CHAPTER 1

Probability

Chapter Outline

1.1 INTRODUCTION

The concept of probability originated from the analysis of the games of chance. Even today, a large number of problems exist which are based on the games of chance, such as tossing of a coin, throwing of dice, and playing of cards. The utility of probability in business and economics is most emphatically revealed in the field of predictions for the future. Probability is a concept which measures the degree of uncertainty and that of certainty as a corollary.

The word *probability* or 'chance' is used commonly in day-to-day life. Daily, we come across the sentences like, 'it may rain today', 'India may win the forthcoming cricket match against Sri Lanka', 'the chances of making profits by investing in shares of Company A are very bright, etc. Each of the above sentences involves an element of uncertainty. A numerical measure of uncertainty is provided by a very important branch of mathematics called *theory of probability*. Before we study the probability theory in detail, it is appropriate to explain certain terms which are essential for the study of the theory of probability.

1.2 SOME IMPORTANT TERMS AND CONCEPTS

1. Random Experiment If an experiment is conducted, any number of times, under identical conditions, there is a set of all possible outcomes associated with it.

If the outcome is not unique but may be any one of the possible outcomes, the experiment is called a random experiment, e.g., tossing a coin, throwing a die.

2. Outcome The result of a random experiment is called an outcome. For example, consider the following:

(a) Suppose a random experiment is 'a coin is tossed'. This experiment gives two possible outcomes—head or tail.
(b) Suppose a random experiment is 'a die is thrown'. This experiment gives six possible outcomes—1, 2, 3, 4, 5 or 6—on the uppermost face of a die.

3. Trial and Event Any particular performance of a random experiment is called a trial and outcome. A combination of outcomes is called an event. For example, consider the following:

(a) Tossing of a coin is a trial, and getting a head or tail is an event.
(b) Throwing of a die is a trial and getting 1 or 2 or 3 or 4 or 5 or 6 is an event.

4. Exhaustive Event The total number of possible outcomes of a random experiment is called an exhaustive event. For example, consider the following:

(a) In tossing of a coin, there are two exhaustive events, viz., head and tail.
(b) In throwing of a die, there are six exhaustive events, getting 1 or 2 or 3 or 4 or 5 or 6.

5. Mutually Exclusive Events Events are said to be mutually exclusive if the occurrence of one of them precludes the occurrence of all others in the same trial, i.e., they cannot occur simultaneously. For example, consider the following:

(a) In tossing a coin, the events head or tail are mutually exclusive since both head and tail cannot occur at the same time.
(b) In throwing a die, all the six events, i.e., getting 1 or 2 or 3 or 4 or 5 or 6 are mutually exclusive events.

6. Equally Likely Events The outcomes of a random experiment are said to be equally likely if the occurrence of none of them is expected in preference to others. For example, consider the following:

(a) In tossing a coin, head or tail are equally likely events.
(b) In throwing a die, all the six faces are equally likely events.

7. Independent Events Events are said to be independent if the occurrence of an event does not have any effect on the occurrence of other events. For example, consider the following:

(a) In tossing a coin, the event of getting a head in the first toss is independent of getting a head in the second, third, and subsequent tosses.
(b) In throwing a die, the result of the first throw does not affect the result of the second throw.

8. Favourable Events The favourable events in a random experiment are the number of outcomes which entail the occurrence of the event. For example, consider the following:

In throwing of two dice, the favourable events of getting the sum 5 is (1, 4), (4, 1), (2, 3), (3, 2), i.e., 4.

1.3 DEFINITIONS OF PROBABILITY

1.3.1 Classical Definition of Probability

Let n be the number of equally likely, mutually exclusive, and exhaustive outcomes of a random experiment. Let m be number of the outcomes which are favourable to the occurrence of an event A. The probability of event A occurring, denoted by $P(A)$, is given by

$$P(A) = \frac{\text{Number of outcomes favourable to } A}{\text{Number of exhaustive outcomes}} = \frac{m}{n}$$

1.3.2 Empirical or Statistical Definition of Probability

If an experiment is repeated a large number of times under identical conditions, the limiting value of the ratio of the number of times the event A occurs to the total number of trials of the experiment as the number of trials increase indefinitely is called the probability of occurrence of the event A.

Let $P(A)$ be the probability of occurrence of the event A. Let m be the number of times in which an event A occurs in a series of n trials.

$$P(A) = \lim_{n\to\infty} \frac{m}{n}, \text{ provided the limit is finite and unique.}$$

1.3.3 Axiomatic Definition of Probability

Before discussing the axiomatic definition of probability, it is necessary to explain certain concepts that are necessary to its understanding.

1. Sample Space A set of all possible outcomes of a random experiment is called a sample space. Each element of the set is called a *sample point* or a *simple event* or an *elementary event*.

The sample space of a random experiment is denoted by S. For example, consider the following:

(a) In a random experiment of tossing of a coin, the sample space consists of two elementary events.

$S = \{H, T\}$

(b) In a random experiment of throwing of a die, the sample space consists of six elementary events.

$S = \{1, 2, 3, 4, 5, 6\}$

The elements of S can either be single elements or ordered pairs. If two coins are tossed, each element of the sample space consists of the following ordered pairs:

$S = \{(\text{H, H}), (\text{H, T}), (\text{T, H}), (\text{T, T})\}$

2. Event Any subset of a sample space is called an event. In the experiment of throwing of a die, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Let A be the event that an odd number appears on the die. Then $A = \{1, 3, 5\}$ is a subset of S. Similarly, let B be the event of getting a number greater than 3. Then $B = \{4, 5, 6\}$ is another subset of S.

Definition of Probability Let S be a sample space of an experiment and A be any event of this sample space. The probability $P(A)$ of the event A is defined as the real-value set function which associates a real value corresponding to a subset A of the sample space S. The probability $P(A)$ satisfies the following three axioms.

Axiom I: $P(A) \geq 0$, i.e., the probability of an event is a nonnegative number.

Axiom II: $P(S) = 1$, i.e., the probability of an event that is certain to occur must be equal to unity.

Axiom III: If $A_1, A_2, ..., A_n$ are finite mutually exclusive events then

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$
$$= \sum_{i=1}^{n} P(A_i)$$

i.e., the probability of a union of mutually exclusive events is the sum of probabilities of the events themselves.

Example 1

What is the probability that a leap year selected at random will have 53 Sundays?

Solution

A leap year has 366 days, i.e., 52 weeks and 2 days. These 2 days can occur in the following possible ways:

(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

Number of exhaustive cases $n = 7$

Number of favourable cases $m = 2$

Let A be the event of getting 53 Sundays in a leap year.

$$P(A) = \frac{m}{n} = \frac{2}{7}$$

Example 2

Three unbiased coins are tossed. Find the probability of getting (i) exactly two heads, (ii) at least one tail, (iii) at most two heads, (iv) a head on the second coin, and (v) exactly two heads in succession.

Solution

When three coins are tossed, the sample space S is given by

$$S = \{\text{HHH, HTH, THH, HHT, TTT, THT, TTH, HTT}\}$$

$$n(s) = 8$$

(i) Let A be the event of getting exactly two heads.

$$A = \{\text{HTH, THH, HHT}\}$$

$$n(A) = 3$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{8}$$

(ii) Let B be the event of getting at least one tail.

$$B = \{\text{HTH, THH, HHT, TTT, THT, TTH, HTT}\}$$

$$n(B) = 7$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{7}{8}$$

(iii) Let C be the event of getting at most two heads.

$$C = \{\text{HTH, THH, HHT, TTT, THT, TTH, HTT}\}$$

$$n(C) = 7$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{7}{8}$$

(iv) Let D be the event of getting a head on the second coin.

$$D = \{\text{HHH, THH, HHT, THT}\}$$

$$n(D) = 4$$

$$P(D) = \frac{n(D)}{n(S)} = \frac{4}{8} = \frac{1}{2}$$

(v) Let E be the event of getting two heads in succession.

$$E = \{\text{HHH, THH, HHT}\}$$

$$n(E) = 3$$

$$P(E) = \frac{n(E)}{n(S)} = \frac{3}{8}$$

Example 3

A fair dice is thrown. Find the probability of getting (i) an even number, (ii) a perfect square, and (iii) an integer greater than or equal to 3.

Solution

When a dice is thrown, the sample space S is given by

$$S = \{1, 2, 3, 4, 5, 6\}$$

$$n(S) = 6$$

(i) Let A be the event of getting an even number.

$$A = \{2, 4, 6\}$$

$$n(A) = 3$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

(ii) Let B be the event of getting a perfect square.

$$B = \{1, 4\}$$

$$n(B) = 2$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{2}{6} = \frac{1}{3}$$

(iii) Let C be the event of getting an integer greater than or equal to 3.

$$C = \{3, 4, 5, 6\}$$

$$n(C) = 4$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{4}{6} = \frac{2}{3}$$

Example 4

A card is drawn from a well-shuffled pack of 52 cards. Find the probability of (i) getting a king card, (ii) getting a face card, (iii) getting a red card, (iv) getting a card between 2 and 7, both inclusive, and (v) getting a card between 2 and 8, both exclusive.

Solution

Total number of cards = 52

One card out of 52 cards can be drawn in ways.

$$n(S) = {}^{52}C_1 = 52$$

(i) Let A be the event of getting a king card. There are 4 king cards and one of them can be drawn in ${}^{4}C_1$ ways.

$$n(A) = {}^{4}C_1 = 4$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{4}{52} = \frac{1}{13}$$

(ii) Let B be the event of getting a face card. There are 12 face cards and one of them can be drawn in ${}^{12}C_1$ ways.

$$n(B) = {}^{12}C_1 = 12$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{12}{52} = \frac{3}{13}$$

(iii) Let C be the event of getting a red card. There are 26 red cards and one of them can be drawn in ${}^{26}C_1$ ways.

$$n(C) = {}^{26}C_1 = 26$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{26}{52} = \frac{1}{2}$$

(iv) Let D be the event of getting a card between 2 and 7, both inclusive. There are 6 such cards in each suit giving a total of $6 \times 4 = 24$ cards. One of them can be drawn in ${}^{24}C_1$ ways.

$$n(D) = {}^{24}C_1 = 24$$

$$P(D) = \frac{n(D)}{n(S)} = \frac{24}{52} = \frac{6}{13}$$

(v) Let E be the event of getting a card between 2 and 8, both exclusive. There are 5 such cards in each suit giving a total of $5 \times 4 = 20$ cards. One of them can be drawn in ${}^{20}C_1$ ways.

$$n(E) = {}^{20}C_1 = 20$$

$$= \frac{n(E)}{n(S)} = \frac{20}{52} = \frac{5}{13}$$

Example 5

A bag contains 2 black, 3 red, and 5 blue balls. Three balls are drawn at random. Find the probability that the three balls drawn (i) are blue (ii) consist of 2 blue and 1 red ball, and (iii) consist of exactly one black ball.

Solution

Total number of balls = 10

3 balls out of 10 balls can be drawn in ${}^{10}C_3$ ways.

$$n(S) = {}^{10}C_3 = 120$$

(i) Let A be the event that the three balls drawn are blue. 3 blue balls out of 5 blue balls can be drawn in ${}^5C_3$ ways.

$$n(A) = {}^5C_3 = 10$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{10}{120} = \frac{1}{12}$$

(ii) Let B be the event that the three balls drawn consist of 2 blue and 1 red ball.
2 blue balls out of 5 blue balls can be drawn in ${}^5C_2$ ways. 1 red ball out of 3 red balls can be drawn in ${}^3C_1$ ways.

$$n(B) = {}^5C_2 \times {}^3C_1 = 30$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{30}{120} = \frac{1}{4}$$

(iii) Let C be the event that three balls drawn consist of exactly one black ball, i.e., remaining two balls can be drawn from 3 red and 5 blue balls. One black ball can be drawn from 2 black balls in ${}^2C_1$ ways and the remaining 2 balls can be drawn from 8 balls in ${}^8C_2$ ways.

$$n(C) = {}^2C_1 \times {}^8C_2 = 56$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{56}{120} = \frac{7}{15}$$

Example 6

A class consists of 6 *girls and* 10 *boys. If a committee of three is chosen at random from the class, find the probability that (i) three boys are selected, and (ii) exactly two girls are selected.*

Solution

Total number of students = 16

A committee of 3 students from 16 students can be selected in ${}^{16}C_3$ ways.

$$n(S) = {}^{16}C_3 = 560$$

(i) Let A be the event that 3 boys are selected.

$$n(A) = {}^{10}C_3 = 120$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{120}{560} = \frac{3}{14}$$

(ii) Let B be the event that exactly 2 girls are selected. 2 girls from 6 girls can be selected in ${}^6C_2$ ways and one boy from 10 boys can be selected in ${}^{10}C_1$ ways.

$$n(B) = {}^6C_2 \times {}^{10}C_1 = 150$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{150}{560} = \frac{15}{16}$$

Example 7

From a collection of 10 *bulbs, of which* 4 *are defective,* 3 *bulbs are selected at random and fitted into lamps. Find the probability that (i) all three bulbs glow, and (ii) the room is lit.*

Solution

Total number of bulbs = 10

3 bulbs can be selected from 10 bulbs in ${}^{10}C_3$ ways.

$$n(S) = {}^{10}C_3 = 120$$

(i) Let A be event that all three bulbs glow. This event will occur when 3 bulbs are selected from 6 nondefective bulbs in ${}^6C_3$ ways.

$$n(A) = {}^6C_3 = 20$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{20}{120} = \frac{1}{6}$$

(ii) Let B be the event that the room is lit. Let $\overline{B}$ be the event that the room is dark. The event $\overline{B}$ will occur when 3 bulbs are selected from 4 defective bulbs in ${}^4C_3$ ways.

$$n(\overline{B}) = {}^4C_3 = 4$$

$$P(\overline{B}) = \frac{n(\overline{B})}{n(S)} = \frac{4}{120} = \frac{1}{30}$$

$$\therefore \quad P(B) = 1 - P(\overline{B}) = 1 - \frac{1}{30} = \frac{29}{30}$$

Example 8

There are 20 *tickets numbered* 1, 2, ..., 20. *One ticket is drawn at random. Find the probability that the ticket bears a number which is (i) even, (ii) a perfect square, and (iii) multiple of* 3.

