdot 90 \times \cdot 10}{50}} = 0 \cdot 04$$

$$|Z| = \frac{|p - P|}{S.E_p} = \frac{|0 \cdot 74 - 0 \cdot 90|}{\cdot 04} = 4$$

At 5% level of significance, the critical value of Z for left tailed test $= 1 \cdot 645$.

Since, the calculated value of $|Z| >$ the critical value of Z, we reject H_0 and conclude that the automobile manufacture's assertion is not correct.

Example 30. In two large populations, there are 30% and 25% respectively of curly haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations ?

Solution. Let P_1 and P_2 be proportions of curly hair people in two large populations.

Given : $P_1 = 30\% = 0 \cdot 30, Q = 1 - 0 \cdot 30 = 0 \cdot 70, n_1 = 1200$

$P_2 = 25\% = 0 \cdot 25, Q_2 = 1 - 0 \cdot 25 = 0 \cdot 75, n_2 = 900$

We set up the null hypothesis $H_0 : P_1 = P_2$ *i.e.,* the difference in population is likely to be hidden in the samples.

Alternative hypothesis $H_1 : P_1 \neq P_2$

$$S.E_{P_1 - P_2} = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}} \quad \text{(⇒Two tailed test)}$$

[Here, proportion of successes in two populations are known]

$$= \sqrt{\frac{\cdot 3 \times \cdot 7}{1200} + \frac{\cdot 25 \times \cdot 75}{900}}$$

$$= \sqrt{\cdot 000175 + 0 \cdot 00021}$$

$$= \sqrt{\cdot 00385} = \cdot 0196$$

$$|z| = \frac{|P_1 - P_2|}{S.E_{P_1 - P_2}} = \frac{0 \cdot 30 - 0 \cdot 25}{\cdot 0196} = \frac{\cdot 05}{\cdot 0196} = 2 \cdot 55$$

At 5% level of significance, the critical value of Z for two tailed test $= 1 \cdot 96$.

Since, $|Z| = 2 \cdot 55 > 1 \cdot 96$, we reject H_0 and conclude that the difference is unlikely to be hidden due to simple sampling fluctuations.

Example 31. The results of IQ test of two groups of girls are as under :

Group of 120 Girls : Mean = 84, S.D. = 10

Group of 80 Girls : Mean = 81, S.D. = 12

Do they belong to the same population?

Solution. To test if two independent samples belong to the same population, we have to test :

(i) The equality of population means and

(ii) The equality of population standard deviations

(i) Let us take the null hypothesis that the mean of two groups of girls is not significant *i.e.* $H_0 : \mu_1 = \mu_2$

$x_1 = 120; n_2 = 80, s_1 = 10, s_2 = 12, \bar{X}_1 = 84, \bar{X}_2 = 81$

$$SE_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(10)^2}{120} + \frac{(12)^2}{80}} = \sqrt{\frac{100}{120} + \frac{144}{80}}$$

$$= \sqrt{0.833 + 1.80} = \sqrt{2.633} = 1.62$$

Applying Z-statistic

$$|Z| = \frac{|\bar{X}_1 - \bar{X}_2|}{SE_{\bar{X}_1 - \bar{X}_2}} = \frac{|84 - 81|}{1.62} = \frac{3}{1.62} = 1.85$$

At 5% level of significance, the critical value of Z for two tailed test = 1.96

Since, the calculated value of |Z| is less than the critical value of Z of 5% l.o.s, we accept the null hypothesis and conclude that there is no significant difference in the mean of two groups of girls.

(ii) Let us take the hypothesis that there is no difference in the S.Ds of the two groups of girls, *i.e.*

$H_0' : \sigma_1 = \sigma_2$

$$SE_{\sigma_1-\sigma_2} = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}} = \sqrt{\frac{(10)^2}{2\times 120} + \frac{(12)^2}{2\times 80}}$$

$$= \sqrt{\frac{100}{240} + \frac{144}{160}} = \sqrt{0.416 + 0.90}$$

$$= \sqrt{1.316} = 1.147$$

$$|Z| = \frac{|\sigma_1 - \sigma_2|}{SE_{\sigma_1-\sigma_2}} = \frac{|10-12|}{1.147} = 1.74$$

Since, the calculated value of $|Z|$ is less than the critical value of Z at 5% l.o.s, we accept H_0 and conclude that there is no significant difference in the standard deviations of two groups of girls.

Since, both the Hypothesis $H_0 : \mu_1 = \mu_2$ and $H_0' : \delta_1 = \delta_2$ are accepted, we may say that the two groups of girls belong to the same population.

IMPORTANT FORMULAE

(i) Test of Hypothesis about population mean :

$$Z = \frac{\overline{X} - \mu}{S.E_{\overline{X}}}$$

Where $S.E_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ or $\frac{s}{\sqrt{n}}$

(ii) Test of Hypothesis about the difference between two population mean :

$$|Z| = \frac{\overline{X}_1 - \overline{X}_2}{S.E_{\overline{X}_1 - \overline{X}_2}}$$

Where $S.E_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

(iii) Test of Hypothesis about population proportion :

$$|Z| = \frac{p - P}{S.E_p}$$

Where $S.E_p = \sqrt{\frac{PQ}{n}}$

(iv) Test of Hypothesis about difference between two population proportions :

$$|Z| = \frac{p_1 - p_2}{S.E_{p_1-p_2}}$$

Where $S.E_{p_1-p_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ and $q = (1 - p)$

(p = polled estimate)

QUESTIONS

1. What is statistical hypothesis ? Discuss the procedure of testing a statistical hypothesis.
2. Explain the following :
   (a) Null hypothesis and alternative hypothesis.
   (b) Type I and Type II Errors.
   (c) One tail and two tailed test.
   (d) Acceptance and rejection regions
   (e) Level of significance.
3. How do you test the equality of two population means and two population proportions in case of large samples ?
4. Explain how the size of α (alpha) is determined for testing an hypothesis.
5. Outline the procedure for large sample tests and discuss their theoretical basis.
6. Discuss the applications of large sample tests.
7. Distinguish between a null hypothesis and an alternative hypothesis. Use example to explain the nature of null and alternative hypothesis in cases of one and two tailed tests.
8. Explain the concept of level of significance in test of hypothesis. Discuss briefly the procedure of testing a hypothesis.
9. State the null and alternative hypothesis regarding the population mean that lead to :
   (i) Left-tailed test (ii) Right-tailed test (iii) Two tailed test;
   [Hint : Left tailed test : $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$, Right-tailed : $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$; Two tailed test $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$].

❀❀❀

4

Tests of Hypothesis – Small Sample Tests

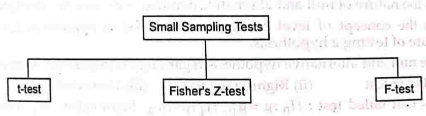
INTRODUCTION

The various tests of hypothesis or test of significance discussed in the previous chapter were related to large samples. Tests of hypothesis relating to large samples are based on two assumptions : (a) Sampling distribution of a statistic approaches a normal distribution whether or not the population distribution is normal or not, (b) The values of the sample statistics are sufficient close to the population values. But these two basic assumptions of large samples don't hold good in case of small samples. Therefore, it becomes necessary to make a separate study of small sample tests. Various small sampling tests are available; the most common being them are :

(1) *t*—test

(2) Fisher's Z—test

(3) *F*—test



(1) *t*—test

t—test is a small sample test. It was developed by William Gosset in 1908. He published this test under the pen name of "Student". Therefore, it is known as Student's *t*–test. For applying *t*–test, the value of *t*–statistic is computed. For this, the following formula is used :

$$t = \frac{\text{Deviation from the population parameter}}{\text{Standard Error of the sample statistic}}$$

The calculated value of *t* is compared with the table value of *t* for given degrees of freedom at certain specified level of significance.

Applications/Uses of *t*—test : The following are some important applications of *t*–test :

(1) Test of hypothesis about the population

(2) Test of hypothesis about the difference between the two means in case of independent samples.

(3) Test of hypothesis about the difference between two means with dependent samples.

(4) Test of hypothesis about an observed coefficient of correlation.

(1) Test of hypothesis about the population mean (σ unknown and sample size is small) : A random sample of size n (≤ 30) drawn from a normal population has a sample mean $\overline{X}$. To test the hypothesis that the population mean μ has a specified value μ_0 when population standard deviation σ is not known, and $n < 30$, we use t-test and the appropriate test statistic t to be used is :

$$t = \frac{\overline{X} - \mu}{S} \cdot \sqrt{n}$$

where, S = Modified S.D. of the sample.

n = Size of the sample

Procedure : The following steps are taken while testing the hypothesis about the population mean μ :

(1) Set up the null hypothesis $H_0 : \mu = \mu_0$ *i.e.*, the population mean is μ_0

Alternative hypothesis $H_1 : \mu \neq \mu_0$ ($\Rightarrow$ Two tailed test)

or $H_1 : \mu > \mu_0$ or $\mu < \mu_0$ ($\Rightarrow$ One tailed test)

(2) Thereafter the value of modified standard deviation of the sample (S) is computed by using any of the following formulae :

(a) When deviations are taken from the actual mean :

$$S = \sqrt{\frac{\Sigma (X - \overline{X})^2}{n-1}}$$

(b) When actual mean is in fraction and deviations are taken from assumed mean :

$$S = \sqrt{\frac{\Sigma d^2 - (\overline{d})^2 \times n}{n-1}}$$

Where, $d = X - A$ (deviation from the Assumed Mean)

(c) When sample standard deviation is given :

$$S = \sqrt{\frac{n}{n-1} s^2} \quad \text{or} \quad \sqrt{\frac{n}{n-1}} \cdot s$$

(3) The values of $\overline{X}$, μ and S are substituted in the above stated formula.

(4) Degrees of freedom are worked out by using the following formula :

Degrees of Freedom $= v = n - 1$

(5) Obtain the table value of t at the stated level of significance (α) and for given degrees of freedom from the table of "values t-statistics"

(6) If the computed value of $t = |t| <$ table value of t at a level of significance α, then we accept the null hypothesis.

If the computed value of $t >$ table value of t at a level of significance α, then we reject the null hypothesis and accept the alternative hypothesis.

Important Note : Since 't' distribution is symmetric about $t = 0$, the significant values at a level of significance 'α' for a single tailed (right or left) test can be obtained from the table of two tailed test by looking the value at the level of significance 2α. For example :

t^v (0·05) for single tailed test $= t^v$ (0·10) for two tailed test.

t^v (0·01) for single tailed test $= t^v$ (0·02) for two tailed test.

t^v (0·025) for single tailed test $= t^v$ (0·05) for two tailed test.

t^v (0·005) for single tailed test $= t^v$ (0·01) for two tailed test.

The testing procedure of population mean is clarified from the following examples :

Example 1. **A group of 5 patients treated with medicine *A* weights : 42, 39, 48, 60 and 41 kg. In the light of the above data, discuss the suggestion that mean weight of the population is 48 kg. Test at 5% level of significance.**
(Given the table value of *t* for 4 d.f. at 5% level is 2·776)

Solution. Let us take the null hypothesis that mean weight in the population is 48 *i.e.*,

$H_0 : \mu = 48$ and $H_1 : \mu \neq 48$ ($\Rightarrow$ Two tailed test)

Weight (X)	$\overline{X} = 46$ $(X - \overline{X})$	$(X - \overline{X})^2$
42	−4	16
39	−7	49
48	2	4
60	14	196
41	−5	25
$\Sigma X = 230, n = 5$		$\Sigma(X - \overline{X})^2 = 290$

$$\overline{X} = \frac{\Sigma X}{n} = \frac{230}{5} = 46$$

$$S = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n-1}} = \sqrt{\frac{290}{5-1}} = \sqrt{\frac{290}{4}} = 8{\cdot}514$$

Applying *t*–test :

$$t = \frac{\overline{X} - \mu}{S} . \sqrt{n}$$

$$|t| = \frac{46 - 48}{8{\cdot}514} . \sqrt{5} = \frac{2 \times 2{\cdot}236}{8{\cdot}514} = \frac{4{\cdot}472}{8{\cdot}514} = 0{\cdot}525$$

Degrees of freedom $= v = n - 1 = 5 - 1 = 4$

For $v = 4, t_{0{\cdot}05}$ for two tailed test $= 2{\cdot}776$

Since, the calculated value of t is less than the table value, we accept the null hypothesis and therefore conclude that the mean weight in the population is 48 kg.

Example 2. **A random sample of 9 boys had heights (inches) : 45, 47, 50, 52, 48, 47, 49, 53 and 51. In the light of the data, discuss the suggestion that the mean height in the population is 47·5.**
(Give the table value of *t* for 8 d.f. at 5% level = 2·306).

Solution. Let us take the null hypothesis that the mean height in the population is 47·5 *i.e.*,

$H_0 : \mu = 47{\cdot}5$ and $H_1 = \mu \neq 47{\cdot}5$ ($\Rightarrow$ Two tailed test)

Height X	$A = 49$ $d = (X - A)$	d^2
45	−4	16
47	−2	4
50	+1	1
52	+3	9
48	−1	1
47	−2	4
49	0	0

53	+4	16
51	+2	4
$\Sigma X = 442$	$\Sigma d = 1$	$\Sigma d^2 = 55$

$$\overline{X} = \frac{\Sigma X}{n} = \frac{442}{9} = 49{\cdot}11$$

Since, the actual mean is in fraction, we should take deviations from assumed mean (49) to simplify the calculations.

$$\overline{d} = \frac{\Sigma d}{n} = \frac{1}{9} = 0{\cdot}11$$

$$S = \sqrt{\frac{\Sigma d^2 - n(\overline{d})^2}{n-1}} = \sqrt{\frac{55 - 9 \times (\cdot 11)^2}{9-1}} = \sqrt{\frac{54{\cdot}8911}{8}} = \sqrt{6{\cdot}8613} = 2{\cdot}62$$

Applying *t*–test :

$$t = \frac{\overline{X} - \mu}{S} . \sqrt{n} = \frac{49{\cdot}11 - 47{\cdot}5}{2{\cdot}62} \sqrt{9}$$

$$= \frac{1{\cdot}61 \times 3}{2{\cdot}62} = \frac{4{\cdot}83}{2{\cdot}62} = 1{\cdot}843$$

Degrees of freedom $= v = 9 - 1 = 8$

For $v = 8, t_{0{\cdot}05}$ for two tailed test $= 2{\cdot}306$

Since, the calculated value of t is less than the table value, we accept the null hypohtesis and therefore conclude that the mean height in the population is 47·5 inches.

Aliter : The vlaue of S can also be calculated as :

$$\Sigma(X - \overline{X})^2 = \Sigma d^2 - \frac{(\Sigma d^2)}{n} = 55 - \frac{(1)^2}{9} = 54{\cdot}89$$

$$\therefore \quad S = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n-1}} = \sqrt{\frac{54{\cdot}89}{9-1}}$$

$$= \sqrt{\frac{54{\cdot}89}{8}} = 2{\cdot}619 \simeq 2{\cdot}62$$

Example 3. **A random sample of 9 values from a normal population showed a mean of 41·5 inches and the sum of squares of deviations from this mean equal to 72 inches. Show whether the assumption of mean of 44·5 inches in the population is reasonable. (For $v = 8$, $t_{.05} = 2{\cdot}776$)**

Solution. $\overline{X} = 41{\cdot}5, \mu = 44{\cdot}5, n = 9, \Sigma(X - \overline{X})^2 = 72$

Let us take the null hypothesis that the population mean is 44·5 *i.e.*, $H_0 : \mu = 44{\cdot}5$ and $H_1 : \mu \neq 445$

$$S = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n-1}} = \sqrt{\frac{72}{9-1}} = \sqrt{\frac{72}{8}} = \sqrt{9} = 3$$

Applying t–test :

$$|t| = \frac{\overline{X} - \mu}{S} \cdot \sqrt{n}$$

$$|t| = \frac{|41 \cdot 5 - 44 \cdot 5|}{3} \times \sqrt{9} = 3$$

Degrees of freedom $= v = n - 1 = 9 - 1 = 8$

For $v = 8$, $t_{0 \cdot 05}$ for two tailed test $= 2 \cdot 306$

Since, the calculated value of $|t|$ > the table value of t, we reject the null hypothesis. We conclude that the population mean is not equal to $44 \cdot 5$.

Example 4. **Sixteen oil tins are taken at random from an automatic filling machine. The mean weight of the tins is 14·5 kg with a standard deviation of 0·40 kg. Does the sample mean differ significantly from the intended weight of 16 kg ?**

Solution. We are given : $n = 16$, $\overline{X} = 14 \cdot 5$, $s = 0 \cdot 40$, $\mu = 16$

Let us take the null hypothesis that there is no difference between the sample mean and intended mean *i.e.,*

$H_0 : \mu = 16$ and $H_1 : \mu \neq 16$ ($\Rightarrow$ Two tailed test)

$$S = \sqrt{\frac{n}{n-1} s^2} = \sqrt{\frac{16}{16-1} \times (\cdot 4)^2} = \sqrt{\frac{2 \cdot 56}{15}} = 0 \cdot 413$$

$$\left[\because s = \sqrt{\frac{\Sigma (X - \overline{X})^2}{n-1}} \right]$$

Applying t–test :

$$\therefore \quad |t| = \frac{|\overline{X} - \mu|}{S} \cdot \sqrt{n} = \frac{|14 \cdot 5 - 16|}{0 \cdot 413} \times \sqrt{16} = 14 \cdot 52$$

Degrees of freedom $= v = n - 1 = 16 - 1 = 15$

For $v = 15$, $t_{0 \cdot 05}$ for two tailed test $= 2 \cdot 131$

Since, the calculated value of $|t| = 14 \cdot 52$ > the table value of t, we reject the null hypothesis. It means that the sample mean differ significant from the intended mean 16 kg.