Solution

There are 20 tickets numbered from 1 to 20.

$n(S) = 20$

(i) Let A be the event that a ticket bears a number which is even.

$$A = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$$

$$n(A) = 10$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{10}{20} = \frac{1}{2}$$

(ii) Let B the event that a ticket bears a number which is a perfect square.

$$B = \{1, 4, 9, 16\}$$

$$n(B) = 4$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{4}{20} = \frac{1}{5}$$

(iii) Let C be the event that a ticket bears a number which is a multiple of 3.

$$C = \{3, 6, 9, 12, 15, 18\}$$

$$n(C) = 6$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{6}{20} = \frac{3}{10}$$

Example 9

Four letters of the word 'THURSDAY' are arranged in all possible ways. Find the probability that the word formed is 'HURT'.

Solution

Total number of letters in the word 'THURSDAY' = 8

Four letters from 8 letters can be arranged in ${}^8P_4$ ways.

$$n(S) = {}^8P_4 = 1680$$

Let A be the event that the word formed is 'HURT'. The word 'HURT' can be formed in one way only.

$$n(A) = 1$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{1680}$$

Example 10

A bag contains 5 red, 4 blue, and m green balls. If the probability of getting two green balls when two balls are selected at random is $\frac{1}{7}$*, find m.*

Solution

Total number of balls = 5 + 4 + m = 9 + m

2 balls out of 9 + m balls can be drawn in ${}^{9+m}C_2$ ways.

$$n(S) = {}^{9+m}C_2$$

Let A be the event that both the balls drawn are green.

2 green balls out of m green balls can be drawn in ${}^{m}C_2$ ways.

$$n(A) = {}^{m}C_2$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{{}^{m}C_2}{{}^{9+m}C_2}$$

But $\quad P(A) = \dfrac{1}{7}$

$$\frac{{}^{m}C_2}{{}^{9+m}C_2} = \frac{1}{7}$$

$$\frac{m(m-1)}{(m+9)(m+8)} = \frac{1}{7}$$

$$(m+9)\,(m+8) = 7\,m\,(m-1)$$

$$m^2 + 17m + 72 = 7m^2 - 7m$$

$$6m^2 - 24m - 72 = 0$$

$$3m^2 - 12m - 36 = 0$$

$$3m^2 - 18m + 6m - 36 = 0$$

$$3m(m-6) + 6(m-6) = 0$$

$$(3m+6)(m-6) = 0$$

$$3m + 6 = 0 \quad \text{or} \quad m - 6 = 0$$

$$m = -2 \quad \text{or} \quad m = 6$$

But $\quad m \neq -2$

$\therefore \quad m = 6$

EXERCISE 1.1

1. A card is drawn at random from a pack of 52 cards. Find the probability that the card drawn is (i) an ace card, and (ii) a club card.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{13} \quad \text{(ii) } \frac{1}{4}\right]$$

2. An unbiased coin is tossed twice. Find the probability of (i) exactly one head, (ii) at most one head, (iii) at least one head, and (iv) same face on both the coins.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{2} \text{ (ii) } \frac{3}{4} \text{ (iii) } \frac{3}{4} \text{ (iv) } \frac{1}{2}\right]$$

3. A fair dice is thrown thrice. Find the probability that the sum of the numbers obtained is 10.

$$\left[\textbf{Ans.:} \frac{1}{8}\right]$$

4. A ball is drawn at random from a box containing 12 red, 18 white, 19 blue, and 15 orange balls. Find the probability that (i) it is red or blue, and (ii) it is white, blue, or orange.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{2}{5} \text{ (ii) } \frac{43}{55}\right]$$

5. Eight boys and three girls are to sit in a row for a photograph. Find the probability that no two girls are together.

$$\left[\textbf{Ans.:} \frac{28}{55}\right]$$

6. If four persons are chosen from a group of 3 men, 2 women, and 4 children, find the probability that exactly two of them will be children.

$$\left[\textbf{Ans.:} \frac{10}{21}\right]$$

7. A box contains 2 white, 3 red, and 5 black balls. Three balls are drawn at random. What is the probability that they will be of different colours?

$$\left[\textbf{Ans.:} \frac{1}{4}\right]$$

8. Two cards are drawn from a well-shuffled pack of 52 cards. Find the probability of getting (i) 2 king cards, (ii) 1 king card and 1 queen card, and (iii) 1 king card and 1 spade card.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{221} \text{ (ii) } \frac{8}{663} \text{ (iii) } \frac{1}{26}\right]$$

9. A four-digit number is to be formed using the digits 0, 1, 2, 3, 4, 5. All the digits are to be different. Find the probability that the digit formed is (i) odd, (ii) greater than 4000, (iii) greater than 3400, and (iv) a multiple of 5.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{12}{25} \text{ (ii) } \frac{2}{5} \text{ (iii) } \frac{12}{25} \text{ (iv) } \frac{9}{25}\right]$$

10. 3 books of physics, 4 books of chemistry, and 5 books of mathematics are arranged in a shelf. Find the probability that (i) no physics books are together, (ii) chemistry books are always together, and (iii) books of the same subjects are together.

$$\left[\textbf{Ans.:}\ \text{(i)}\ \frac{6}{11}\quad \text{(ii)}\ \frac{1}{55}\quad \text{(iii)}\ \frac{1}{4620}\right]$$

11. 8 boys and 2 girls are to be seated at random in a row for a photograph. Find the probability that (i) the girls sit together, and (ii) the girls occupy 3^{rd} and 7^{th} seats.

$$\left[\textbf{Ans.:}\ \text{(i)}\ \frac{1}{5}\quad \text{(ii)}\ \frac{1}{45}\right]$$

12. A committee of 4 is to be formed from 15 boys and 3 girls. Find the probability that the committee contains (i) 2 boys and 2 girls, (ii) exactly one girl, (iii) one particular girl, and (iv) two particular girls.

$$\left[\textbf{Ans.:}\ \text{(i)}\ \frac{7}{68}\quad \text{(ii)}\ \frac{91}{204}\quad \text{(iii)}\ \frac{2}{9}\quad \text{(iv)}\ \frac{2}{51}\right]$$

13. If the letters of the word REGULATIONS are arranged at random, what is the probability that there will be exactly four letters between *R* and *E*?

$$\left[\textbf{Ans.:}\ \frac{6}{55}\right]$$

14. Find the probability that there will be 5 Sundays in the month of October.

$$\left[\textbf{Ans.:}\ \frac{3}{7}\right]$$

1.4 THEOREMS ON PROBABILITY

Theorem 1 The probability of an impossible event is zero, i.e., $P(\phi) = 0$, where ϕ is a null set.

Proof An event which has no sample points is called an impossible event and is denoted by ϕ.

For a sample space S of an experiment,

$$S \cup \phi = S$$

Taking probability of both the sides,

$$P(S \cup \phi) = P(S)$$

Since S and ϕ are mutually exclusive events,

$$P(S) + P(\phi) = P(S) \qquad \text{[Using Axiom III]}$$

$$\therefore \quad P(\phi) = 0$$

Theorem 2 The probability of the complementary event $\bar{A}$ of A is

$$P(\bar{A}) = 1 - P(A)$$

Proof Let A be an event in the sample space S.

$$A \cup \bar{A} = S$$

$$P(A \cup \bar{A}) = P(S)$$

Since A and $\bar{A}$ are mutually exclusive events,

$$P(A) + P(\bar{A}) = P(S)$$

$$P(A) + P(\bar{A}) = 1 \qquad [\because \quad P(S) = 1]$$

$$\therefore \quad P(\bar{A}) = 1 - P(A)$$

Note Since A and $\bar{A}$ are mutually exclusive events,

$A \cup \bar{A} = S$ and $A \cap \bar{A} = \phi$

Corollary Probability of an event is always less than or equal to one, i.e., $P(A) \leq 1$

Proof $P(A) = 1 - P(\bar{A})$

$$P(A) \leq 1 \qquad [\because \quad P(\bar{A}) \geq 0 \text{ by Axiom I}]$$

De Morgan's Laws Since an event is a subset of a sample space, De Morgan's laws are applicable to events.

$$P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$$

$$P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$$

Theorem 3 For any two events A and B in a sample space S,

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

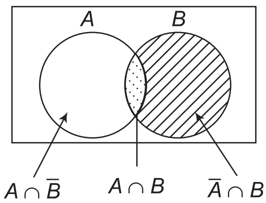


Fig. 1.1

Proof From the Venn diagram (Fig. 1.1),

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

$$P(B) = P\left[(A \cap B) \cup (\bar{A} \cap B)\right]$$

Since $(A \cap B)$ and $(\bar{A} \cap B)$ are mutually exclusive events,

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Similarly, it can be shown that

$$P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

Theorem 4 Additive Law of Probability (Addition Theorem)

The probability that at least one of the events A and B will occur is given by

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

Proof From the Venn diagram (Fig. 1.1),

$$A \cup B = A \cup (\bar{A} \cap B)$$

$$P(A \cup B) = P\left[A \cup (\bar{A} \cap B\right]$$

Since A and $(\bar{A} \cap B)$ are mutually exclusive events,

$$P(A \cup B) = P(A) + P(\bar{A} \cap B) \qquad \text{[Using Axiom III]}$$

$$= P(A) + P(B) - P(A \cap B) \qquad \text{[Using Theorem 3]}$$

Remarks

1. If A and B are mutually exclusive events, i.e., $A \cap B = \phi$ then $P(A \cap B) = 0$ according to Theorem 1.

   Hence, $P(A \cup B) = P(A) + P(B)$

2. The event $A \cup B$ (i.e., A or B) denotes the occurrence of either A or B or both. Alternately, it implies the occurrence of at least one of the two events.

   $A \cup B = A + B$

3. The event $A \cap B$ (i.e., A and B) is a compound or joint event that denotes the simultaneous occurrence of the two events.

   $A \cap B = AB$

Corollary 1 From the Venn diagram (Fig. 1.1),

$$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$$

where $P(\bar{A} \cap \bar{B})$ is the probability that none of the events A and B occur simultaneously.

Corollary 2 P(Exactly one of A and B occurs) $= P\left[(A \cap \bar{B}) \cup (\bar{A} \cap B)\right]$

$$= P(A \cap \bar{B}) + P(\bar{A} \cap B) \qquad \left[\because \ (A \cap \bar{B}) \cap (\bar{A} \cap B) = \phi\right]$$

$$= P(A) - P(A \cap B) + P(B) - P(A \cap B) \qquad \text{[Using Theorem 3]}$$

$$= P(A) + P(B) - 2P(A \cap B)$$

$$= P(A \cup B) - P(A \cap B) \qquad \text{[Using Theorem 4]}$$

$= P$ (at least one of the two events occur)

$\quad - P$ (the two events occur simultaneously)

Corollary 3 The addition theorem can be applied for more than two events. If A, B, and C are three events of a sample space S then the probability of occurrence of at least one of them is given by

$$\begin{aligned} P(A\cup B\cup C) &= P\left[A\cup(B\cup C)\right] \\ &= P(A)+P(B\cup C)-P\left[A\cap(B\cup C)\right] \\ &= P(A)+P(B\cup C)-P\left[A\cap B)\cup(A\cap C)\right] \\ &= P(A)+P(B)+P(C)-P(B\cap C)-P(A\cap B)-P(A\cap C)+P(A\cap B\cap C) \\ &\qquad \left[\text{Applying Theorem 4 on second and third term}\right] \end{aligned}$$

Alternately, the probability of occurrence of at least one of the three events can also be written as

$$P(A\cup B\cup C)=1-P(\overline{A}\cap\overline{B}\cap\overline{C})$$

If A, B, and C are mutually exclusive events,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

Corollary 4 The probability of occurrence of at least two of the three events is given by

$$\begin{aligned} P\left[A\cap B)\cup(B\cap C)\cup(A\cap C)\right] &= P(A\cap B)+P(B\cap C)+P(A\cap C)-3P(A\cap B\cap C) \\ &\quad +P(A\cap B\cap C) \qquad \text{[Using Corollary 3]} \\ &= P(A\cap B)+P(B\cap C)+P(A\cap C)-2P(A\cap B\cap C) \end{aligned}$$

Corollary 5 The probability of occurrence of exactly two of the three events is given by

$$\begin{aligned} &P\left[A\cap B\cap\overline{C})\cup(A\cap\overline{B}\cap C)\cup(\overline{A}\cap B\cap C)\right] \\ &\quad = P\left[(A\cap B)\cup(B\cap C)\cup(A\cap C)\right]-P(A\cap B\cap C) \quad \text{[Using Corollary 2]} \\ &\quad = P(A\cap B)+P(B\cap C)+P(A\cap C)-3P(A\cap B\cap C) \quad \text{[Using Corollary 4]} \end{aligned}$$

Corollary 6 The probability of occurrence of exactly one of the three events is given by

$$P\left[(A\cap\overline{B}\cap\overline{C})\cup(\overline{A}\cap B\cap\overline{C})\cup(\overline{A}\cap\overline{B}\cap C)\right]$$

$= P$(at least one of the three event occur) – P (at least two of the three events occur)

$$= P(A)+P(B)+P(C)-2P(A\cap B)-2P(B\cap C)-2P(A\cap C)+3P(A\cap B\cap C)$$

Example 1

A card is drawn from a well-shuffled pack of cards. What is the probability that it is either a spade or an ace?

Solution

Let A and B be the events of getting a spade and an ace card respectively.

$$P(A) = \frac{^{13}C_1}{^{52}C_1} = \frac{13}{52}$$

$$P(B) = \frac{^{4}C_1}{^{52}C_1} = \frac{4}{52}$$

$$P(A \cap B) = \frac{^{1}C_1}{^{52}C_1} = \frac{1}{52}$$

Probability of getting either a spade or an ace card

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} \\ &= \frac{4}{13} \end{aligned}$$

Example 2

Two cards are drawn from a pack of cards. Find the probability that they will be both red or both pictures.

Solution

Let A and B be the events that both cards drawn are red and pictures respectively.

$$P(A) = \frac{^{26}C_2}{^{52}C_2} = \frac{325}{1326}$$

$$P(B) = \frac{^{12}C_2}{^{52}C_2} = \frac{66}{1326}$$

$$P(A \cap B) = \frac{^{6}C_2}{^{52}C_2} = \frac{15}{1326}$$

Probability that both cards drawn are red or pictures

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

$$= \frac{325}{1326} + \frac{66}{1326} - \frac{15}{1326}$$
$$= \frac{188}{663}$$

Example 3

The probability that a contractor will get a plumbing contract is $\frac{2}{3}$ *and the probability that he will not get an electric contract is* $\frac{5}{9}$. *If the probability of getting any one contract is* $\frac{4}{5}$, *what is the probability that he will get both the contracts?*

Solution

Let A and B be the events that the contractor will get plumbing and electric contracts respectively.

$$P(A) = \frac{2}{3}, \quad P(\overline{B}) = \frac{5}{9}, \quad P(A \cup B) = \frac{4}{5}$$
$$P(B) = 1 - P(\overline{B}) = 1 - \frac{5}{9} = \frac{4}{9}$$

Probability that the contractor will get any one contract

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

Probability that the contractor will get both the contracts

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$
$$= \frac{2}{3} + \frac{4}{9} - \frac{4}{5}$$
$$= \frac{14}{45}$$

Example 4

A person applies for a job in two firms A and B, the probability of his being selected in the firm A is 0.7 and being rejected in the firm B is 0.5. The probability of at least one of the applications being rejected is 0.6. What is the probability that he will be selected in one of the two firms?

Solution

Let A and B be the events that the person is selected in firms A and B respectively.

$$P(A) = 0.7, \qquad P(\overline{B}) = 0.5, \qquad P(\overline{A} \cup \overline{B}) = 0.6$$

$$P(\overline{A}) = 1 - P(A) = 1 - 0.7 = 0.3$$

$$P(B) = 1 - P(\overline{B}) = 1 - 0.5 = 0.5$$

$$P(\overline{A} \cup \overline{B}) = P(\overline{A}) + P(\overline{B}) - P(\overline{A} \cap \overline{B}) \qquad \ldots(1)$$

Probability that the person will be selected in one of the two firms

$$\begin{aligned} P(A \cup B) &= 1 - P(\overline{A} \cap \overline{B}) \\ &= 1 - \left[P(\overline{A}) + P(\overline{B}) - P(\overline{A} \cup \overline{B})\right] \qquad \text{[Using Eq. (1)]} \\ &= 1 - (0.3 + 0.5 - 0.6) \\ &= 0.8 \end{aligned}$$

Example 5

In a group of 1000 *persons, there are* 650 *who can speak Hindi,* 400 *can speak English, and* 150 *can speak both Hindi and English. If a person is selected at random, what is the probability that he speaks (i) Hindi only, (ii) English only, (iii) only of the two languages, and (iv) at least one of the two languages?*

Solution

Let A and B be the events that a person selected at random speaks Hindi and English respectively.

$$P(A) = \frac{650}{1000}, \quad P(B) = \frac{400}{1000}, P(A \cap B) = \frac{150}{1000}$$

(i) Probability that a person selected at random speaks Hindi only

$$\begin{aligned} P(A \cap \overline{B}) &= P(A) - P(A \cap B) \\ &= \frac{650}{1000} - \frac{150}{1000} \\ &= \frac{1}{2} \end{aligned}$$

(ii) Probability that a person selected at random speaks English only

$$\begin{aligned} P(\overline{A} \cap B) &= P(B) - P(A \cap B) \\ &= \frac{400}{1000} - \frac{150}{1000} \\ &= \frac{1}{4} \end{aligned}$$

(iii) Probability that a person selected at random speaks only one of the languages.

$$P\left[(A\cap\overline{B})\cup(\overline{A}\cap B)\right]=P(A)+P(B)-2P(A\cap B)$$
$$=\frac{650}{1000}+\frac{400}{1000}-2\left(\frac{150}{1000}\right)$$
$$=\frac{3}{4}$$

(iv) Probability that a person selected at random speaks at least one of the two languages

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$
$$=\frac{650}{1000}+\frac{400}{1000}-\frac{150}{1000}$$
$$=\frac{9}{10}$$

Example 6

A box contains 4 *white,* 6 *red,* 5 *black balls, and* 5 *balls of other colours. Two balls are drawn from the box at random. Find the probability that (i) both are white or both are red, and (ii) both are red or both are black.*

Solution

Let A, B, and C be the events of drawing white, red and black balls from the box respectively.

$$P(A)=\frac{^{4}C_2}{^{20}C_2}=\frac{3}{95}$$

$$P(B)=\frac{^{6}C_2}{^{20}C_2}=\frac{3}{38}$$

$$P(C)=\frac{^{5}C_2}{^{20}C_2}=\frac{1}{19}$$

(i) Probability that the both balls are white or both are red

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$
$$=\frac{3}{95}+\frac{3}{38}-0$$
$$=\frac{21}{190}$$

(ii) Probability that both balls are red or both are black

$$P(B \cup C) = P(B) + P(C) - P(B \cap C)$$
$$= \frac{3}{38} + \frac{1}{19} - 0$$
$$= \frac{5}{38}$$

Example 7

Three students A, B, C are in a running race. A and B have the same probability of winning and each is twice as likely to win as C. Find the probability that B or C wins.

Solution

Let A, B, and C be the events that students A, B, and C win the race respectively.

$$P(A) = P(B) = 2P(C)$$
$$P(A) + P(B) + P(C) = 1$$
$$2P(C) + 2P(C) + P(C) = 1$$
$$P(C) = \frac{1}{5}$$

$\therefore \qquad P(A) = \frac{2}{5}$ and $P(B) = \frac{2}{5}$

Probability that student B or C wins

$$P(B \cup C) = P(B) + P(C) - P(B \cap C)$$
$$= \frac{2}{5} + \frac{1}{5} - 0$$
$$= \frac{3}{5}$$

Example 8

A card is drawn from a pack of 52 cards. Find the probability of getting a king or a heart or a red card.

Solution

Let A, B and C be the events that the card drawn is a king, a heart and a red card respectively.

$$P(A)=\frac{{}^{4}C_1}{{}^{52}C_1}=\frac{4}{52}$$

$$P(B)=\frac{{}^{13}C_1}{{}^{52}C_1}=\frac{13}{52}$$

$$P(C)=\frac{{}^{26}C_1}{{}^{52}C_1}=\frac{26}{52}$$

$$P(A\cap B)=\frac{{}^{1}C_1}{{}^{52}C_1}=\frac{1}{52}$$

$$P(B\cap C)=\frac{{}^{13}C_1}{{}^{52}C_1}=\frac{13}{52}$$

$$P(A\cap C)=\frac{{}^{2}C_1}{{}^{52}C_1}=\frac{2}{52}$$

$$P(A\cap B\cap C)=\frac{{}^{1}C_1}{{}^{52}C_1}=\frac{1}{52}$$

Probability that the card drawn is a king or a heart or a red card.

$$\begin{aligned}P(A\cup B\cup C) &= P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(A\cap C)\\ &\qquad +P(A\cap B\cap C)\\ &=\frac{4}{52}+\frac{13}{52}+\frac{26}{52}-\frac{1}{52}-\frac{13}{52}-\frac{2}{52}+\frac{1}{52}\\ &=\frac{7}{3}\end{aligned}$$

Example 9

From a city, 3 newspapers A, B, C are being published. A is read by 20%, B is read by 16%, C is read by 14%, both A and B are read by 8%, both A and C are read by 5%, both B and C are read by 4% and all three A, B, C are read by 2%. What is the probability that a randomly chosen person (i) reads at least one of these newspapers, and (ii) reads one of these newspapers?

Solution

Let A, B, and C be the events that the person reads newspapers A, B, and C respectively.

$$P(A)=0.2, \qquad P(B)=0.16 \qquad P(C)=0.14$$

$$P(A\cap B)=0.08, \qquad P(A\cap B)=0.05, \qquad P(B\cap C)=0.04$$

$$P(A\cap B\cap C)=0.02$$

(i) Probability that the person reads at least one of these newspapers

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

$$= 0.2 + 0.16 + 0.14 - 0.08 - 0.05 - 0.04 + 0.02$$

$$= 0.35$$

(ii) Probability that the person reads none of these newspapers

$$P(\bar{A} \cap \bar{B} \cap \bar{C}) = 1 - P(A \cup B \cup C)$$

$$= 1 - 0.35$$

$$= 0.65$$

Alternatively, the problem can be solved by a Venn diagram (Fig. 1.2).

(i) P(the person reads at least one paper) = $1 - \dfrac{65}{100} = 0.35$

(ii) P(the person reads none of these papers) = 0.65

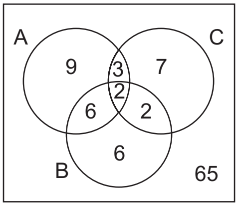


Fig. 1.2

EXERCISE 1.2

1. The probability that a student passes a Physics test is $\dfrac{2}{3}$ and the probability that he passes both Physics and English tests is $\dfrac{14}{45}$. The probability that he passes at least one test is $\dfrac{4}{5}$. What is the probability that the student passes the English test?

$$\left[\textbf{Ans.: } \frac{4}{9}\right]$$

2. What is the probability of drawing a black card or a king from a well-shuffled pack of playing cards?

$$\left[\textbf{Ans.: } \frac{7}{13}\right]$$

3. A pair of unbiased dice is thrown. Find the probability that (i) the sum of spots is either 5 or 10, and (ii) either there is a doublet or a sum less than 6.

$$\left[\textbf{Ans.: } \text{(i) } \frac{7}{36} \quad \text{(ii) } \frac{7}{18}\right]$$

4. From a pack of well-shuffled cards, a card is drawn at random. What is the probability that the card drawn is a diamond card or a king card?

$$\left[\textbf{Ans.: } \frac{4}{13}\right]$$

5. A bag contains 6 red, 5 blue, 3 white, and 4 black balls. A ball is drawn at random. Find the probability that the ball is (i) red or black, and (ii) neither red or black.

$$\left[\textbf{Ans.: } \text{(i) } \frac{5}{9} \quad \text{(ii) } \frac{4}{9}\right]$$

6. There are 100 lottery tickets, numbered from 1 to 100. One of them is drawn at random. What is the probability that the number on it is a multiple of 5 or 7?

$$\left[\textbf{Ans.: } \frac{8}{25}\right]$$

7. From a group of 6 boys and 4 girls, a committee of 3 is to be formed. Find the probability that the committee will include (i) all three boys or all three girls, (ii) at most two girls, and (iii) at least one girl.

$$\left[\textbf{Ans.: } \text{(i) } \frac{1}{5} \quad \text{(ii) } \frac{29}{30} \quad \text{(iii) } \frac{5}{6}\right]$$

8. From a pack of 52 cards, three cards are drawn at random. Find the probability that (i) all three will be aces or all three kings, (ii) all three are pictures or all three are aces, (iii) none is a picture, (iv) at least one is a picture, (v) none is a spade, (vi) at most two are spades, and (vii) at least one is a spade.

$$\left[\begin{array}{l}\textbf{Ans.: } \text{(i) } \dfrac{2}{5225} \quad \text{(ii) } \dfrac{56}{5225} \quad \text{(iii) } \dfrac{38}{85} \quad \text{(iv) } \dfrac{47}{85} \\ \quad\quad \text{(v) } \dfrac{703}{1700} \quad \text{(vi) } \dfrac{839}{850} \quad \text{(vii) } \dfrac{997}{1700}\end{array}\right]$$

9. From a set of 16 cards numbered 1 to 16, one card is drawn at random. Find the probability that (i) the number obtained is divisible by 3 or 7, and (ii) not divisible by 3 and 7.

$$\left[\textbf{Ans.: } \text{(i) } \frac{7}{16} \quad \text{(ii) } \frac{9}{16}\right]$$

10. There are 12 bulbs in a basket of which 4 are working. A person tries to fit them in 3 sockets choosing 3 of the bulbs at random. What is the probability that there will be (i) some light, and (ii) no light in the room?

$$\left[\textbf{Ans.:} \text{ (i) } \frac{41}{55} \quad \text{(ii) } \frac{14}{55}\right]$$

Theorem 5 Multiplicative Law or Compound Law of Probability

A compound event is the result of the simultaneous occurrence of two or more event. The probability of a compound event depends upon whether the events are independent or not. Hence, there are two theorems:

(a) Conditional Probability Theorem

(b) Multiplicative Theorem for Independent Events

(a) Conditional Probability Theorem For any two events A and B in a sample space S, the probability of their simultaneous occurrence, i.e., both the events occurrings simultaneously is given by

$$P(A \cap B) = P(A)\, P(B/A)$$

or $$P(A \cap B) = P(B)\, P(A/B)$$

where $P(B/A)$ is the conditional probability of B given that A has already occurred. $P(A/B)$ is the conditional probability of A given that B has already occurred.