Aliter : The value of can also be calculated using the female;

$$|t| = \frac{|\overline{X} - \mu|}{s} \cdot \sqrt{n-1}$$

$$= \frac{|14 \cdot 5 - 16|}{0 \cdot 4} \times \sqrt{16 - 1} = \frac{1 \cdot 5 - 3 \cdot 872}{0 \cdot 4} = 14 \cdot 52$$

Example 5. **A consumer testing agency while examining a new automobile for gasoline mileage performance found that 12 readings of miles covered per gallon under the normal conditions resulted in a average of 16 miles per gallon with a standard deviation of 1·8 miles. Do the sample results support the manufacturer's claim that the new automobile gives a performance of more than 15 miles per gallon ? Use $\alpha = 0 \cdot 10$, assuming that the distribution of mileage performance per gallon is approximately normal.**

Solution. We are given : $n = 12$, $\overline{X} = 16$, $s = 1 \cdot 8$, $\mu = 15$

Null hypothesis $H_0 : \mu = 15$

Alternative hypothesis $H_1 : \mu > 15$ ($\Rightarrow$ Right tailed test)

$$S = \sqrt{\frac{n}{n-1} s^2} = \sqrt{\frac{12}{12-1} \times (1 \cdot 8)^2} = 1 \cdot 88$$

Applying t–test :

$$\therefore \quad t = \frac{\overline{X} - \mu}{S} \cdot \sqrt{n} = \frac{16 - 15}{1 \cdot 88} \sqrt{12} = 1 \cdot 84$$

Degrees of freedom $= v = n - 1 = 12 - 1 = 11$

For $v = 11$, $t_{0 \cdot 10}$ for one tailed test $= 1 \cdot 363$

Since, the calculated value of $|t| = 1 \cdot 84$ > table vlaue of t at 10% level of significance, the null hypothesis is rejected and conclude that the sample data support the manufacturer's claim of improved mileage performance.

Aliter : The value of t can also be calculated by using the formula :

$$|t| = \frac{|\overline{X} - \mu|}{s} \cdot \sqrt{n-1}$$

$$= \frac{16 - 15}{1 \cdot 8} \times \sqrt{12 - 1} = 1 \cdot 84.$$

EXERCISE – 1

1. Six boys are chosen at random from a school and their heights are found to be in inches : 63, 63, 64, 66, 60, 68. Discuss the suggestion that the mean height in the population is 65 inches ? (Given the table value of t for 5df at 5% level is 2·57).
   [Ans. $t = 0 \cdot 888$, Accept H_0]
2. Ten specimens of copper wires drawn from two large lots have the following breaking strength (in kg. wt) :
   578, 572, 570, 568, 572, 578, 570, 572, 596, 584
   Test whether the mean breaking strength of the lot may be taken to be 580 kg. wt.
   [Ans. $t = 1 \cdot 481$, Accept H_0]
3. 10 students are selected at random from a college and their marks in Hindi are found to be as follows :
   71, 72, 73, 75, 76, 77, 78, 79, 79, 80
   In the light of the marks, test whether, the average marks in Hindi of the college are 75. (The value of t at 5% for $v = 9$ is $2 \cdot 262$) **[Ans. $t = 1 \cdot 054$, Accept H_0]**
4. A random sample of 16 values from a normal population showed a mean of 41·5 and the sum of squares of deviations from the mean equal to 135. Can it be assumed that the mean of the population is 43·5. Use 5% level of significance. **[Ans. $t = 2 \cdot 67$, Reject H_0]**
5. Ten cartons are taken from an automatic filling machines. The mean weight is 11·802 and standard deviation 0·1502. Does the sample mean differ significanctly from the intended weight 11·6 ? (Given for 9 degree of freedom at 5% level is 2·262).
   [Ans. $t = 4 \cdot 034$, Reject H_0]

6. A random sample of size 10 has mean as 40 and standard devaition=5. Can this sample be regarded as taken from the population having 42 as mean ? (For $v=9, t_{0.05}=2 \cdot 262$).
[Ans. $t=1 \cdot 2$, Accept H_0]

7. A soft drink vending machine is set to dispense 8 ounces per cup. If the machine is tested 9 times yielding a mean cup fill of 8·2 ounces with a standard deviation of 0·3 ounces, what can we conclude about the hull hypothesis of $\mu=8$ ounces against the alternative hypothesis of $\mu>8$ ounces at $\alpha=0 \cdot 01$.
[Ans. $|t|=1 \cdot 885$ Accept H_0]

(2) Test of hypothesis about diffeence between two means in case of independent samples: Let two independent random samples of sizes n_1 and n_2 ($n_1 < 30$ and $n_2 < 30$) be drawn from two normal populations with means μ_1 and μ_2 and equal standard deviation ($\sigma_1=\sigma_2=\sigma_3$). To test whether the two populations means are equal or whether the difference $\overline{X}_1-\overline{X}_2$ is significant, we use *t*-test and the appropriate test statistic '*t*' to be used is :

$$t=\frac{\overline{X}_1-\overline{X}_2}{S} \cdot \sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

Procedure : The following steps are taken while testing the hypothesis about difference between two means :

(1) Set up the null hypothesis $H_0 : \mu_1=\mu_2$ *i.e.,* there is no significant difference between two populations means :
Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ (⇒Two tailed test)
or $H_1 : \mu_1 > \mu_2$ or $\mu_1 < \mu_2$ (⇒One tailed test)

(2) If the population standard deviation σ_1 and σ_2 are equal *i.e.,* $\sigma_1=\sigma_2=\sigma$, the values of S is computed by using any of the following formula :

(a) When deviations are taken from actual mean :

$$S=\sqrt{\frac{\Sigma(X_1-\overline{X}_1)^2+\Sigma(X_2-\overline{X}_2)^2}{n_1+n_2-2}}$$

(b) When actual mean is in fraction and deviations are taken from the assumed mean:

$$S=\sqrt{\frac{\Sigma(X_1-A_1)^2+\Sigma(X_2-A_2)^2-n_1(\overline{X}_1-A_1)^2-n_2(\overline{X}_2-A_2)^2}{n_1+n_2-2}}$$

Where, A_1 and A_2 are the assumed means of two samples

(c) When standard deviations of two samples s_1 and s_2 are given :

$$S=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

(3) The values of $\overline{X}_1, \overline{X}_2$ n_1 and n_2 and S are substituted in the above stated formula.

(4) Degrees of freedom are worked out by using the following formula :
Degrees of freedom $=v=n_1+n_2-2$

The other steps such as (i) level of significance (ii) table value of t (iii) decision making for testing the difference between the two means are the same as those given in testing of the hypothesis.

The testing procedure of the differnece of two population means is clarified by the following examples :

Example 6. **In a test given to two groups of students, the marks obtained are as follows:**

First Group :	18	20	36	50	49	36	34	49	41
Second Groups :	29	28	26	35	30	44	46		

Examine the significance of difference between the mean marks secured by students of the above two groups. (The value of *t* at 5% level for 4 d.f. = 2·14)

Solution

Let us take the hypothesis that there is no significant difference in the mean marks of the two groups of students *i.e.,*

$H_0 : \mu_1=\mu_2$ and $H_1 : \mu_1 \neq \mu_2$ (⇒Two tailed test)

Group I X_1	$\overline{X}_1=37$ $(X_1-\overline{X}_1)$	$(X_1-\overline{X}_1)^2$	Group II X_2	$\overline{X}_2=34$ $(X^2-\overline{X}_2)$	$(X_2-\overline{X}_2)^2$
18	−19	361	29	−5	25
20	−17	289	28	−6	36
36	−1	1	26	−8	64
50	+13	169	35	+1	1
49	+12	144	30	−4	16
36	−1	1	44	+10	100
34	−3	9	46	+12	144
49	+12	144			
41	+4	16			
$\Sigma X_1=333$ $n_1=9$		$\Sigma(X_1-\overline{X}_1)^2$ =1,134	$\Sigma X_2=238$ $n_2=7$		$\Sigma(X^2-\overline{X}_2)^2$ = 386

$$\overline{X}_1=\frac{\Sigma X_1}{n_1}=\frac{333}{9}=37, \quad \overline{X}_2=\frac{\Sigma X_2}{n_2}=\frac{238}{7}=34$$

$$S=\sqrt{\frac{\Sigma(X_1-\overline{X}_1)^2+(X_2-\overline{X}_2)^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{1134+386}{9+7-2}}=\sqrt{\frac{1520}{14}}=\sqrt{108 \cdot 571}=10 \cdot 42$$

Applying *t*–test :

$$t=\frac{\overline{X}_1-\overline{X}_2}{S}\sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$=\frac{37-34}{10 \cdot 42}\sqrt{\frac{9 \times 7}{9+7}}=\frac{3}{10 \cdot 42} \times 1 \cdot 984=0 \cdot 571$$

Degrees of freedom $=v=n_1+n_2-2=9+7-2=14$

For $v=14, t_{0.05}$ for two tailed test $=2 \cdot 14$.

Since, the calculated value of t is less than the table value, we accept the null hypothesis and conclude that the mean marks of the students of two groups do not differ significantly.

Example 7. **Two independent samples of 8 and 7 items gave the following values :**

Sample A :	9	11	13	11	15	9	12	14
Sample B :	10	12	10	14	9	8	10	

Examine whether the differnce between the means of two samples is significant at 5% level.

Solution. Let us take the null hypothesis that there is no significant difference in the two means *i.e.,*

$H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$ (⇒ Two tailed test)

X_1	$A_1 = 12$ $(X_1 - A_1)$	$(X_1 - A_1)^2$	X_2	$A_2 = 10$ $(X_2 - A_2)$	$(X_2 - A_2)^2$
9	−3	9	10	0	0
11	−1	1	12	+2	4
13	+1	1	10	0	0
11	−1	1	14	+4	16
15	+3	9	9	−1	1
9	−3	9	8	−2	4
12	0	0	10	0	0
14	+2	4			
$\Sigma X_1 = 94$ $n_1 = 8$ $\overline{X}_1 = \frac{94}{8} = 11 \cdot 75$		$\Sigma(X_1 - A_1)^2 = 34$	$\Sigma X_2 = 73$ $n_2 = 7$ $\overline{X}_2 = \frac{73}{7} = 10 \cdot 43$		$\Sigma(X_2 - A_2)^2 = 25$

Since, the actual means are not whole numbers, we take 12 as assumed for X_1 and 10 as assumed for X_2 :

$$S = \sqrt{\frac{\Sigma(X_1 - A_1)^2 + \Sigma(X_2 - A_2)^2 - n_1(\overline{X}_1 - A_1)^2 - n_2(\overline{X}_2 - A_2)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{34 + 25 - 8(11 \cdot 75 - 12)^2 - 7(10 \cdot 43 - 10)^2}{8 + 7 - 2}}$$

$$= \sqrt{\frac{34 + 25 - 0 \cdot 5 - 1 \cdot 2943}{13}} = \sqrt{4 \cdot 4004} = 2 \cdot 0977$$

Applying *t*-test,

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{11 \cdot 75 - 10 \cdot 43}{2 \cdot 0977} \times \sqrt{\frac{8 \times 7}{8 + 7}}$$

$$= \frac{1.32}{2.0977} \times 1 \cdot 932 = 1 \cdot 2158$$

Degrees of freedom $= v \Rightarrow 8 + 7 - 2 = 13$

For $v = 13$, $t_{0 \cdot 05}$ for two tailed test $= 2 \cdot 16$

Since, the calculated value of t is less than the table value, we accept the null hypothesis and conclude that there is no significant difference in the means of the two samples.

Aliter : The value of S can also be calculated as:

$$\Sigma(X_1 - \overline{X}_1)^2 = \Sigma d_1^2 - (\Sigma d_1)^2 / n_1 = 34 - (2)^2 / 8 = 33 \cdot 5$$

$$\Sigma(X_2 - \overline{X}_2)^2 = \Sigma d_2^2 - (\Sigma d_2)^2 / n_2 = 25 - (3)^2 / 7 = 23 \cdot 71$$

$$S = \sqrt{\frac{\Sigma(X_1 - \overline{X}_1)^2 + \Sigma(X_2 - \overline{X}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{33 \cdot 5 + 23 \cdot 71}{8 + 7 - 2}}$$

$$= \sqrt{\frac{57 \cdot 21}{13}} = 2 \cdot 097$$

Example 8. **The mean life of a sample of 10 electric light bulbs was found to be 1456 hours with $s = 423$ hours. A second sample of 17 bulbs chosen from a differnt batch showed a mean life of 1280 hours with $s = 398$ hours. Is there a significant differnce between the means of the two batches ? Here s is the standard deviation.**

Solution. We are given : $n_1 = 10$, $\overline{X}_1 = 1456$, $s_1 = 423$

$n_2 = 17$, $\overline{X}_2 = 1280$, $s_2 = 398$

Let us assume that there is no significant difference between the means of two batches, *i.e.,*

Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ (⇒ Two tailed test)

$$S = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(10 - 1) \times 423^2 + (17 - 1) \times 398^2}{10 + 17 - 2}}$$

$$= \sqrt{\frac{1610361 + 2534464}{25}}$$

$$= \sqrt{\frac{4144825}{25}} = 407 \cdot 17$$

Applying *t*–test :

$$|t| = \frac{\overline{X}_1 - \overline{X}_2}{S} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$= \frac{1456 - 1280}{407 \cdot 17} \times \sqrt{\frac{10 \times 17}{10 + 17}}$$

$$= \frac{176}{407 \cdot 17} \times 2 \cdot 50 = \frac{440}{407 \cdot 17} = 1 \cdot 0806$$

Degrees of freedom $= v = n_1 + n_2 - 2 = 10 + 17 - 2 = 25$

For $v = 25$, $t_{.05}$ for two tailed test $= 2 \cdot 06$

Since, the calculated value of $|t| = 1 \cdot 0806 <$ the table value of t, we accept H_0 and conclude that there is no significant difference between the means of two batches.

Example 9. **Two salesmen A and B are wroking in a certain district. From a sample survey conducted by the Head Office, the following results were obtained. State whether there is any significance difference in the average sales between the two salesmen :**

	A	B
No. of sales	20	18
Average	170	205
Standard deviation	20	25

Solution. We are given : $n_1 = 20,\ \bar{X}_1 = 170,\ s_1 = 20 \Rightarrow s_1^2 = 400$

$n_2 = 18,\ \bar{X}_2 = 205,\ s_2 = 25 \Rightarrow s_2^2 = 625$

Let us assume that there is no significant differnce in the average sales between two salesmen *i.e.*, $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ ($\Rightarrow$ Two tailed test)

$$S = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(20-1) \times 400 + (18-1) \times 625}{20 + 18 - 2}}$$

$$= \sqrt{\frac{7600 + 10625}{36}} = \sqrt{\frac{18225}{36}} = 22 \cdot 5$$

Applying *t*-test :

$$|t| = \frac{\bar{X}_1 - \bar{X}_2}{S} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$|t| = \frac{|170 - 205|}{22 \cdot 5} \cdot \sqrt{\frac{20 \times 18}{20 + 18}}$$

$$= \frac{35}{22 \cdot 5} \times 3 \cdot 077 = 4 \cdot 786$$

Degrees of freedom $= v = n_1 + n_2 - 2 = 20 + 18 - 2 = 36$

For $v = 36$, $t_{.05}$ for two tailed test $= 1 \cdot 96$

Since, the calculated value of $|t| = 4 \cdot 786 >$ the table value of t, we reject H_0 and conclude that there is a significant difference in the sales between two salesmen.

Example 10. **On an examination in statistics, 10 students in one class showed a mean grade of 80 with a standard deviation of 8, while 12 students in another class showed a mean grade of 76 with a standard deviation of 10. Using $\alpha = .01$ level of significance, determine whether the first group is superior to the second group.**

Solution. We are given : $n_1 = 10,\ \bar{X}_1 = 80,\ s_1 = 8 \Rightarrow s_1^2 = 64$

$n_2 = 12,\ \bar{X}_2 = 76,\ s_2 = 10 \Rightarrow s_2^2 = 100$

Let us assume that mean grade of two groups don't differ significantly *i.e.*, $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 > \mu_2$ ($\Rightarrow$ Right tailed test)

$$S = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(10-1) \times 64 + (12-1) \times 100}{10 + 12 - 2}}$$

$$= \sqrt{\frac{576 + 1100}{20}} = \sqrt{\frac{1676}{20}} = \sqrt{83 \cdot 8} = 9 \cdot 15$$

Applying *t*-test :

$$|t| = \frac{\bar{X}_1 - \bar{X}_2}{S} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$= \frac{80 - 76}{9 \cdot 15} \times \sqrt{\frac{10 \times 12}{10 + 12}}$$

$$= \frac{4}{9 \cdot 15} \times 2 \cdot 33 = 1 \cdot 0185 \approx 1 \cdot 02$$

Degrees of freedom $= v = n_1 + n_2 - 2 = 12 + 10 - 2 = 20$

For $v = 20$, $t_{.01}$ for one tailed test $= 2 \cdot 528$

Since, the calculated value of $|t| <$ the table value of t, we accept H_0 and conclude that the mean of the two groups are the same *i.e.*, first group is not superior to the second group.

Aliter : This question can also be solved using two–tailed test. Let us have; $H_0 : \mu_1 = \mu_2$ *i.e.*, the first group is not superior to the second group.

And $H_1 : \mu_1 \neq \mu_2$ ($\Rightarrow$ Two tailed test)

In the light of the facts given, the relevant test statistic '*t*' computed is :

$$|t| = 1 \cdot 0185 \approx 1 \cdot 02$$

Degrees of freedom $= v = n_1 + n_2 - 2 = 12 + 10 - 2 = 20$

For $v = 20$, $t_{.01}$ for two tailed test $= 2 \cdot 845$.

Since, the calculated value of $|t| <$ table value of t, we accept H_0 and conclude that the first group is not superior to the second group.

Example 11. **The means of two random samples of sizes 9 and 7 are 196·42 and 198·82 respectively. The sum of the squares of the deviations from the means are 26·94 and 18·73 respectively. Can the samples be considered to have been drawn from the same normal population ?**

Solution. Given : $n_1 = 9,\ \bar{X}_1 = 196 \cdot 42,\ \Sigma(X_1 - \bar{X}_1)^2 = 26 \cdot 94$

$n_2 = 7,\ \bar{X}_2 = 198 \cdot 82,\ \Sigma(X_2 - \bar{X}_2)^2 = 18 \cdot 73$

Let us take the null hypothesis that the samples are drawn from the same normal population *i.e.*, $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$ (Two tailed test)

$$S = \sqrt{\frac{\Sigma(X_1 - \overline{X}_1)^2 + \Sigma(X_1 - \overline{X}_2)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{26\cdot94 + 18\cdot73}{9 + 7 - 2}} = \sqrt{\frac{45\cdot67}{14}} = \sqrt{3\cdot26} = 1\cdot81$$

Applying *t*–test :

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$|t| = \frac{|196\cdot42 - 198\cdot82|}{1\cdot81} \sqrt{\frac{9 \times 7}{9 + 7}}$$

$$= \frac{2\cdot40 \times 1\cdot98}{1\cdot81} = 2\cdot6254$$

Degrees of freedom $(v) = n_1 + n_2 - 2 = 9 + 7 - 2 = 14$

For $v = 14$, $t_{.05}$ for two tailed test $= 2\cdot145$

Since, the calculated value of t is greater than the table vlaue, we reject the null hypothesis and conclude that the difference is significant. Thus, both the samples have not been taken from same population.