(b) Multiplicative Theorem for Independent Events If A and B are two independent events, the probability of their simultaneous occurrence is given by

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

$$P(A \cap B) = P(B)\, P(A/B) \qquad ...(1.1)$$

Proof $A = (A \cap B) \cup (A \cap \overline{B})$

Since $(A \cap B)$ and $(A \cap \overline{B})$ are mutually exclusive events,

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) \qquad \text{[Using Axiom III]}$$

$$= P(B)\, P(A/B) + P(\overline{B})\, P(A/\overline{B})$$

If A and B are independent events, the proportion of A's in B is equal to proportion of A's in $\overline{B}$, i.e., $P(A/B) = P(A/\overline{B})$.

$$P(A) = P(A/B)\left[P(B) + P(\overline{B})\right]$$

$$= P(A/B)$$

Substituting in Eq. (1.1),

$\therefore \quad P(A \cap B) = P(A)\, P(B)$

Remark The additive law is used to find the probability of A or B, i.e., $P(A \cup B)$. The multiplicative law is used to find the probability of A and B, i.e., $P(A \cap B)$.

Corollary 1 If A, B and C are three events then

$$P(A \cap B \cap C) = P(A)\, P(B/A)\, P\left[C/(A \cap B)\right]$$

If A, B and C are independent events,

$$P(A \cap B \cap C) = P(A)\, P(B)\, P(C)$$

Corollary 2 If A and B are independent events then A and $\overline{B}$, $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$ are also independent.

Corollary 3 The probability of occurrence of at least one of the events A, B, C is given by

$$P(A \cup B \cup C) = 1 - P(\overline{A} \cap \overline{B} \cap \overline{C})$$

If A, B, and C are independent events, their complements will also be independent.

$$P(A \cup B \cup C) = 1 - P(\overline{A})\, P(\overline{B})\, P(\overline{C})$$

Pairwise Independence and Mutual Independence The events A, B and C are mutually independent if the following conditions are satisfied simultaneously:

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

$$P(B \cap C) = P(B)\, P(C)$$

$$P(A \cap C) = P(A)\, P(C)$$

and $$P(A \cap B \cap C) = P(A)\, P(B)\, P(C)$$

If the last condition is not satisfied, the events are said to be pairwise independent. Hence, mutually independent events are always pairwise independent but not vices versa.

Example 1

If A and B are two events such that $P(A) = \dfrac{2}{3}, P(\overline{A} \cap B) = \dfrac{1}{6}$ *and* $P(A \cap B) = \dfrac{1}{3}$, *find* $P(B), P(A \cup B), P(A/B), P(B/A), P(\overline{A} \cup B)$ *and* $P(\overline{B})$. *Also, examine whether the events A and B are (i) equally likely, (ii) exhaustive, (iii) mutually exclusive, and (iv) independent.*

Solution

$$\begin{aligned} P(B) &= P(\overline{A} \cap B) + P(A \cap B) \\ &= \frac{1}{6} + \frac{1}{3} \\ &= \frac{1}{2} \end{aligned}$$

$$\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
&= \frac{2}{3} + \frac{1}{2} - \frac{1}{3} \\
&= \frac{5}{6} \\
P(A/B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{\left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right)} \\
&= \frac{2}{3} \\
P(B/A) &= \frac{P(A \cap B)}{P(A)} \\
&= \frac{\left(\frac{1}{3}\right)}{\left(\frac{2}{3}\right)} \\
&= \frac{1}{2} \\
P(\overline{A} \cup B) &= P(\overline{A}) + P(B) - P(\overline{A} \cap B) \\
&= \frac{1}{3} + \frac{1}{2} - \frac{1}{6} \\
&= \frac{2}{3} \\
P(\overline{A} \cap \overline{B}) &= 1 - P(A \cup B) \\
&= 1 - \frac{5}{6} \\
&= \frac{1}{6} \\
P(\overline{B}) &= 1 - P(B) \\
&= 1 - \frac{1}{2} \\
&= \frac{1}{2}
\end{aligned}$$

(i) Since $P(A) \neq P(B)$, A and B are not equally like events.

(ii) Since $P(A \cup B) \neq 1$, A and B are not exhaustive events.

(iii) Since $P(A \cap B) \neq 0$, A and B are not mutually exclusive events.

(iv) Since $P(A \cap B) = P(A)\,P(B)$, A and B are independent events.

Example 2

If A and B are two events such that $P(A) = 0.3$, $P(B) = 0.4$, $P(A \cap B) = 0.2$, *find (i)* $P(A \cup B)$, *(ii)* $P(\overline{A}/B)$, *and (iii)* $P(A/\overline{B})$.

Solution

(i) $$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.3 + 0.4 - 0.2 \\ &= 0.5 \end{aligned}$$

(ii) $$\begin{aligned} P(\overline{A}/B) &= \frac{P(\overline{A} \cap B)}{P(B)} \\ &= \frac{P(B) - P(A \cap B)}{P(B)} \\ &= \frac{0.4 - 0.2}{0.4} \\ &= 0.5 \end{aligned}$$

(iii) $$\begin{aligned} P(A/\overline{B}) &= \frac{P(A \cap \overline{B})}{P(\overline{B})} \\ &= \frac{P(A) - P(A \cap B)}{1 - P(B)} \\ &= \frac{0.3 - 0.2}{1 - 0.4} \\ &= \frac{1}{6} \end{aligned}$$

Example 3

If A and B are two events with $P(A) = \frac{1}{3}, P(B) = \frac{1}{4}, P(A \cap B) = \frac{1}{12}$.

Find (i) P(A/B), (ii) P(B/A), (iii) $P(B/\overline{A})$, *and (iv)* $P(A \cap \overline{B})$.

Solution

(i) $$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{12}}{\frac{1}{4}} = \frac{1}{3}$$

(ii) $P(B/A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{\frac{1}{12}}{\frac{1}{3}} = \dfrac{1}{4}$

(iii) $$P(B/\bar{A}) = \frac{P(B \cap \bar{A})}{P(\bar{A})}$$

$$= \frac{P(B) - P(B \cap A)}{1 - P(A)}$$

$$= \frac{\frac{1}{4} - \frac{1}{12}}{1 - \frac{1}{3}}$$

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

(iv) $$P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

$$= \frac{1}{3} - \frac{1}{12}$$

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

Example 4

Find the probability of drawing a queen and a king from a pack of cards in two consecutive draws, the cards drawn not being replaced.

Solution

Let A be the event that the card drawn is a queen.

$$P(A) = \frac{{}^4C_1}{{}^{52}C_1} = \frac{4}{52} = \frac{1}{13}$$

Let B be the event that the cards drawn are a king in the second draw given that the first card drawn is a queen.

$$P(B/A) = \frac{{}^4C_1}{{}^{51}C_1} = \frac{4}{51}$$

Probability that the cards drawn are a queen and a king

$$P(A \cap B) = P(A)\, P(B/A)$$

$$= \frac{4}{52} \times \frac{4}{51}$$

$$= \frac{4}{663}$$

Example 5

A bag contains 3 red and 4 white balls. Two draws are made without replacement. What is the probability that both the balls are red?

Solution

Let A be the event that the ball drawn is red in the first draw.

$$P(A) = \frac{3}{7}$$

Let B be the event that the ball drawn is red in the second draw given that the first ball drawn is red.

$$P(B/A) = \frac{2}{6}$$

Probability that both the balls are red

$$\begin{aligned} P(A \cap B) &= P(A)\, P(B/A) \\ &= \frac{3}{7} \times \frac{2}{6} \\ &= \frac{1}{7} \end{aligned}$$

Example 6

A bag contains 8 red and 5 white balls. Two successive draws of 3 balls each 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 draw will give 3 white and the second, 3 red balls.

Solution

Let A be the event that all 3 balls obtained at the first draw are white, and B be the event that all the 3 balls obtained at the second draw are red.

(i) When balls are replaced before the second trial,

$$P(A) = \frac{{}^{5}C_3}{{}^{13}C_3} = \frac{5}{143}$$

$$P(B) = \frac{{}^{8}C_3}{{}^{13}C_3} = \frac{28}{143}$$

Probability that the first draw will give 3 white and the second, 3 red balls

$$P(A \cap B) = P(A)\, P(B)$$
$$= \frac{5}{143} \times \frac{28}{143}$$
$$= \frac{140}{20449}$$

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

$$P(B/A) = \frac{8_{C_3}}{10_{C_3}} = \frac{7}{15}$$

Probability that the first draw will give 3 white and the second, 3 red balls

$$P(A \cap B) = P(A)\, P(B/A)$$
$$= \frac{5}{143} \times \frac{7}{15}$$
$$= \frac{7}{429}$$

Example 7

From a bag containing 4 white and 6 black balls, two balls are drawn at random. If the balls are drawn one after the other without replacements, find the probability that the first ball is white and the second ball is black.

Solution

Let A be the event that the first ball drawn is white and B be the event that the second ball drawn is black given that the first ball drawn is white.

$$P(A) = \frac{4}{10}$$
$$P(B/A) = \frac{6}{9}$$

Probability that the first ball is white and the second ball is black.

$$P(A \cap B) = P(A)\, P(B/A)$$
$$= \frac{4}{10} \times \frac{6}{9}$$
$$= \frac{4}{15}$$

Example 8

Data on readership of a certain magazine show that the proportion of male readers under 35 *is* 0.40 *and that over* 35 *is* 0.20*. If the proportion of readers under* 35 *is* 0.70*, find the probability of subscribers that are females over* 35 *years. Also, calculate the probability that a randomly selected male subscriber is under* 35 *years of age.*

Solution

Let A be the event that the reader of the magazine is a male. Let B be the event that reader of the magazine is over 35 years of age.

$$P(A \cap \overline{B}) = 0.40, \qquad P(A \cap B) = 0.20, \qquad P(\overline{B}) = 0.7$$

$$\begin{aligned} P(B) &= 1 - P(\overline{B}) \\ &= 1 - 0.7 \\ &= 0.3 \end{aligned}$$

(i) Probability of subscribers that are females over 35 years

$$\begin{aligned} P(\overline{A} \cap B) &= P(B) - P(A \cap B) \\ &= 0.3 - 0.2 \\ &= 0.1 \end{aligned}$$

(ii) Probability that a randomly selected male subscriber is under 35 years of age

$$\begin{aligned} P(\overline{B}/A) &= \frac{P(A \cap \overline{B})}{P(A)} \\ &= \frac{P(A \cap \overline{B})}{P(A \cap B) + P(A \cap \overline{B})} \\ &= \frac{0.4}{0.2 + 0.4} \\ &= \frac{0.4}{0.6} \\ &= \frac{2}{3} \end{aligned}$$

Example 9

From a city population, the probability of selecting (a) a male or a smoker is $\frac{7}{10}$*, (b) a male smoker is* $\frac{2}{5}$*, and (c) a male, if a smoker is* *already selected, is* $\frac{2}{3}$. *Find the probability of selecting (i) a nonsmoker, (ii) a male, and (iii) a smoker, if a male is first selected.*

Solution

Let A be the event that a male is selected. Let B be the event that a smoker is selected.

$$P(A \cup B) = \frac{7}{10},\ \ P(A \cap B) = \frac{2}{5},\ \ P(A/B) = \frac{2}{3}$$

(i) Probability of selecting a nonsmoker

$$\begin{aligned} P(\overline{B}) &= 1 - P(B) \\ &= 1 - \frac{P(A \cap B)}{P(A/B)} \\ &= 1 - \frac{\left(\frac{2}{5}\right)}{\left(\frac{2}{3}\right)} \\ &= \frac{2}{5} \end{aligned}$$

(ii)

$$\begin{aligned} P(B) &= 1 - P(\overline{B}) \\ &= 1 - \frac{2}{5} \\ &= \frac{3}{5} \end{aligned}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \ldots(1)$$

Probability of selecting a male

$$\begin{aligned} P(A) &= P(A \cup B) + P(A \cap B) - P(B) \qquad \text{[Using Eq. (1)]} \\ &= \frac{7}{10} + \frac{2}{5} - \frac{3}{5} \\ &= \frac{1}{2} \end{aligned}$$

(iii) Probability of selecting a smoker if a male is first selected

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{\left(\frac{2}{5}\right)}{\left(\frac{1}{2}\right)}$$

$$= \frac{4}{5}$$

Example 10

Sixty per cent of the employees of the XYZ corporation are college graduates. Of these, ten percent are in sales. Of the employee who did not graduate from college, eighty percent are in sales. What is the probability that

(i) an employee selected at random is in sales?

(ii) an employee selected at random is neither in sales nor a college graduate?

Solution

Let A be the event that an employee is a college graduate. Let B be the event that an employee is in sales.

$$P(A) = 0.6,\ \ P(B/A) = 0.10,\ \ P(B/\overline{A}) = 0.8$$

$$P(\overline{A}) = 1 - P(A) = 1 - 0.60 = 0.40$$

(i) Probability that an employee is in sales

$$\begin{aligned} P(B) &= P(A \cap B) + P(\overline{A} \cap B) \\ &= P(A)\ P(B/A) + P(\overline{A})\ P(B/\overline{A}) \\ &= (0.6 \times 0.1) + (0.40 \times 0.80) \\ &= 0.38 \end{aligned}$$

(ii) Probability that an employee is neither in sales nor a college graduate

$$\begin{aligned} P(\overline{A} \cap \overline{B}) &= 1 - P(A \cup B) \\ &= 1 - \left[P(A) + P(B) - P(A \cap B)\right] \\ &= 1 - \left[P(A) + P(B) - P(A)\ P(B/A)\right] \\ &= 1 - \left[0.60 + 0.38 - (0.60 \times 0.10)\right] \\ &= 0.08 \end{aligned}$$

Example 11

If A and B are two events such that $P(A)=\frac{3}{8}, P(B)=\frac{5}{8}$ *and* $P(A\cup B)=\frac{3}{4}$, *find P(A/B) and P(B/A). Show whether A and B are independent.*

Solution

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

$$\frac{3}{4}=\frac{3}{8}+\frac{5}{8}-P(A\cap B)$$

$$P(A\cap B)=\frac{1}{4}$$

$$P(A/B)=\frac{P(A\cap B)}{P(B)}$$

$$=\frac{\left(\frac{1}{4}\right)}{\left(\frac{5}{8}\right)}$$

$$=\frac{2}{5}$$

$$P(B/A)=\frac{P(A\cap B)}{P(A)}$$

$$=\frac{\left(\frac{1}{4}\right)}{\left(\frac{3}{8}\right)}$$

$$=\frac{2}{3}$$

$$P(A)\ P(B)=\frac{3}{8}\times\frac{5}{8}=\frac{15}{64}$$

$$P(A\cap B)\neq P(A)\ P(B)$$

Hence, the events A and B are not independent.

Example 12

The probability that a student A solves a mathematics problem is $\frac{2}{5}$ *and the probability that a student B solves it is* $\frac{2}{3}$. *What is the probability* *that (i) the problem is not solved, (ii) the problem is solved, and (iii) both A and B, working independently of each other, solve the problem?*

Solution

Let A and B be events that students A and B solve the problem respectively.

$$P(A)=\frac{2}{5},\ \ P(B)=\frac{2}{3}$$

Events A and B are independent.

Probability that the student A does not solve the problem

$$\begin{aligned}P(\overline{A})&=1-P(A)\\&=1-\frac{2}{5}\\&=\frac{3}{5}\end{aligned}$$

Probability that the student B does not solve the problem

$$\begin{aligned}P(\overline{B})&=1-P(B)\\&=1-\frac{2}{3}\\&=\frac{1}{3}\end{aligned}$$

(i) Probability that the problem is not solved

$$\begin{aligned}P(\overline{A}\cap\overline{B})&=P(\overline{A})\,P(\overline{B})\\&=\frac{3}{5}\times\frac{1}{3}\\&=\frac{1}{5}\end{aligned}$$

(ii) Probability that the problem is solved

$$\begin{aligned}P(A\cup B)&=1-P(\overline{A}\cap\overline{B})\\&=1-\frac{1}{5}\\&=\frac{4}{5}\end{aligned}$$

(iii) Probability that both A and B solve the problem

$$\begin{aligned}P(A\cap B)&=P(A)\,P(B)\\&=\frac{2}{5}\times\frac{2}{3}\\&=\frac{4}{15}\end{aligned}$$

Example 13

The probability that the machine A will perform a usual function in 5 years' time is $\frac{1}{4}$, *while the probability that the machine B will perform the function in 5 years' time is* $\frac{1}{3}$. *Find the probability that both machines will perform the usual function.*

Solution

Let A and B be the events that machines A and B will perform the usual function respectively.

$$P(A) = \frac{1}{4}$$

$$P(B) = \frac{1}{3}$$

Events A and B are independent.

Probability that both machines will perform the usual function

$$\begin{aligned} P(A \cap B) &= P(A)\,P(B) \\ &= \frac{1}{4} \times \frac{1}{3} \\ &= \frac{1}{12} \end{aligned}$$

Example 14

A person A is known to hit a target in 3 *out of* 4 *shots, whereas another person B is known to hit the same target in* 2 *out of* 3 *shots. Find the probability of the target being hit at all when they both try.*

[Summer 2015]

Solution

Let A and B be the events that the persons A and B hit the target respectively.

$$P(A) = \frac{3}{4}$$

$$P(B) = \frac{2}{3}$$

Events A and B are independent.

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

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

Probability that the target is not hit at all

$$P(\overline{A} \cap \overline{B}) = P(\overline{A})\, P(\overline{B})$$
$$= \frac{1}{4} \times \frac{1}{3}$$
$$= \frac{1}{12}$$

Probability that the target is hit at all when they both try

$$P(A \cup B) = 1 - P(\overline{A} \cap \overline{B})$$
$$= 1 - \frac{1}{12}$$
$$= \frac{11}{12}$$

Aliter

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= P(A) + P(B) - P(A)\, P(B) \quad [\because \quad A \text{ and } B \text{ independent}]$$
$$= \frac{3}{4} + \frac{2}{3} - \frac{3}{4} \times \frac{2}{3}$$
$$= \frac{11}{12}$$

Example 15

The odds against A speaking the truth are 4 : 6 *while the odds in favour of B speaking the truth are* 7 : 3. *What is the probability that A and B contradict each other in stating the same fact?*

Solution

Let A and B be events that A and B speak the truth respectively.

$$P(A) = \frac{6}{10}$$
$$P(B) = \frac{7}{10}$$

Events A and B are independent.

Probability that A speaks a lie $= P(\overline{A}) = 1 - P(A) = 1 - \frac{6}{10} = \frac{4}{10}$

Probability that B speaks a lie $= P(\overline{B}) = 1 - P(B) = 1 - \frac{7}{10} = \frac{3}{10}$

Probability that A and B contradict each other

$$P\left[(A\cap \bar{B})\cup(\bar{A}\cap B)\right] = P(A\cap \bar{B}) + P(\bar{A}\cap B) \quad \left[\because\ (A\cap \bar{B}) \text{ and } (\bar{A}\cap B) \text{ are mutually exclusive events}\right]$$

$$= P(A)\, P(\bar{B}) + P(\bar{A})\, P(B)$$

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

$$= \frac{23}{50}$$

Example 16

An urn contains 10 red, 5 white and 5 blue balls. Two balls are drawn at random. Find the probability that they are not of the same colour.

Solution

Let A, B, and C be the events that two balls drawn at random be of the same colour, i.e., red, white, and blue respectively.

$$P(A) = \frac{10_{C_2}}{20_{C_2}} = \frac{9}{38}$$

$$P(B) = \frac{5_{C_2}}{20_{C_2}} = \frac{1}{19}$$

$$P(C) = \frac{5_{C_2}}{20_{C_2}} = \frac{1}{19}$$

Events A, B, and C are independent.

Probability that both balls drawn are of same colour

$$P(A\cup B\cup C) = P(A)+P(B)+P(C)$$

$$= \frac{9}{38}+\frac{1}{19}+\frac{1}{19}$$

$$= \frac{13}{38}$$

Probability that both balls drawn are not of the same colour

$$P(\bar{A}\cap\bar{B}\cap\bar{C}) = 1 - P(A\cup B\cup C)$$

$$= 1-\frac{13}{38}$$

$$= \frac{25}{38}$$

Example 17

A problem in statistics is given to three students A, B and C, whose chances of solving it independently are $\frac{1}{2}, \frac{1}{3}$, and $\frac{1}{4}$ respectively. Find the probability that

(i) the problem is solved

(ii) at least two of them are able to solve the problem

(iii) exactly two of them are able to solve the problem

(iv) exactly one of them is able to solve the problem

Solution

Let A, B, and C be the events that students A, B, and C solve the problem respectively.

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{3}, \quad P(C) = \frac{1}{4}$$

Events A, B, and C are independent.

(i) Probability that the problem is solved or at least one of them is able to solve the problem is same.

$$\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
&\quad + P(A \cap B \cap C) \\
&= P(A) + P(B) + P(C) - P(A)\,P(B) - P(A)\,P(C) - P(B)\,P(C) \\
&\quad + P(A)\,P(B)\,P(C) \\
&= \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \left(\frac{1}{2} \times \frac{1}{3}\right) - \left(\frac{1}{2} \times \frac{1}{4}\right) - \left(\frac{1}{3} \times \frac{1}{4}\right) + \left(\frac{1}{2} \times \frac{1}{3} \times \frac{1}{4}\right) \\
&= \frac{3}{4}
\end{aligned}$$

(ii) Probability that at least two of them are able to solve the problem

$$\begin{aligned}
P\left[(A \cap B) \cup (B \cap C) \cup (A \cap C)\right] &= P(A \cap B) + P(B \cap C) + P(A \cap C) - 2P(A \cap B \cap C) \\
&= P(A)\,P(B) + P(B)\,P(C) + P(A)\,P(C) \\
&\quad - 2P(A)\,P(B)\,P(C) \\
&= \left(\frac{1}{2} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{1}{4}\right) + \left(\frac{1}{2} \times \frac{1}{4}\right) - 2\left(\frac{1}{2} \times \frac{1}{3} \times \frac{1}{4}\right) \\
&= \frac{7}{24}
\end{aligned}$$

(iii) Probability that exactly two of them are able to solve the problem

$$P\left[(A\cap B\cap \bar{C})\cup(A\cap \bar{B}\cap C\cup(\bar{A}\cap B\cap C\right]$$
$$= P(A\cap B)+P(B\cap C)+P(A\cap C)-3P(A\cap B\cap C)$$
$$= P(A)\,P(B)+P(B)\,P(C)+P(A)\,P(C)-3P(A)\,P(B)\,P(C)$$
$$=\left(\frac{1}{2}\times\frac{1}{3}\right)+\left(\frac{1}{3}\times\frac{1}{4}\right)+\left(\frac{1}{2}\times\frac{1}{4}\right)-3\left(\frac{1}{2}\times\frac{1}{3}\times\frac{1}{4}\right)$$
$$=\frac{1}{4}$$

(iv) Probability that exactly one of them is able to solve the problem

$$P\left[A\cap \bar{B}\cap \bar{C})\cup(\bar{A}\cap B\cap \bar{C})\cup(\bar{A}\cap \bar{B}\cap C)\right]$$
$$= P(A)+P(B)+P(C)-2P(A\cap B)-2P(B\cap C)-2P(A\cap C)+3P(A\cap B\cap C)$$
$$=\frac{1}{2}+\frac{1}{3}+\frac{1}{4}-2\left(\frac{1}{2}\times\frac{1}{3}\right)-2\left(\frac{1}{3}\times\frac{1}{4}\right)-2\left(\frac{1}{2}\times\frac{1}{4}\right)+3\left(\frac{1}{2}\times\frac{1}{3}\times\frac{1}{4}\right)$$
$$=\frac{11}{24}$$

Example 18

A husband and wife appeared in an interview for two vacancies in an office. The probability of the husband's selection is $\frac{1}{7}$ *and that of the wife's selection is* $\frac{1}{5}$. *Find the probability that (i) both of them are selected, (ii) only one of them is selected, (iii) none of them is selected, and (iv) at least one of them is selected.*

Solution

Let A and B be the events that the husband and wife are selected respectively.

$$P(A)=\frac{1}{7},\qquad P(B)=\frac{1}{5}$$

Events A and B are independent.

(i) Probability that both of them are selected

$$P(A\cap B)=P(A)\,P(B)$$
$$=\frac{1}{7}\times\frac{1}{5}$$
$$=\frac{1}{35}$$

(ii) Probability that at least one of them is selected

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

$$= \frac{1}{7} + \frac{1}{5} - \frac{1}{35}$$

$$= \frac{11}{35}$$

(iii) Probability that none of them is selected

$$P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B)$$

$$= 1 - \frac{11}{35}$$

$$= \frac{24}{35}$$

(iv) Probability that only one of them is selected

$$P\left[A \cap \bar{B}) \cup (\bar{A} \cap B)\right] = P(A \cup B) - P(A \cap B)$$

$$= \frac{11}{35} - \frac{1}{35}$$

$$= \frac{10}{35}$$

$$= \frac{2}{7}$$

Example 19

There are two bags. The first contains 2 red and 1 white ball, whereas the second bag has only 1 red and 2 white balls. One ball is taken out at random from the first bag and put in the second. Then a ball is chosen at random from the second bag. What is the probability that this last ball is red?

Solution

There are two mutually exclusive cases.

Case I: A red ball is transferred from the first bag to the second bag and a red ball is drawn from it.

Case II: A white ball is transferred from the first bag to the second bag and then a red ball is drawn from it.

Let A be the event of transferring a red ball from the first bag, and B be the event of transferring a white ball from the first bag.

$$P(A) = \frac{2}{3}$$

$$P(B) = \frac{1}{3}$$

Let E be the event of drawing a red ball from the second bag.

$$P(E/A) = \frac{2}{4}$$

$$P(E/B) = \frac{1}{4}$$

$$\begin{aligned}
P(\text{Case I}) &= P(A \cap E) \\
&= P(A)\ P(E/A) \\
&= \frac{2}{3} \times \frac{2}{4} \\
&= \frac{1}{3}
\end{aligned}$$

$$\begin{aligned}
P(\text{Case II}) &= P(B \cap E) \\
&= P(B)\ P(E/B) \\
&= \frac{1}{3} \times \frac{1}{4} \\
&= \frac{1}{12}
\end{aligned}$$

$$\begin{aligned}
P\big[(A \cap E) \cup (B \cap E)\big] &= P(A \cap E) + P(B \cap E) \\
&= \frac{1}{3} + \frac{1}{12} \\
&= \frac{5}{12}
\end{aligned}$$

Example 20

An urn contains four tickets marked with numbers 112, 121, 211, *and* 222, *and one ticket is drawn. Let* A_i $(i = 1, 2, 3)$ *be the event that the* i^{th} *digit of the ticket drawn is* 1. *Show that the events* A_1, A_2, A_3 *are pairwise independent but not mutually independent.*

Solution

$$A_1 = \{112, 121\},\ \ A_2 = \{112, 211\},\ \ A_3 = \{121, 211\}$$

$$A_1 \cap A_2 = \{112\},\ \ A_1 \cap A_3 = \{121\},\ \ A_2 \cap A_3 = \{211\}$$

$$P(A_1) = \frac{2}{4} = \frac{1}{2} = P(A_2) = P(A_3)$$

$$P(A_1 \cap A_2) = \frac{1}{4} = P(A_1 \cap A_3) = P(A_2 \cap A_3)$$

$$P(A_1 \cap A_2) = P(A_1)\,P(A_2) = \frac{1}{4}$$

$$P(A_2 \cap A_3) = P(A_2)\,P(A_3) = \frac{1}{4}$$

$$P(A_1 \cap A_3) = P(A_1)\,P(A_3) = \frac{1}{4}$$

Hence, events A_1, A_2, and A_3 are pairwise independent.

$$P(A_1 \cap A_2 \cap A_3) = P(\phi) = 0$$

$$P(A_1 \cap A_2 \cap A_3) \neq P(A_1)\,P(A_2)\,P(A_3)$$

Hence, events A_1, A_2, and A_3 are not mutually independent.