Example 12. **A random sample of seven week old chickens reared on a high protein diet weight : 12, 15, 11, 16, 14, 14 and 16 ounces, another random sample of five chickens similarly treated except that they received a low protein diet weight : 8, 10, 14, 10 and 13 ounces. Test whether there is significant evidence that additional protein has increased the weight of chickens.**

(The table value of *t* for 10 degrees of freedom at 5% level of significance is 2.228)

Solution. Let us take the hypothesis that additional protein has not increased the weight of chickens, *i.e.*, $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$ (Two tailed test)

X_1	$\overline{X}_1 = 14$ $(X_1 - \overline{X}_1)$	$(X_1 - \overline{X}_1)^2$	X_2	$\overline{X}_2 = 11$ $(X_2 - \overline{X}_2)$	$(X_2 - \overline{X}_2)^2$
12	− 2	4	8	− 3	9
15	+ 1	1	10	− 1	1
11	− 3	9	14	+ 3	9
16	+ 2	4	10	− 1	1
14	0	0	13	+ 2	4
14	0	0			
16	+ 2	4			
$\Sigma X_1 = 98$ $n_1 = 7$		$\Sigma(X_1 - \overline{X}_1)^2 = 22$	$\Sigma X_2 = 55$ $n_2 = 5$		$\Sigma(X_2 - \overline{X}_2)^2 = 24$

$$\overline{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{98}{7} = 14 \quad \overline{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{55}{5} = 11$$

$$S = \sqrt{\frac{\Sigma(X_1 - \overline{X}_1)^2 + \Sigma(X_2 - \overline{X}_2)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{22 + 24}{7 + 5 - 2}} = \sqrt{\frac{46}{10}} = 2\cdot14$$

Applying *t*–test,

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$= \frac{14 - 11}{2\cdot14} \times \sqrt{\frac{7 \times 5}{7 + 5}}$$

$$= \frac{3}{2\cdot14} \times 1\cdot708 = 2\cdot394$$

Degrees of freedom $(v) = n_1 + n_2 - 2 = 7 + 5 - 2 = 10$

For $v = 10$, $t_{.05}$ for two tailed test $= 2\cdot228$

Since, the calculated value of t is more than the table value, the null hypothesis is rejected. Hence, there is a significant evidence that additional protein has increased the weight of chickens.

Aliter : This question can also be solved by using one tailed test.

Let us have, $H_0 : \mu_1 = \mu_2$ *i.e.*, additional protein has not increased the weight of chickens.

And $H_1 : \mu_1 > \mu_2$ (as we want to conclude that additional protein has increased the weight)

It is a one tailed test.

In the light of the facts given, the relevant test statistic computed is :

$$t = 2\cdot394$$

Degrees of freedom $= v = n_1 + n_2 - 2 = 7 + 5 - 2 = 10$

For $v = 10$, $t_{.05}$ for one tailed test $= 1\cdot812$.

Since, the calculated vlaue of $|t| = 2\cdot394$ is more than the table value, the null hypothesis is rejected. Hence, there is a significant evidence that additional protein has increased the weight of chickens.

EXERCISE – 2

1. The height of six randomly chosen soldiers are in inches : 76, 70, 68, 69, 69 and 68. Those of 6 randomly chosen sailors are 68, 64, 65, 69, 72 and 64. Discuss in the light of these data the suggestion that soldiers are on the average taller than sailors. Use *t*–test. (Table value of *t* at 5% level for 10 d.f. $= 2\cdot23$) **[Ans.** $t = 1\cdot66$, **Accept** H_0**]**
2. Below are given the gain in weights (lbs) of lions fed on two diets x_1 and x_2 :

Gain in weights (ibs)

Diet x_1 :	25	32	30	32	24	14	32			
Diet x_2 :	24	34	22	30	42	31	40	30	32	35

Test at 5% level whether the two diets differ as regards their effect of the mean increase in weight.
[Ans. $t = 1{\cdot}585$, Reject H_0]

3. Two kinds of fertilizers were applied to 15 plots. Other conditions remaining the same, the yields in quintals are given below :

Fertilizer I :	18	31	28	22	26	40	45	
Fertilizer II :	20	14	48	40	44	34	32	30

Examine the significance of difference between the mean yields due to different kinds of fertilizers.
[Ans. $t = 0{\cdot}535$, Accept H_0]

4. Test the sgnificance of the difference of means of the two samples at 5% level of significance from the following data :

	No. of Items	Mean	S.D.
Sample *A* :	6	40	8·0
Sample *B* :	5	50	10·0

(The table value of t for 9 d.f. at 5% level is 2·262) **[Ans.** $t = 1{\cdot}84$, Accept H_0]

5. Test the significance of the difference of the means of two samples at 5% level of significance from the following data :

	Sample Size	Mean	Variance
Sample *A* :	10	1000	100
Sample *B* :	12	1020	121

(The table value of t for 20 d.f. at 5% level is 2·086) **[Ans.** $t = 4{\cdot}42$, Reject H_0]

6. Two salesmen *A* and *B* are working in a certain district. From a sample survery conducted by the head office, the following results were obtained. State whether there is any significant difference in the average sales between the two salesmen.

	A	*B*
No. of sales	10	18
Average sales (Rs.)	170	205
Standard deviation (in Rs.)	20	25

[Ans. $t = 3{\cdot}79$, Reject H_0]

7. In a sample of 10 electric lamps the mean life and S.D. was found to be 1456 hours and 423 hours respectively. Another 17 rnadomly selected lamps had a mean life of 1200 hours and S.D. 398 hours. Is there any significant difference in their mean values ?
[Ans. $t = 1{\cdot}578$, Accept H_0]

8. Two salesmen are working in a shop. The number of items sold by them in a week are given as :

S_1	25	32	30	32	24	14	32
S_2	24	34	22	30	42	31	40

Test at 5% level whether the performance of two salesmen differ significantly.
($t_{0.05} = 2.179$ at df = 12)

9. Measurement performed on random samples of two kinds of cigarettes yielded the following results on their nicotine content (in mgs.) :

Brand A :	21.4	23.6	24.8	22.4	26.3
Brand B :	22.4	27.7	23.5	29.1	25.8

Assuming that nicotine content is distributed normally, test the hypothesis that Brand B has a higher nicotine content than Brand A.
(**Hint :** Use one tailed Test) **[Ans.** $|t| = 1{\cdot}315$, Accept H_0]

10. The nicotine contents in milligrams of two samples of tobacco were found to be as follows :

Sample A :	24	27	26	21	25	
Sample B :	27	30	28	31	22	26

Can it be said that two samples come from normal population having the same mean ?
[Ans. $|t| = 1{\cdot}92$, Accept H_0]

---

(3) Test of hypothesis about diffeence between two means with dependent samples (*i.e.,* paired data) or paired *t*-test : t–test is also used to test the hypothesis about difference between two means in case of paired data *i.e.,* when the sample items are the same but different situations are beings analysed. For example, performance of some students are taken down before and after extra coaching and we want to find the effectiveness of extra coaching. To test the significance of the difference between two means in two situations in case of paired data, the appropriate test statistic 't' to be used is :

$$t = \frac{\bar{d}}{S} \cdot \sqrt{n}$$

Where $\bar{d}$ = mean of the difference, *i.e.,* $\frac{\Sigma d}{n}$, n = size of the sample

S = standard deviation of the difference.

Procedure : The following steps are taken while testing the hypothesis of difference between two means in case of paired data :

(i) Set up the null hypothesis $H_0 : \bar{d} = 0$ or $\mu_1 = \mu_2$

Alternative hypothesis $H_1 : \bar{d} \neq 0$ or $\mu_1 \neq \mu_2$ (two tailed test)

or $H_1 : \bar{d} > 0$ *i.e.,* $\mu_1 < \mu_2$ or $\mu_2 > \mu_1$ (one tailed test)

(ii) Find the difference between each matched pair as :

$$d = I - II \quad \text{or} \quad II - I$$

(iii) Calculate the mean of the difference as :

$$\bar{d} = \frac{\Sigma d}{n}$$

(iv) The value of the standard deviation of the difference is computed by using the following formula :

$$S=\sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}}$$

or

$$S=\sqrt{\frac{\Sigma d^2-(\bar{d})^2\times n}{n-1}}$$

(4) The values of $\bar{d}$, S and n are substituted in the above stated formula.

(5) Degrees of freedom are worked by using the following formula :

Degrees of freedom $=v=n-1$

The other steps such as (i) level of significance, (ii) table value of t (iii) decision making for testing the difference between the two means are the same as given in test of hypothesis.

Example 13. **Ten students of M. Com. were given a test in the Business Statistics. They were imparted a month's special coaching and a second test was conducted at the end of it. The results were as follows :**

Students	Marks in Ist Test	Marks in IInd Test
1	36	32
2	40	42
3	38	30
4	36	24
5	42	18
6	38	64
7	40	32
8	46	40
9	58	52
10	62	38

Do the marks give an evidence that the students have benefited by extra coaching ?

(Given : $v=d.f.=9$, $t_{.05}=2\cdot202$)

Solution. This is a problem on paired observations from two dependent samples. Here the paired t–test is used.

Computation of Mean and S.D.