EXERCISE 1.3

1. Find the probability of drawing 2 red balls in succession from a bag containing 4 red and 5 black balls when the ball that is drawn first is (i) not replaced, and (ii) replaced.

$\left[\textbf{Ans.:}\text{ (i) } \frac{1}{6} \quad \text{(ii) } \frac{16}{81}\right]$

2. Two aeroplanes bomb a target in succession. The probability of each correctly scoring a hit is 0.3 and 0.2 respectively. The second will bomb only if the first misses the target. Find the probability that (i) the target is hit, and (ii) both fail to score hits.

[**Ans.:** (i) 0.44 (ii) 0.56]

3. Box *A* contains 5 red and 3 white marbles and Box *B* contains 2 red and 6 white marbles. If a marble is drawn from each box, what is the probability that they are both of the same colour?

[**Ans.:** 0.109]

4. Two marbles are drawn in succession from a box containing 10 red, 30 white, 20 blue, and 15 orange marbles, with replacement being made after each draw. Find the probability that (i) both are white, and (ii) the first is red and the second is white.

$\left[\textbf{Ans.:}\text{ (i) } \frac{4}{25} \quad \text{(ii) } \frac{4}{75}\right]$

5. *A*, *B*, *C* are aiming to shoot a balloon. *A* will succeed 4 times out of 5 attempts. The chance of *B* to shoot the balloon is 3 out of 4, and that of C is 2 out of 3. If the three aim the balloon simultaneously, find the probability that at least two of them hit the balloon.

$$\left[\textbf{Ans.: } \frac{5}{6}\right]$$

6. There are 12 cards numbered 1 to 12 in a box. If two cards are selected, what is the probability that the sum is odd (i) with replacement, and (ii) without replacement?

$$\left[\textbf{Ans.: } \text{(i) } \frac{1}{2} \quad \text{(ii) } \frac{6}{11}\right]$$

7. 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, and (ii) not replaced.

$$\left[\textbf{Ans.: } \text{(i) } \frac{1}{169} \quad \text{(ii) } \frac{1}{221}\right]$$

8. A can hit a target 2 times in 5 shots; B, 3 times in 4 shots; and C, 2 times in 3 shots. They fire a volley. What is the probability that at least 2 shots hit the target?

$$\left[\textbf{Ans.: } \frac{2}{3}\right]$$

9. There are two bags. The first bag contains 5 red and 7 white balls and the second bag contains 3 red and 12 white balls. One ball is taken out at random from the first bag and is put in the second bag. Now, a ball is drawn from the second bag. What is the probability that this last ball is red?

$$\left[\textbf{Ans.: } \frac{41}{192}\right]$$

10. In a shooting competition, the probability of A hitting the target is $\frac{1}{2}$; of B, is $\frac{2}{3}$; and of C, is $\frac{3}{4}$. If all of them fire at the target, find the probability that (i) none of them hits the target, and (ii) at least one of them hits the target.

$$\left[\textbf{Ans.: } \text{(i) } \frac{1}{24} \quad \text{(ii) } \frac{23}{24}\right]$$

11. The odds against a student X solving a statistics problem are 12 to 10 and the odds in favour of a student Y solving the problem are 6 to 9.

What is the probability that the problem will be solved when both try independently of each other?

$$\left[\textbf{Ans.: } \frac{37}{55}\right]$$

12. A bag contains 6 white and 9 black balls. Four balls are drawn at random twice. Find the probability that the first draw will give 4 white balls and the second draw will give 4 black balls if (i) the balls are replaced, and (ii) the balls are not replaced before the second draw.

$$\left[\textbf{Ans.: } \text{(i) } \frac{6}{5915} \quad \text{(ii) } \frac{3}{715}\right]$$

13. An urn contains 10 white and 3 black balls. Another urn contains 3 white and 5 black balls. Two balls are transferred from the first urn to the second urn and then one ball is drawn from the latter. What is the probability that the ball drawn is white?

$$\left[\textbf{Ans.: } \frac{5}{26}\right]$$

14. A man wants to marry a girl having the following qualities: fair complexion—the probability of getting such a girl is $\frac{1}{20}$, handsome dowry—the probability is $\frac{1}{50}$, westernized manners and etiquettes—the probability of this is $\frac{1}{100}$. Find the probability of his getting married to such a girl when the possessions of these three attributes are independent.

$$\left[\textbf{Ans.: } \frac{1}{100000}\right]$$

15. A small town has one fire engine and one ambulance available for emergencies. The probability that the fire engine is available when needed is 0.98 and the probability that the ambulance is available when called is 0.92. In the event of an injury resulting from a burning building, find the probability that both the fire engine and ambulance will be available.

$$[\textbf{Ans.: } 0.9016]$$

16. In a certain community, 36% of the families own a dog and 22% of the families that own a dog also own a cat. In addition, 30% of the families own a cat. What is the probability that (i) a randomly selected family owns both a dog and a cat, and (ii) a randomly selected family owns a dog given that it owns a cat?

[**Ans.:** (i) 0.0792 (ii) 0.264]

1.5 BAYES' THEOREM

Let $A_1, A_2, ..., A_n$ be n mutually exclusive and exhaustive events with $P(A_i) \neq 0$ for $i = 1, 2, ..., n$ in a sample space S. Let B be an event that can occur in combination with any one of the events $A_1, A_2, ..., A_n$ with $P(B) \neq 0$. The probability of the event A_i when the event B has actually occurred is given by

$$P(A_i/B) = \frac{P(A_i)\,P(B/A_i)}{\sum_{i=1}^{n} P(A_i)\,P(B/A_i)}$$

Proof Since $A_1, A_2, ..., A_n$ are n mutually exclusive and exhaustive events of the sample space S,

$$S = A_1 \cup A_2 \cup ... \cup A_n$$

Since B is another event that can occur in combination with any of the mutually exclusive and exhaustive events $A_1, A_2, ..., A_n$,

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)$$

Taking probability of both the sides,

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

The events $(A_1 \cap B), (A_2 \cap B),$ etc., are mutually exclusive.

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i)\,P(B/A_i)$$

The conditional probability of an event A given that B has already occurred is given by

$$\begin{aligned} P(A_i/B) &= \frac{P(A_i \cap B)}{P(B)} \\ &= \frac{P(A_i)\,P(B/A_i)}{P(B)} \\ &= \frac{P(A_i)\,P(B/A_i)}{\sum_{i=1}^{n} P(A_i)\,P(B/A_i)} \end{aligned}$$

Example 1

A company has two plants to manufacture hydraulic machines. Plant I manufactures 70% of the hydraulic machines, and Plant II manufactures 30%. At Plant I, 80% of hydraulic machines are rated standard quality; and at Plant II, 90% of hydraulic machines are rated standard quality. A machine is picked up at random and is found to be of standard quality. What is the chance that it has come from Plant I? **[Summer 2015]**

Solution

Let A_1 and A_2 be the events that the hydraulic machines are manufactured in Plant I and Plant II respectively. Let B be the event that the machine picked up is found to be of standard quality.

$$P(A_1) = \frac{70}{100} = 0.7$$

$$P(A_2) = \frac{30}{100} = 0.3$$

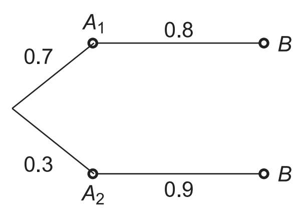


Fig. 1.3

Probability that the machine is of standard quality given that it is manufactured in Plant I

$$P(B/A_1) = \frac{80}{100} = 0.8$$

Probability that the machine is of standard quality given that it is manufactured in Plant II

$$P(B/A_2) = \frac{90}{100} = 0.9$$

Probability that a machine is manufactured in Plant I given that it is of standard quality

$$P(A_1/B) = \frac{P(A_1)\,P(B/A_1)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2)}$$

$$= \frac{0.7 \times 0.8}{0.7 \times 0.8 + 0.3 \times 0.9}$$

$$= 0.6747$$

Example 2

A bag A contains 2 white and 3 red balls, and a bag B contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that the red ball is drawn from the bag B.

Solution

Let A_1 and A_2 be the events that the ball is drawn from bags A and B respectively. Let B be the event that the ball drawn is red.

$$P(A_1) = \frac{1}{2}$$

$$P(A_2) = \frac{1}{2}$$

Probability that the ball drawn is red given that it is drawn from the bag A

$$P(B/A_1) = \frac{3}{5}$$

Probability that the ball drawn is red given that it is drawn from the bag B

$$P(B/A_2) = \frac{5}{9}$$

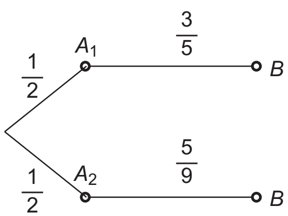


Fig. 1.4

Probability that the ball is drawn from the bag B given that it is red

$$P(A_2/B) = \frac{P(A_2)\,P(B/A_2)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2)}$$

$$= \frac{\frac{1}{2} \times \frac{5}{9}}{\left(\frac{1}{2} \times \frac{3}{5}\right) + \left(\frac{1}{2} \times \frac{5}{9}\right)}$$

$$= \frac{25}{52}$$

Example 3

The chances that Doctor A will diagnose a disease X correctly is 60%. The chances that a patient will die by his treatment after correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. A patient of Doctor A, who had the disease X, died. What is the chance that his disease was diagnosed correctly?

Solution

Let A_1 be the event that the disease X is diagnosed correctly by Doctor A. Let A_2 be the event that the disease X is not diagnosed correctly by Doctor A. Let B be the event that a patient of Doctor A who has the disease X, dies.

$$P(A_1) = \frac{60}{100} = 0.6$$

$$P(A_2) = P(\overline{A}_1) = 1 - P(A_1) = 0.4$$

Probability that the patient of Doctor A who has the disease X dies given that the disease X is diagnosed correctly

$$P(B/A_1) = \frac{40}{100} = 0.4$$

Probability that the patient of Doctor A who has the disease X dies given that the disease X is not diagnosed correctly

$$P(B/A_2) = \frac{70}{100} = 0.7$$

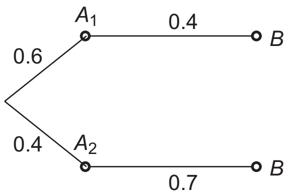


Fig. 1.5

Probability that the disease X is diagnosed correctly given that a patient of Doctor A who has the disease X dies

$$P(A_1/B) = \frac{P(A_1)\,P(B/A_1)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2)}$$

$$= \frac{0.6 \times 0.4}{(0.6 \times 0.4) + (0.4 \times 0.7)}$$

$$= \frac{6}{13}$$

Example 4

In a bolt factory, machines A, B, C manufacture 25%, 35%, and 40% of the total output and out of the total manufacturing, 5%, 4%, and 2% are defective bolts. A bolt is drawn at random from the product and is found to be defective. Find the probabilities that it is manufactured from (i) Machine A, (ii) Machine B, and (iii) Machine C.

Solution

Let A_1, A_2 and A_3 be the events that bolts are manufactured by machines A, B, and C respectively. Let B be the event that the bolt drawn is defective.

$$P(A_1) = \frac{25}{100} = 0.25$$

$$P(A_2) = \frac{35}{100} = 0.35$$

$$P(A_3) = \frac{40}{100} = 0.4$$

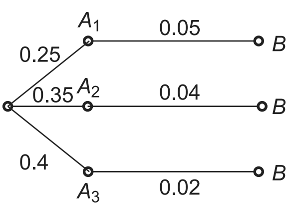


Fig. 1.6

Probability that the bolt drawn is defective given that it is manufactured from Machine A

$$P(B/A_1) = \frac{5}{100} = 0.05$$

Probability that the bolt drawn is defective given that it is manufactured from Machine B

$$P(B/A_2) = \frac{4}{100} = 0.04$$

Probability that the bolt drawn is defective given that it is manufactured from Machine C

$$P(B/A_3) = \frac{2}{100} = 0.02$$

(i) Probability that a bolt is manufactured from Machine A given that it is defective

$$P(A_1/B) = \frac{P(A_1)\,P(B/A_1)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3)}$$

$$= \frac{0.25 \times 0.05}{(0.25 \times 0.05) + (0.35 \times 0.04) + (0.4 \times 0.02)}$$

$$= 0.3623$$

(ii) Probability that a bolt is manufactured from Machine B given that it is defective

$$P(A_2/B) = \frac{P(A_2)\,P(B/A_2)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3)}$$

$$= \frac{0.35 \times 0.04}{(0.25 \times 0.05) + (0.35 \times 0.04) + (0.4 \times 0.02)}$$

$$= 0.4058$$

(iii) Probability that a bolt is manufactured from Machine C given that it is defective

$$P(A_3/B) = \frac{P(A_3)\,P(B/A_3)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3)}$$

$$= \frac{0.4 \times 0.02}{(0.25 \times 0.05) + (0.35 \times 0.04) + (0.4 \times 0.02)}$$

$$= 0.2319$$

Example 5

A businessman goes to hotels X, Y, Z for 20%, 50%, 30% *of the time respectively. It is known that* 5%, 4%, 8% *of the rooms in X, Y, Z hotels have faulty plumbings. What is the probability that the businessman's room having faulty plumbing is assigned to Hotel Z?*

Solution

Let A_1, A_2 and A_3 be the events that the businessman goes to hotels X, Y, Z respectively. Let B be the event that the rooms have faulty plumbings.

$$P(A_1) = \frac{20}{100} = 0.2$$

$$P(A_2) = \frac{50}{100} = 0.5$$

$$P(A_3) = \frac{30}{100} = 0.3$$

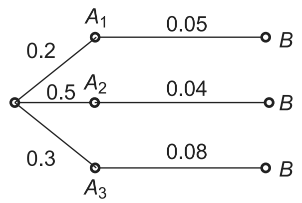


Fig. 1.7

Probability that rooms have faulty plumbings given that rooms belong to Hotel X

$$P(B/A_1) = \frac{5}{100} = 0.05$$

Probability that rooms have faulty plumbing given that rooms belong to Hotel Y

$$P(B/A_2) = \frac{4}{100} = 0.04$$

Probability that rooms have faulty plumbings given that rooms belong to Hotel Z

$$P(B/A_3) = \frac{8}{100} = 0.08$$

Probability that the businessman's room belongs to Hotel Z given that the room has faulty plumbing

$$P(A_3/B) = \frac{P(A_3)\,P(B/A_3)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3)}$$

$$= \frac{0.3 \times 0.08}{(0.2 \times 0.05) + (0.5 \times 0.04) + (0.3 \times 0.08)}$$

$$= \frac{4}{9}$$

Example 6

Of three persons the chances that a politician, a businessman, or an academician would be appointed the Vice Chancellor (VC) of a university are 0.5, 0.3, 0.2 *respectively. Probabilities that research is promoted by these persons if they are appointed as VC are* 0.3, 0.7, 0.8 *respectively.*

(i) Determine the probability that research is promoted.