Students	Marks in Ist Test X_1	Marks in IInd Test (after extra coaching) X_2	$d=X_2-X_1$	d^2
1	36	32	– 4	16
2	40	42	2	4
3	38	30	– 8	64
4	36	24	– 12	144
5	42	18	– 24	576

6	38	64	26	676
7	40	32	– 8	64
8	46	40	– 6	36
9	58	52	– 6	36
10	62	38	– 24	576
$n=10$			$\Sigma d=-64$	$\Sigma d^2=2192$

Mean difference $=\bar{d}=\frac{\Sigma d}{n}=\frac{-64}{10}=-6\cdot4$

$$S=\sqrt{\frac{\Sigma d^2-(\bar{d})^2\times n}{n-1}}=\sqrt{\frac{2192-(-6\cdot4)^2\times10}{10-1}}$$

$$=\sqrt{\frac{2192-409\cdot6}{9}}=14\cdot07$$

Let us have the null hypothesis $H_0:\bar{d}=0$ *i.e.,* $\mu_2-\mu_1=0$ (*i.e.,* students are not benefited by extra coaching).

Alternative Hypothesis $H_1:\bar{d}>0$ *i.e.,* $\mu_2-\mu_1>0\Rightarrow\mu_2>\mu_1$ (*i.e.,* students are benefited by extra coaching)

Here one tailed test is to be applied

Aliter : The value of S can also be calculated as :

$$\Sigma(d-\bar{d})^2=\Sigma d^2-(\Sigma d)^2/n=2192-(-64)^2/10$$

$$=1782\cdot4$$

$$S=\sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}}=\sqrt{\frac{178\cdot4}{9}}=14\cdot07$$

Applying t–test :

$$|t|=\frac{\bar{d}}{S}\cdot\sqrt{n}$$

$$=\frac{-6\cdot4}{14\cdot07}\times\sqrt{10}=\frac{20\cdot238}{14\cdot07}=1\cdot438$$

Degrees of freedom $=v=n-1=10-1=9$

For $v=9$, $t_{.05}$ for one tailed test $=1\cdot833$

Since the calculated value of $|t|=1\cdot438$ in less than the table value of t, we accept H_0 and conclude that extra coaching has not benefited the students.

Aliter : This question can also be solved using two tailed test :

Let us take the hypothesis that there is no difference in the marks before and after special coaching *i.e.,* extra coaching has not benefitted the students *i.e.,*

$H_0:\bar{d}=0$

And $H_1:\bar{d}\neq0$ ($\Rightarrow$ Two tailed test)

In the light of the facts given, the relevant test statistics 't' computed is

$$|t| = 1 \cdot 438$$

For $v = 9$, $t_{.05}$ for two tailed test $= 2 \cdot 262$

Sincet he calculated value of $|t| = 1 \cdot 438$ is less than the table value of t, we accept H_0 and conclude that extra coaching has not benefited the students.

Example 14. **A certain medicine given to each of the 9 patients resulted in the following increase in blood pressure :**

7, 3, −1, 4, −3, 5, 6, −4, −1

Can it be concluded that the medicine will, in general, be accompained by an increase in blood pressure ? (Given $t_{.05}$ (8) = 2·0306)

Solution. Let us assume that the medicine does not increase the blood pressure *i.e.,*

$H_0 : \bar{d} = 0$ *i.e.,* $\mu_1 = \mu_2$ and $H_1 = \bar{d} \neq 0$ (⇒ Two tailed test)

d	7	3	−1	4	−3	5	6	−4	−1	$\Sigma d = 16$
d^2	49	9	1	16	9	25	36	16	1	$\Sigma d^2 = 162$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{16}{9} = 1 \cdot 778$$

$$S = \sqrt{\frac{\Sigma d^2 - (\bar{d})^2 \times n}{n-1}} = \sqrt{\frac{162 - (1 \cdot 778)^2 \times 9}{9-1}} = 4 \cdot 086$$

$$t = \frac{\bar{d}}{S} \cdot \sqrt{n} = \frac{1 \cdot 778}{4 \cdot 086} \times \sqrt{9} = 1 \cdot 30$$

Degrees of freedom $(v) = n - 1 = 9 - 1 = 8$.

For $v = 8$ d.f., $t_{0.05}$ for two tailed test $= 2 \cdot 0306$

Since, the calculated value of $|t| = 1 \cdot 30$ is less than the table value of t, we accept the null hypothesis and conclude that the medicine in general does not increase the blood pressure.

Aliter : This question can also be solved using one tailed test.

$H_0 : \bar{d} \leq 0$ *i.e.,* there is no increase in blood pressure after the medicine.

$H_1 : \bar{d} > 0$ (⇒ One tailed test)

In the light of the facts given, the relevant test-statistics 't' computed is :

$$|t| = \frac{\bar{d}}{S} \cdot \sqrt{n} = 1 \cdot 30$$

For d.f. $= v = 8$, $t_{.05}$ for one tailed test $= 1 \cdot 86$

Since, the calculated value of $|t| = 1 \cdot 30$ is less than the table value of t, we accept H_0 and conclude that medicine in general does not increase the blood pressure.

Aliter : The value S can also be calculated on :

$$\Sigma(d - \bar{d})^2 = \Sigma d^2 - (\Sigma d)^2 / n = 162 - (16)^2 / 9 = 133 \cdot 55$$

$$S = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n-1}} = \sqrt{\frac{133 \cdot 55}{9-1}} = 4 \cdot 085$$

Example 15. **10 prsons were appointed in a clerical position in an office. Their performance were noted giving a test and the mark recorded out of 50. They were given 6 months training and again they were given a test and marks were recorded out of 50.**

Employees :	A	B	C	D	E	F	G	H	I	J
Before taining :	25	20	35	15	42	28	26	44	35	48
After training :	26	20	34	13	43	40	29	41	36	46

By applying the t–test, can it be concluded that the employees have been benefitted by the training ? (Given for d.f. = 9, $t_{.05} = 2 \cdot 262$)

Solution. Let us take the hypothesis that the employees have not been benefitted by the training *i.e.,* $H_0 : \bar{d} = 0$ and $H_1 : \bar{d} \neq 0$. (⇒ two tailed test)

Before Training (I)	After Training (II)	d (II – I)	d^2
25	26	+ 1	1
20	20	0	0
35	34	− 1	1
15	13	− 2	4
42	43	+ 1	1
28	40	+ 12	144
26	29	+ 3	9
44	41	− 3	9
35	36	+ 1	1
48	46	− 2	4
$n = 10$		$\Sigma d = 10$	$\Sigma d^2 = 174$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{10}{10} = 1$$

$$S = \sqrt{\frac{\Sigma d^2 - (\bar{d})^2 \times n}{n-1}}$$

$$= \sqrt{\frac{174 - 10(1)^2}{10-1}} = 4 \cdot 269$$

Applying t–test :

$$t = \frac{\bar{d}}{S} \cdot \sqrt{n} = \frac{1}{4 \cdot 269} \times \sqrt{10} = \frac{3 \cdot 1692}{4 \cdot 269} = 0 \cdot 741$$

Degrees of freedom $= v = n - 1 = 10 - 1 = 9$

For $v=9$, $t_{.05}$ for two tailed test $=2 \cdot 262$

Since, the calculated value of t is less than the table value, we accept the null hypothesis and conclude that the employees have not been benefitted from the training.

Aliter : The value S can also be calculated on :

$$\Sigma(d-\bar{d})^2=\Sigma d^2-(\Sigma d)^2 / n=174-(10)^2 / 10$$

$$S=\sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}}=\sqrt{\frac{164}{10-1}}=4 \cdot 269$$

EXERCISE – 3

1. A certain stimulus administered to each of 12 patients resulted in the following increase in blood pressure : 5, 2, 8, –1, 3, 0, –2, 1, 5, 0, 4 and 6. Can it be concluded that the stimulus, will in general be accompanied by an increase in blood pressure ? (For $v=11$, $t_{0.01}=2 \cdot 21$)
   [Ans. $t=2 \cdot 90$, Reject H_0]
2. 12 Students were given intensive coaching and two tests were conducted, first test before coaching and second test after coaching. The scores of two tests are given below :

Student	1	2	3	4	5	6	7	8	9	10	11	12
Marks in 1st test	50	42	51	26	35	42	60	41	70	55	62	38
Marks in 2nd test	62	40	61	35	30	52	68	51	84	63	72	50

   Has the coaching helped in improving the scores ?
   [Ans. $t=4 \cdot 885$ Reject H_0, Coaching is carefull]
3. The sales data of an item in six shops before and after a special promotional compaign are as under :

Shops :	A	B	C	D	E	F
Before campaign :	53	28	31	48	50	42
After campaign :	58	29	30	55	56	45

   Can the campaign be judged to be a success ? Test at 5% level of significance.
   [Ans. $t=2 \cdot 78$, Reject H_0]
4. An I.Q. (Intelligence Quotient) test was conducted for 5 officers before and after a training. The reuslts are given below :

Officer :	I	II	III	IV	V
I.Q. before training :	110	120	123	132	125
I.Q. after training :	120	118	125	136	121

   Test whether there is any change in I.Q. after the training programme.
   (Give $t_{.01}(4)=4 \cdot 604$) **[Ans.** $t=\cdot 817$, Accept H_0]
5. Eight students were given Test in statistics, and after one month's coaching, they were given another test of the similar nature. The following table gives the difference in their marks in the second Test over the first :

Roll Number :	1	2	3	4	5	6	7	8
Difference in Marks :	4	– 2	6	– 8	12	5	–7	2

Is the difference in marks statistically significant ?
[Ans. $t=0 \cdot 623$, Accept H_0 and training is not effective *i.e.* difference in marks is statistically significant]

(4) Test of Hypothesis about Coefficient of Correlation : Suppose that a random sample (x_i, y_i) of size n has been drawn from a bivariate normal population and let r be the observed sample correlation coefficient. To test the hypothesis that the correlation coefficient in the population is zero *i.e.*, $\rho=0$ or the observed correlation is significant or not, we use t–test and the appropriate test statistic t to be used is

$$t=\frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}$$

Where, r = sample correlation coefficient, n = size of the sample.