(ii) If research is promoted, what is the probability that the VC is an academician?

Solution

Let A_1, A_2 and A_3 be the events that a politician, a businessman or an academician will be appointed as the VC respectively. Let B be the event that research is promoted by these persons if they are appointed as VC.

$$P(A_1) = 0.5$$
$$P(A_2) = 0.3$$
$$P(A_3) = 0.2$$

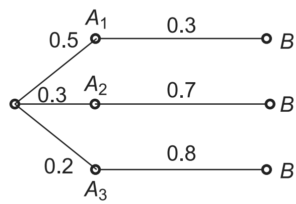


Fig. 1.8

Probability that research is promoted given that a politician is appointed as VC

$$P(B/A_1) = 0.3$$

Probability that research is promoted given that a businessman is promoted as VC

$$P(B/A_2) = 0.7$$

Probability that research is promoted given that an academician is appointed as VC

$$P(B/A_3) = 0.8$$

(i) Probability that research is promoted

$$\begin{aligned} P(B) &= P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3) \\ &= (0.5 \times 0.3) + (0.3 \times 0.7) + (0.2 \times 0.8) \\ &= 0.52 \end{aligned}$$

(ii) Probability that the VC is an academician given that research is promoted by him

$$\begin{aligned} P(A_3/B) &= \frac{P(A_3)\,P(B/A_3)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3)} \\ &= \frac{0.2 \times 0.8}{0.52} \\ &= \frac{4}{13} \end{aligned}$$

Example 7

The contents of urns I, II, and III are as follows:

1 *white,* 2 *red, and* 3 *black balls,*

2 *white,* 3 *red, and* 1 *black ball, and*

3 *white,* 1 *red, and* 2 *black balls.*

One urn is chosen at random and two balls are drawn. They happen to be white and red. Find the probability that they came from (i) Urn I, (ii) Urn II, and (iii) Urn III.

Solution

Let A_1, A_2, and A_3 be the events that urns I, II and III are chosen respectively. Let B be the event that 2 balls drawn are white and red.

$$P(A_1) = \frac{1}{3}$$

$$P(A_2) = \frac{1}{3}$$

$$P(A_3) = \frac{1}{3}$$

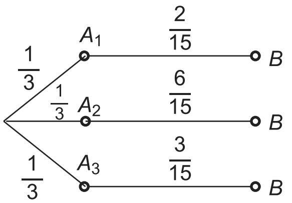


Fig. 1.9

Probability that 2 balls drawn are white and red given that they are chosen from the urn I

$$P(B/A_1) = \frac{{}^1C_1 \times {}^2C_1}{{}^6C_2} = \frac{1\times 2}{15} = \frac{2}{15}$$

Probability that 2 balls drawn are white and red given that they are chosen from the urn II

$$P(B/A_2) = \frac{{}^2C_1 \times {}^3C_1}{{}^6C_2} = \frac{2\times 3}{15} = \frac{6}{15}$$

Probability that 2 balls drawn are white and red given that they are chosen from the urn III

$$P(B/A_3) = \frac{{}^3C_1 \times {}^1C_1}{{}^6C_2} = \frac{3\times 1}{15} = \frac{3}{15}$$

(i) Probability that 2 balls came from the urn I given that they are white and red

$$P(A_1/B) = \frac{P(A_1)\,P(B/A_1)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3)}$$

$$= \frac{\frac{1}{3}\times\frac{2}{15}}{\left(\frac{1}{3}\times\frac{2}{15}\right)+\left(\frac{1}{3}\times\frac{6}{15}\right)+\left(\frac{1}{3}\times\frac{3}{15}\right)}$$

$$= \frac{2}{11}$$

(ii) Probability that 2 balls came from the urn II given that they are white and red

$$P(A_2/B) = \frac{P(A_2)\,P(B/A_2)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3)}$$

$$= \frac{\frac{1}{3}\times\frac{6}{15}}{\left(\frac{1}{3}\times\frac{2}{15}\right)+\left(\frac{1}{3}\times\frac{6}{15}\right)+\left(\frac{1}{3}\times\frac{3}{15}\right)}$$

$$= \frac{6}{11}$$

(iii) Probability that 2 balls came from the urn III given that they are white and red

$$P(A_3/B) = \frac{P(A_3)\,P(B/A_3)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + P(A_3)\,P(B/A_3)}$$

$$= \frac{\frac{1}{3} \times \frac{3}{15}}{\left(\frac{1}{3} \times \frac{2}{15}\right) + \left(\frac{1}{3} \times \frac{6}{15}\right) + \left(\frac{1}{3} \times \frac{3}{15}\right)}$$

$$= \frac{3}{11}$$

EXERCISE 1.4

1. There are 4 boys and 2 girls in Room *A* and 5 boys and 3 girls in Room *B*. A girl from one of the two rooms laughed loudly. What is the probability the girl who laughed was from Room *B*?

$\left[\textbf{Ans.: } \frac{9}{17}\right]$

2. The probability 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*?

$\left[\textbf{Ans.: } \text{(i) } \frac{23}{45} \quad \text{(ii) } \frac{6}{23}\right]$

3. A factory has two machines, *A* and *B*. Past records show that the machine *A* produces 30% of the total output and the machine *B*, the remaining 70%. Machine *A* produces 5% defective articles and Machine *B* produces 1% defective items. An item is drawn at random and found to be defective. What is the probability that it was produced (i) by the machine *A*, and (ii) by the Machine *B*?

[**Ans.:** (i) 0.682 (ii) 0.318]

4. A company has two plants to manufacture scooters. Plant I manufactures 80% of the scooters, and Plant II manufactures 20%. At Plant I, 85 out of 100 scooters are rated standard quality or better. At Plant II, only 65 out of 100 scooters are rated standard quality or better. What is the probability that a scooter selected at random came from (i) Plant I, and (ii) Plant II if it is known that the scooter is of standard quality?

[**Ans.:** (i) 0.84 (ii) 0.16]

5. A new pregnancy test was given to 100 pregnant women and 100 non-pregnant women. The test indicated pregnancy in 92 of the 100 pregnant women and in 12 of the 100 non-pregnant women. If a randomly selected woman takes this test and the test indicates she is pregnant, what is the probability she was not pregnant?

$$\left[\textbf{Ans.: } \frac{3}{26}\right]$$

6. An insurance company insured 2000 scooter drivers, 4000 car drivers, and 6000 truck drivers. The probability of an accident is 0.01, 0.03, and 0.15 in the respective category. One of the insured drivers meets with an accident. What is the probability that he is a scooter driver?

$$\left[\textbf{Ans.: } \frac{1}{52}\right]$$

7. Consider a population of consumers consisting of two types. The upper-income class of consumers comprise 35% of the population and each member has a probability of 0.8 of purchasing Brand *A* of a product. Each member of the rest of the population has a probability of 0.3 of purchasing Brand *A* of the product. A consumer, chosen at random, is found to be the buyer of Brand *A*. What is the probability that the buyer belongs to the middle-income and lower-income classes of consumers?

$$\left[\textbf{Ans.: } \frac{39}{95}\right]$$

8. There are two boxes of identical appearance, each containing 4 spark plugs. It is known that the box I contains only one defective spark plug, while all the four spark plugs of the box II are non-defective. A spark plug drawn at random from a box, selected at random, is found to be non-defective. What is the probability that it came from the box I?

$$\left[\textbf{Ans.: } \frac{3}{7}\right]$$

9. Vijay has 5 one-rupee coins and one of them is known to have two heads. He takes out a coin at random and tosses it 5 times—it always falls head upward. What is the probability that it is a coin with two heads?

$$\left[\textbf{Ans.: } \frac{8}{9}\right]$$

10. Stores *A*, *B*, and *C* have 50, 75, and 100 employees and, respectively 50, 60, 70 per cent of these are women. Resignations are equally likely among all employees, regardless of sex. One employee resigns and this is a woman. What is the probability that she works in Store *C*?

[**Ans.:** 0.5]

Points to Remember

Theorems on Probability

Theorem 1

The probability of an impossible event is zero, i.e., $P(\phi) = 0$, where ϕ is a null set.

Theorem 2

The probability of the complementary event $\bar{A}$ of A is

$$P(\bar{A}) = 1 - P(A)$$

De Morgan's Laws

$$P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$$

$$P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$$

Theorem 3

For any two events A and B in a sample space S,

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Theorem 4 Additive Law of Probability (Addition Theorem)

The probability that at least one of the events A and B will occur is given by

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

Corollary 1 From the Venn diagram,

$$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$$

Corollary 2 $P(\text{Exactly one of } A \text{ and } B \text{ occurs}) = P(A \cup B) - P(A \cap B)$

Corollary 3 If A, B, and C are three events of a sample space S then the probability of occurrence of at least one of them is given by

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

Alternately,

$$P(A \cup B \cup C) = 1 - P(\overline{A} \cap \overline{B} \cap \overline{C})$$

If A, B, and C are mutually exclusive events,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

Corollary 4 The probability of occurrence of at least two of the three events is given by

$$P\left[A \cap B) \cup (B \cap C) \cup (A \cap C)\right] = P(A \cap B) + P(B \cap C) + P(A \cap C) - 2P(A \cap B \cap C)$$

Corollary 5 The probability of occurrence of exactly two of the three events is given by

$$P\left[A \cap B \cap \overline{C}) \cup (A \cap \overline{B} \cap C) \cup (\overline{A} \cap B \cap C)\right]$$

$$= P(A \cap B) + P(B \cap C) + P(A \cap C) - 3P(A \cap B \cap C)$$

Corollary 6 The probability of occurrence of exactly one of the three events is given by

$$P\left[(A \cap \overline{B} \cap \overline{C}) \cup (\overline{A} \cap B \cap \overline{C}) \cup (\overline{A} \cap \overline{B} \cap C)\right]$$

$$= P(A) + P(B) + P(C) - 2P(A \cap B) - 2P(B \cap C) - 2P(A \cap C) + 3P(A \cap B \cap C)$$

Theorem 5 Multiplicative Law or Compound Law of Probability

(a) Conditional Probability Theorem

$$P(A \cap B) = P(A)\ P(B/A)$$

$$P(A \cap B) = P(B)\ P(A/B)$$

(b) Multiplicative Theorem for Independent Events

$$P(A \cap B) = P(A)\ P(B)$$

Corollary 1 If A, B and C are three events then

$$P(A \cap B \cap C) = P(A)\ P(B/A)\ P\left[C/(A \cap B)\right]$$

If A, B and C are independent events,

$$P(A \cap B \cap C) = P(A)\ P(B)\ P(C)$$

Corollary 2 If A and B are independent events then A and $\overline{B}$, $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$ are also independent.

Corollary 3 The probability of occurrence of at least one of the events A, B, C is given by

$$P(A \cup B \cup C) = 1 - P(\overline{A} \cap \overline{B} \cap \overline{C})$$

If A, B, and C are independent events, their complements will also be independent.

$$P(A \cup B \cup C) = 1 - P(\overline{A})\, P(\overline{B})\, P(\overline{C})$$

Pairwise Independence and Mutual Independence The events A, B and C are mutually independent if the following conditions are satisfied simultaneously:

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

$$P(B \cap C) = P(B)\, P(C)$$

$$P(A \cap C) = P(A)\, P(C)$$

and $$P(A \cap B \cap C) = P(A)\, P(B)\, P(C)$$

If the last condition is not satisfied, the events are said to be pairwise independent.

Bayes' Theorem

The probability of the event A_i when the event B has actually occurred is given by

$$P(A_i/B) = \frac{P(A_i)\, P(B/A_i)}{\sum_{i=1}^{n} P(A_i)\, P(B/A_i)}$$

CHAPTER 2

Random Variables and Probability Distributions

Chapter Outline

2.1 INTRODUCTION

The outcomes of random experiments are, in general, abstract quantities or, in other words, most of the time they are not in any numerical form. However, the outcomes of a random experiment can be expressed in quantitative terms, in particular, by means of real numbers. Hence, a function can be defined that takes a definite real value corresponding to each outcome of an experiment. This gives a rationale for the concept of random variables about which probability statements can be made.

In probability and statistics, a probability distribution assigns a probability to each measurable subset of the possible outcomes of a random experiment. Important and commonly encountered probability distributions include binomial distribution, Poisson distribution, and normal distribution.

2.2 RANDOM VARIABLES

A random variable X is a real-valued function of the elements of the sample space of a random experiment. In other words, a variable which takes the real values, depending on the outcome of a random experiment is called a *random variable*, e.g.,

(i) When a fair coin is tossed, $S = \{H, T\}$. If X is the random variable denoting the number of heads,
$X(H) = 1$ and $X(T) = 0$
Hence, the random variable X can take values 0 and 1.

(ii) When two fair coins are tossed, $S = \{HH, HT, TH, TT\}$. If X is the random variable denoting the number of heads,
$X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0.$
Hence, the random variable X can take values 0, 1, and 2.