Procedure : The following steps are taken while testing the hypothesis about coefficient of correlation :

(1) Set up the null hypothesis $H_0 : \rho=0$, *i.e.*, correlation coefficient in the population is zero or the observed correlation coefficient is not significant.
Alternative hypothesis $H_1 : \rho \neq 0$ ($\Rightarrow$ Two tailed test)

(2) Substituting the values of r and n in the above stated formula.

(3) Degrees of freedom are worked out by using the following formula :
Degrees of freedom $=v=n-2$

The other steps such as (i) level of significance, (ii) table value of t and (iii) decision making for testing the population correlation coefficient are the same as those given for the test of hypothesis.

Example 16. A correlation coefficient of $0 \cdot 6$ is discovered in a sample of 18 observations. Is it significant at 1% level ?

Solution. Given, $r=0 \cdot 6, n=18$

$H_0 : \rho=0$

and $H_1 : \rho \neq 0$ (Two tailed test)

Applying t–test :

$$t=\frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}=\frac{0 \cdot 6}{\sqrt{1-(\cdot 6)^2}} \times \sqrt{18-2}=\frac{\cdot 6}{\cdot 8} \times 4=3$$

Degrees of freedom $=v=n-2=18-2=16$

For $v=16$, $t_{.05}$ for two tailed test $=2 \cdot 92$

Since the calculated value of $|t|=3$ is greater than the table value of t, we reject H_0 and conclude that the observed value of the correlation coefficient is significant.

Example 17. A random sample of 27 observations from a normal populations gives a correlation coefficient of $-0 \cdot 4$. Is this significant of the existence of correlation in the population ?
(Given for $v=25$, $t_{.01}=2 \cdot 79$)

Solution. We are given : $n=27, r=-0 \cdot 4$

$H_0 : \rho=0$

And $H_1 : \rho \neq 0$ ($\Rightarrow$ Two tailed test)

Aplying t–test,

$$t=\frac{r}{\sqrt{1-r^2}}\cdot\sqrt{n-2}$$

$$=\frac{0\cdot4}{\sqrt{1-(-\cdot4)^2}}\sqrt{27-2}$$

$$=\frac{\cdot4}{\cdot917}\times5=2\cdot18$$

Degres of freedom $=v=n-1=27-2=25$

For $v=25, t_{.05}$ for two tailed test $=2\cdot79$

Since the calculated value of $|t|$ is less than the table value, we accept H_0 and conclude that the correlation of the sample is not significant to warrant the existence of such correlation in the population.

Example 18. **(a) How many pairs of observations must be included in a sample in order that an observed correlation coefficient of value 0·42 shall have a calculated value of t greater than 2·72?**

(b) Find the least value of r in a random sample of 27 pairs of values from a bivariate population which would be significant at 5% level.

Solution. (a) Given : $r=0\cdot42$, Critical value, 2·72

We have $$t=\frac{r}{\sqrt{1-r^2}}\times\sqrt{n-2}$$

In order that the calculated value of t may be greater than 2·72,

$$\frac{r}{\sqrt{1-r^2}}\times\sqrt{n-2}>2\cdot72 \quad\Rightarrow\quad \frac{0\cdot4}{\sqrt{1-(\cdot42)^2}}\times\sqrt{n-2}>2\cdot72$$

$$\Rightarrow\quad \frac{0\cdot42\sqrt{n-2}}{\sqrt{1-\cdot1764}}>2\cdot72 \quad\Rightarrow\quad \frac{0\cdot42\sqrt{n-2}}{\sqrt{\cdot8236}}>2\cdot72$$

$$\Rightarrow\quad \frac{0\cdot42\sqrt{n-2}}{0\cdot908}>2\cdot72 \quad\Rightarrow\quad 0\cdot4625\sqrt{n-2}>2\cdot72$$

$$\Rightarrow\quad \sqrt{n-2}>\frac{2\cdot72}{0\cdot4625} \quad\Rightarrow\quad \sqrt{n-2}>5\cdot88\Rightarrow n-2>34\cdot57$$

(Squaring both sides)

or $\quad n>34\cdot57+2 \quad$ or $\quad 36\cdot57$

Hence, the required value of n would be greater than 36·57 *i.e.,* 37 at least.

(b) In order that the 'r' may be significant, the calculated value of t must be more than the table value of r for 25 d.f. at 5% for two tailed test.

We have $$t=\frac{r}{\sqrt{1-r^2}}\times\sqrt{n-2} \quad\text{and}\quad t_{.05}(25)=2\cdot06$$

Thus, according to given condition, $t>2\cdot06$

$$\frac{5r}{\sqrt{1-r^2}}>2\cdot06 \quad\text{or}\quad 5r>2\cdot06\sqrt{1-r^2}$$

$$\text{or}\quad 25r^2>4\cdot2436(1-r^2) \quad\text{or}\quad 25r^2>4\cdot2436-4\cdot2436r^2$$

$$\Rightarrow\quad 29\cdot2436r^2>4\cdot2436 \quad\text{or}\quad r^2>\frac{4\cdot2436}{29\cdot2436}$$

$$\Rightarrow\quad r>0\cdot38$$

Hence, any value of r which is more than 0·38 would be significant for 25 d.f. at 5% level.

EXERCISE – 4

1. A random sample of 18 pairs from a normal population showed a correlation coefficient of 0·4. Is this value significant of correlation in the population ?
   [Ans. $t=1\cdot76$, Accept H_0]
2. Find the least value of r in a sample of 18 pairs a bivariate normal population significant at 5% level.
   [Ans. $|r|=\cdot468$]
3. How many pairs of observations must be included in a sample in order that an observed correlation coefficient of value 0·52 shall have a calculated value of t greater than 2·82 ?
   [Ans. $n=24$]
4. A random sample of 25 from a normal universe gives correlation coefficient of $-0\cdot48$. Is this significant of the existance of correlation in the population ?
   [Ans. $|t|=2\cdot624$, Reject H_0]
5. A random sample of 27 pairs of observations from a normal population gave a correlation coefficient of 0·58. Is this significant of correlation in the population ?
   [Ans. $t=3\cdot56$, Reject H_0]
6. A study of the heights of 25 pairs of husbands and their wives in a factory shows that the coefficient of correlation is 0·37. Test whether correlation is significant or not (The value of t at 5% for 23 degrees of freedom is 2·069)
   [Ans. $t=1\cdot903$, Accept H_0]
7. Is a value of $r=-0\cdot48$ significant if obtained from a sample of 25 pairs of values from a normal population ?
   [Ans. $t=2\cdot624$, Reject H_0]

(2) Fisher's Z—Transformation

t-test is used to test the significance of the correlation coefficient if the value of the population correlation coefficient is zero. If the population correlation coefficient is other than zero or difference between two sample correlation coeffiicents are to be tested, then the t-rest can not used and in that case Fisher Z-test is used. Fisher's Z-test has two applications :

(1) To test whether an observed value of r differs significantly from some hyptothetical value of population correlation coefficient, other than zero, the appropriate test statistic to used is :

$$|Z|=\frac{Z_r-Z_\rho}{SE_z}$$

Procedure

Its testing procedure is as follows ;

(i) Set up the null hypothesis $H_0:\rho=\rho_0$ *i.e.,* population correlation coefficient is equal to a specified value or there is no difference between r and ρ.

(ii) Thereafter, the value of r (sample correlation coefficient) and ρ (population correlation coefficient) are changed into Z–transformation by using the following formula :

$$Z = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) = 1{\cdot}1513\log_{10}\left(\frac{1+r}{1-r}\right)$$

(iii) Z-transformation of r

$$Z_s = 1{\cdot}1513 \,.\, \log_{10}\left(\frac{1+r}{1-r}\right)$$

Z-transformation of ρ

$$Z_p = 1{\cdot}1513 \,.\, \log_{10}\left(\frac{1+\rho}{1-\rho}\right)$$

Note : If ρ is not known, then it is taken as zero in which case $\rho = 0$.

(iv) The standard error of Z is worked out as under :

$$S.E_Z = \frac{1}{\sqrt{n-3}}$$

(v) Finally, we compute the value of Z as follows :

$$Z = \frac{Z_s - Z_p}{\frac{1}{\sqrt{n-3}}} = (Z_s - Z_p) \times \sqrt{n-3}$$

(vi) The calculated value of Z is compared with 1·96 at 5% level of significance and 2·58 at 1% level of significance.

If $|Z| > 1{\cdot}96$, the difference is considered significant at 5% level of significance otherwise insignificant.

If $|Z| > 2{\cdot}58$, the difference is considered significant at 1% level of significance otherwise insignificant.

The following examples illustrate the application of Z-test.

Example 19. **Test the significance of the coefficient of correlation $r = 0{\cdot}5$ discovered in a sample of 19 paired of observation against hypothetical correlation $\rho = 0{\cdot}7$. Apply Fisher's Z—test.**

Solution. Let us take the hypothesis that correlation coefficient in the population is 0·7 *i.e.,* $H_0 : \rho = 0{\cdot}7$ and $H_1 : \rho \neq 0{\cdot}7$

Given : $r = 0{\cdot}5$, $\rho = 0{\cdot}7$

Applying Z-transformation, we obtain :

Z-transformation of r

$$Z_s = 1{\cdot}1513\log_{10}\left(\frac{1+r}{1-r}\right)$$
$$= 1{\cdot}1513 \,.\, \log_{10}\left(\frac{1+0{\cdot}5}{1-0{\cdot}5}\right)$$
$$= 1{\cdot}1513\log_{10}\left(\frac{1{\cdot}5}{0{\cdot}5}\right)$$
$$= 1{\cdot}1513 \times \log 3$$
$$= 1{\cdot}1513 \times {\cdot}4771 = 0{\cdot}549$$

Z-transformation of ρ

$$Z_\rho = 1{\cdot}1513\log_{10}\left(\frac{1+\rho}{1-\rho}\right)$$
$$= 1{\cdot}1513 \,.\, \log_{10}\left(\frac{1+0{\cdot}7}{1-0{\cdot}7}\right)$$
$$= 1{\cdot}1513\log_{10}\left(\frac{1{\cdot}7}{0{\cdot}3}\right)$$
$$= 1{\cdot}1513 \times \log(5{\cdot}67)$$
$$= 1{\cdot}1513 \times 0{\cdot}7536 = 0{\cdot}868$$

Applying Fisher's Z-test

$$\therefore \quad |Z| = \frac{Z_s - Z_p}{\frac{1}{\sqrt{n-3}}} = \frac{|0{\cdot}549 - 0{\cdot}868|}{\frac{1}{\sqrt{19-3}}}$$
$$= \frac{0{\cdot}319}{1} \times \sqrt{16} = 0{\cdot}319 \times 4 = 1{\cdot}276$$

Since, the calculated value of $|Z|$ is less than 1·96, we accept the null hypothesis at 5% level and conclude that the population correlation coefficient is 0·7.

(2) Testing the significance of the difference between two independent sample correlation coefficients.

Z-test can also be used to test the significance of the difference of the two sample correlation coefficients.

Procedure

Its testing procedure is as follows :

(i) First of all, set up the null hypothesis that there is no difference between two correlation coefficients *i.e.,* $H_0 : \rho_1 = \rho_2$

(iii) Thereafter, the two values r_1 and r_2 are changed into Z-transformation by using the following formula :

$$Z = 1{\cdot}1513\log_{10}\left(\frac{1+r}{1-r}\right)$$

(iv) The standard error of the difference between Z_1 and Z_2 is worked out as under :

$$S.E_{Z_1 - Z_2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

(v) Finally, we compute the value of $|Z|$ as follows :

$$|Z| = \frac{Z_1 - Z_2}{\frac{1}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}}$$

Where $Z_1 = 1{\cdot}1513\log_{10}\left(\frac{1+r_1}{1-r_1}\right)$

$$Z_2 = 1{\cdot}1513\log_{10}\left(\frac{1+r_2}{1-r_2}\right)$$

(v) If the calculated value of $|Z|$ is greater than 1·96 at 5% level of significane, the difference between two r's is significant.

Example 20. **The following data give sample sizes and correlation coefficient. Test the significance of the difference between the values using Fisher's Z-test.**

Sample size	Value of r
5	0·87
12	0·56

Solution : Let us take the hypothesis that two correlation coefficients do not differ significantly *i.e.*, $H_0 : \rho_1 = \rho_2$ and $H_1 : \rho_1 \neq \rho_2$ ($\Rightarrow$ Two tailed test)

Z–transformation of r_1

$$Z_1 = 1\cdot1513 \log_{10}\left(\frac{1+r_1}{1-r_1}\right) = 1\cdot1513 \log_{10}\left(\frac{1+0\cdot87}{1-0\cdot87}\right) = 1\cdot1513 \log_{10}\left(\frac{1\cdot87}{0\cdot13}\right)$$

$$= 1\cdot1513 \log_{10}(14\cdot384) = 1\cdot1513 \times 1\cdot1577 = 1\cdot33$$

Z–transformation of r_2

$$Z_2 = 1\cdot1513 \log_{10}\left(\frac{1+r_2}{1-r_2}\right) = 1\cdot1513 \log_{10}\left(\frac{1+0\cdot56}{1-0\cdot56}\right) = 1\cdot1513 \log_{10}\left(\frac{1\cdot56}{0\cdot44}\right)$$

$$= 1\cdot1513 \times \log_{10} 3\cdot545 = 1\cdot1513 \times 0\cdot5495 = 0\cdot63$$

Applying Fisher's Z–test,

$$|Z| = \frac{Z_1 - Z_2}{\dfrac{1}{\sqrt{\dfrac{1}{n_1-3}+\dfrac{1}{n_2-3}}}} = \frac{1\cdot33-0\cdot63}{\dfrac{1}{\sqrt{\dfrac{1}{5-3}+\dfrac{1}{12-3}}}} = \frac{0\cdot70}{\sqrt{0\cdot61}} = \frac{0\cdot70}{0\cdot78} = 0\cdot9$$

As the value of Z is less than $1\cdot96$ at 5% level of significance, we accept the null hypothesis and conclude that the difference is not significant.

EXERCISE – 5

1. Test the significance of the correlation $r = +0\cdot75$ from a sample of size 30 against hypothetical correlation $\rho = 0\cdot55$. **[Ans.** $|Z| = 1\cdot8$, Accept H_0]
2. From a sample of 10 pair of observations the correlation is $0\cdot5$ and the corresponding population value is $0\cdot3$. Is this difference significant at 5% level of significance ? Apply Z–test. **[Ans.** $|Z| = 0\cdot96$, Accept H_0]
3. The following data give sample sizes and correlation coefficients Test the significance of the difference between two values using Fisher's Z–test.

Sample size	Value of *r*
23	0·87
28	0·56

[Ans. $|Z| = 1.$, Accept H_0]

4. A correlation coeffcient of $0\cdot6$ is obtained from a sample of 19 paired observations. Is it significantly different from $0\cdot4$? **[Ans.** $|Z| = 1\cdot076$, Accept H_0]

(3) F–Test (Variance Ratio Test)

F–test is named after the greater statistician **R.A. Fisher.** F–test is used to test whether the two independent estimates of population variance differ significantly or whether the two samples may be regarded as drawn from the normal population having the same variance. For carrying out the test, we calculate F–statistic. F–statistic is defined as :

$$F = \frac{\text{Larger estimate of population variance}}{\text{Smaller estimate of population variance}} = \frac{S_1^2}{S_2^2} \quad \text{where, } S_1^2 > S_2^2$$

Procedure

Its testing procedure is as follows :

(i) Set up null hypothesis that the two population variances are equal *i.e.*, $H_0 : \sigma_1^2 = \sigma_2^2$

(ii) The variances of the random samples are calculated by using formula :

$$S_1^2 = \frac{\Sigma(X_1-\overline{X}_1)^2}{n_1-1} \quad \text{or} \quad \frac{n_1}{n_1-1}s_1^2 \quad \text{or} \quad \frac{1}{n_1-1}\left[\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n_1}\right]$$

$$S_2^2 = \frac{\Sigma(X_2-\overline{X}_2)^2}{n_1-1} \quad \text{or} \quad \frac{n_2}{n_2-1}s_2^2 \quad \text{or} \quad \frac{1}{n_2-1}\left[\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n_2}\right]$$

(iii) The variance ratio F is computed as :

$$F = \frac{S_1^2}{S_2^2} \quad \text{where, } S_1^2 > S_2^2$$

(iv) The degrees of freedom are computed. The degrees of freedom of the larger estimate of the population variance is denoted by v_1 and the smaller estimate by v_2. That is,

v_1 = degrees of freedom for sample having larger variance $= n_1 - 1$

v_2 = degrees of freedom for sample having smaller variance $= n_2 - 1$

(v) Then from the F–table given at the end of the book, the value of F is found for v_1 and v_2 with 5% level of significance.

(vi) Then we compare the calculated value of F with the table value of F_{05} for v_1 and v_2 degrees of freedom. If the calculated value of F exceeds the table value of F, we reject the null hypothesis and conclude that the differnce between the two variances is significant. On the other hand, if the calculated value of F is less than the table value, the null hypothesis is accepted and conclude that both the samples have come from the population having same varaince.

The following examples illustrate the applications of F–test :

Example 21. **In a sample of 8 observations, the sum of squared deviations of items from the mean was $94\cdot5$. In another sample of 10 observations, the value was found to be $101\cdot7$ Test whether the difference is significant at 5% level. (You are given that at 5% level of significance, the critical value of F for $v_1 = 7$ and $v_2 = 9$ df is $3\cdot29$).**

Solution. Let us take the hypothesis that the difference in the variances of the two samples is not significant *i.e.*, $H_0 : \sigma_1^2 = \sigma_2^2$

We are given : $n_1 = 8$, $\Sigma(X_1-\overline{X}_1)^2 = 94\cdot5$

$n_2 = 10$, $\Sigma(X_2-\overline{X}_2)^2 = 101\cdot7$

$$S_1^2 = \frac{\Sigma(X_1-\overline{X}_1)^2}{n_1-1} = \frac{94\cdot5}{8-1} = \frac{94\cdot5}{7} = 13\cdot5$$

$$S_2^2 = \frac{\Sigma(X_2-\overline{X}_1)^2}{n_1-1} = \frac{101\cdot7}{10-1} = \frac{101\cdot7}{9} = 11\cdot3$$

Applying F–test,