(iii) When a fair die is tossed, $S = \{1, 2, 3, 4, 5, 6\}$.
If X is the random variable denoting the square of the number obtained,
$X(1) = 1, X(2) = 4, X(3) = 9, X(4) = 16, X(5) = 25, X(6) = 36$
Hence, the random variable X can take values 1, 4, 9, 16, 25, and 36.

Types of Random Variables

There are two types of random variables:

(i) Discrete random variables
(ii) Continuous random variables

Discrete Random Variables A random variable X is said to be discrete if it takes either finite or countably infinite values. Thus, a discrete random variable takes only isolated values, e.g.,

(i) Number of children in a family
(ii) Number of cars sold by different companies in a year
(iii) Number of days of rainfall in a city
(iv) Number of stars in the sky
(v) Profit made by an investor in a day

Continuous Random Variables A random variable X is said to be continuous if it takes any values in a given interval. Thus, a continuous random variable takes uncountably infinite values, e.g.,

(i) Height of a person in cm
(ii) Weight of a bag in kg
(iii) Temperature of a city in degree Celsius

(iv) Life of an electric bulb in hours
(v) Volume of a gas in cc.

Example 1

Identify the random variables as either discrete or continuous in each of the following cases:

(i) A page in a book can have at most 300 *words*
X = Number of misprints on a page
(ii) Number of students present in a class of 50 *students*
(iii) A player goes to the gymnasium regularly
X = Reduction in his weight in a month
(iv) Number of attempts required by a candidate to clear the IAS examination
(v) Height of a skyscraper

Solution

(i) X = Number of misprints on a page
The page may have no misprint or 1 misprint or 2 misprint … or 300 misprints. Thus, X takes values 0, 1, 2, …, 300. Hence, X is a discrete random variable.
(ii) Let X be the random variable denoting the number of students present in a class. X takes values 0, 1, 2, …, 50. Hence, X is a discrete random variable.
(iii) Reduction in weight cannot take isolated values 0, 1, 2, etc., but it takes any continuous value.
Hence, X is a continuous random variable.
(iv) Let X be a random variable denoting the number of attempts required by a candidate. Thus, X takes values 1, 2, 3, …. Hence, X is a discrete random variable.
(v) Since height can have any fractional value, it is a continuous random variable.

2.3 DISCRETE PROBABILITY DISTRIBUTION

Probability distribution of a random variable is the set of its possible values together with their respective probabilities. Let X be a discrete random variable which takes the values $x_1, x_2, \ldots x_n$. The probability of each possible outcome x_i is $p_i = p(x_i) = P(X = x_i)$ for $i = 1, 2, \ldots, n$. The number $p(x_i)$, $i = 1,2, \ldots$. must satisfy the following conditions:

(i) $p(x_i) \geq 0$ for all values of i

(ii) $\sum_{i=1}^{\infty} p(x_i) = 1$

The function $p(x_i)$ is called the probability function or probability mass function or probability density function of the random variable X. The set of pairs $\{x, p(x_i)\}$, $i = 1, 2, \ldots, n$ is called the probability distribution of the random variable which can be displayed in the form of a table as shown below:

$X = x_i$	x_1	x_2	x_3	$\ldots x_i$	$\ldots x_n$
$p(x_i) = P(X = x_i)$	$p(x_1)$	$p(x_2)$	$p(x_3)$	$\ldots p(x_i)$	$\ldots p(x_n)$

2.4 DISCRETE DISTRIBUTION FUNCTION

Let X be a discrete random variable which takes the values $x_1, x_2, \ldots$ such that $x_1 < x_2 < \ldots$ with probabilities $p(x_1), p(x_2) \ldots$ such that $p(x_i) \geq 0$ for all values of i and $\sum_{i=1}^{x} p(x_i) = 1.$

The distribution function $F(x)$ of the discrete random variable X is defined by

$$F(x) = P(X \leq x) = \sum_{i=1}^{x} p(x_i)$$

where x is any integer. The function $F(x)$ is also called the cumulative distribution function. The set of pairs $\{x_i, F(x)\}$, $i = 1, 2, \ldots$ is called the cumulative probability distribution.

X	x_1	x_2	...
$F(x)$	$p(x_1)$	$p(x_1) + p(x_2)$	...

Example 1

A fair die is tossed once. If the random variable is getting an even number, find the probability distribution of X.

Solution

When a fair die is tossed,

$$S = \{1, 2, 3, 4, 5, 6\}$$

Let X be the random variable of getting an even number. Hence, X can take the values 0 and 1.

$$P(X = 0) = P(1, 3, 5) = \frac{3}{6} = \frac{1}{2}$$

$$P(X = 1) = P(2, 4, 6) = \frac{3}{6} = \frac{1}{2}$$

Hence, the probability distribution of X is

$X = x$	0	1
$P(X = 1)$	$\frac{1}{2}$	$\frac{1}{2}$

Also, $$\sum P(X = x) = \frac{1}{2} + \frac{1}{2} = 1$$

Example 2

Find the probability distribution of the number of heads when three coins are tossed.

Solution

When three coins are tossed,

$$S = \{\text{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}\}$$

Let X be the random variable of getting heads in tossing of three coins. Hence X can take the values 0, 1, 2, 3.

$$P(X = 0) = P(\text{no head}) = P(\text{TTT}) = \frac{1}{8}$$

$$P(X = 1) = P(\text{one head}) = P(\text{HTT, THT, TTH}) = \frac{3}{8}$$

$$P(X = 2) = P(\text{two heads}) = P(\text{HHT, THH, HTH}) = \frac{3}{8}$$

$$P(X = 3) = P(\text{three heads}) = P(\text{HHH}) = \frac{1}{8}$$

Hence, the probability distribution of X is

$X = x$	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Also, $$\sum P(X = x) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

Example 3

State with reasons whether the following represent the probability mass function of a random variable:

(i)

$X = x$	0	1	2	3
$P(X = x)$	0.4	0.3	0.2	0.1

(ii)

$X = x$	0	1	2	3
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{4}$

(iii)

$X = x$	0	1	2	3
$P(X = x)$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{4}$

Solution

(i) Here, $0 \le P(X = x) \le 1$ is satisfied for all values of X.

$$\begin{aligned}\sum P(X = x) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\\ &= 0.4 + 0.3 + 0.2 + 0.1\\ &= 1\end{aligned}$$

Since $\sum P(X = x) = 1$, it represents probability mass function.

(ii) Here, $0 \le P(X = x) \le 1$ is satisfied for all values of X.

$$\begin{aligned}\sum P(X = x) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\\ &= \frac{1}{2} + \frac{1}{3} + \frac{1}{6} + \frac{1}{4}\\ &= \frac{5}{4} > 1\end{aligned}$$

Since $\sum(P(X = x) > 1$, it does not represent a probability mass function.

(iii) Here, $0 \le P(X = x) \le 1$ is not satisfied for all the values of X as $P(X = 0) = -\frac{1}{2}$.

Hence, $P(X = x)$ does not represent a probability mass function.

Example 4

Verify whether the following functions can be regarded as the probability mass function for the given values of X:

(i) $P(X = x) = \dfrac{1}{5}$ *for* $x = 0, 1, 2, 3, 4$

$= 0$ *for* *otherwise*

(ii) $P(X = x) = \dfrac{x-2}{5}$ *for* $x = 1, 2, 3, 4, 5$

$= 0$ *for* *otherwise*

(iii) $P(X = x) = \dfrac{x^2}{30}$ *for* $x = 0, 1, 2, 3, 4$

$= 0$ *for* *otherwise*

Solution

(i) $P(X = 0) = P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = \dfrac{1}{5}$

$P(X = x) \geq 0$ for all values of x

$$\sum P(X = x) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5}$$

$$= 1$$

Hence, $P(X = x)$ is a probability mass function.

(ii) $P(X = 1) = \dfrac{1-2}{5} = -\dfrac{1}{5} < 0$

Hence, $P(X = x)$ is not a probability mass function.

(iii) $P(X = 0) = 0$

$P(X = 1) = \dfrac{1}{30}$

$P(X = 2) = \dfrac{4}{30}$

$P(X = 3) = \dfrac{9}{30}$

$P(X = 4) = \dfrac{16}{30}$

$P(X = x) \geq 0$ for all values of x

$$\sum P(X = x) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$
$$= 0 + \frac{1}{30} + \frac{4}{30} + \frac{9}{30} + \frac{16}{30}$$
$$= 1$$

Hence, $P(X = x)$ is a probability mass function.

Example 5

A random variable X has the probability mass function given by

X	1	2	3	4
$P(X = x)$	0.1	0.2	0.5	0.2

Find (i) $P(2 \leq x < 4)$, *(ii)* $P(X > 2)$, *(iii)* $P(X$ *is odd), and (iv)* $P(X$ *is even).*

Solution

(i) $P(2 \leq X < 4) = P(X = 2) + P(X = 3)$
$= 0.2 + 0.5$
$= 0.7$

(ii) $P(X > 2) = P(X = 3) + P(X = 4)$
$= 0.5 + 0.2$
$= 0.7$

(iii) $P(X \text{ is odd}) = P(X = 1) + P(X = 3)$
$= 0.1 + 0.5$
$= 0.6$

(iv) $P(X \text{ is even}) = P(X = 2) + P(X = 4)$
$= 0.2 + 0.2$
$= 0.4$

Example 6

If the random variable X takes the value 1, 2, 3, *and* 4 *such that* $2P(X = 1) = 3P(X = 2) = P(X = 3) = 5P(X = 4)$. *Find the probability distribution.*

Solution

Let $2P(X = 1) = 3P(X = 2) = P(X = 3) = 5P(X = 4) = k$

$$P(X = 1) = \frac{k}{2}$$

$$P(X = 2) = \frac{k}{3}$$

$$P(X=3)=k$$
$$P(X=4)=\frac{k}{5}$$

Since $\sum(P(X=x)=1$,

$$\frac{k}{2}+\frac{k}{3}+k+\frac{k}{5}=1$$
$$k=\frac{30}{61}$$

Hence, the probability distribution is

X	1	2	3	4
$P(X=x)$	$\frac{15}{61}$	$\frac{10}{61}$	$\frac{30}{61}$	$\frac{6}{61}$

Example 7

A random variable X has the following probability distribution:

X	0	1	2	3	4	5	6	7
$P(X=x)$	a	$4a$	$3a$	$7a$	$8a$	$10a$	$6a$	$9a$

(i) Find the value of a.
(ii) Find $P(X<3)$.
(iii) Find the smallest value of m for which $P(X \leq m) \geq 0.6$.

Solution

(i) Since $P(X=x)$ is a probability distribution function,

$$\sum(P(X=x)=1$$
$$P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)+P(X=6)+P(X=7)=1$$
$$a+4a+3a+7a+8a+10a+6a+9a=1$$
$$a=\frac{1}{48}$$

(ii) $$P(X<3)=P(X=0)+P(X=1)+P(X=2)$$
$$=a+4a+3a$$
$$=8a$$
$$=8\left(\frac{1}{48}\right)$$
$$=\frac{1}{6}$$

(iii) $P(X \le 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$$= a + 4a + 3a + 7a + 8a$$
$$= 23a$$
$$= 23\left(\frac{1}{48}\right)$$
$$= 0.575$$

$$P(X \le 5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$
$$= a + 4a + 3a + 7a + 8a + 10a$$
$$= 33a$$
$$= 33\left(\frac{1}{48}\right)$$
$$= 0.69$$

Hence, the smallest value of m for which $P(X \le m) \ge 0.6$ is 5.

Example 8

The probability mass function of a random variable X is zero except at the points X = 0, 1, 2. At these points, it has the values $P(X = 0) = 3c^3$, $P(X = 1) = 4c - 10c^2$, $P(X = 2) = 5c - 1$. *Find (i) c, (ii)* $P(X < 1)$, *(iii)* $P(1 < X \le 2)$, *and (iv)* $P(0 < X \le 2)$.

Solution

(i) Since $P(X = x)$ is a probability mass function,

$$\sum (P(X = x) = 1$$
$$P(X=0) + P(X=1) + P(X=2) = 1$$
$$3c^3 + 4c - 10c^2 + 5c - 1 = 1$$
$$3c^3 - 10c^2 + 9c - 2 = 0$$
$$(3c-1)(c-2)(c-1) = 0$$
$$c = \frac{1}{3},\quad 2,\quad 1$$

But $c < 1$, otherwise given probabilities will be greater than one or less than zero.

$$\therefore \qquad c = \frac{1}{3}$$

Hence, the probability distribution is

X	0	1	2
$P(X = x)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{3}$

(ii) $P(X < 1) = P(X = 0) = \dfrac{1}{9}$

(iii) $P(1 < X \leq 2) = P(X = 2) = \dfrac{2}{3}$

(iv) $P(0 < X \leq 2) = P(X = 1) + P(X = 2)$

$$= \frac{2}{9} + \frac{2}{3}$$

$$= \frac{8}{9}$$

Example 9

From a lot of 10 items containing 3 defectives, a sample of 4 items is drawn at random. Let the random variable X denote the number of defective items in the sample. Find the probability distribution of X.

Solution

The random variable X can take the value 0, 1, 2, or 3.

Total number of items = 10

Number of good items = 7

Number of defective items = 3

$$P(X = 0) = P(\text{no defective}) = \frac{{}^{7}C_4}{{}^{10}C_4} = \frac{1}{6}$$

$$P(X = 1) = P(\text{one defective and three good items}) = \frac{{}^{3}C_1\ {}^{7}C_3}{{}^{10}C_4} = \frac{1}{2}$$

$$P(X = 2) = P(\text{two defectives and two good items}) = \frac{{}^{3}C_2\ {}^{7}C_2}{{}^{10}C_4} = \frac{3}{10}$$

$$P(X = 3) = P(\text{three defectives and one good item}) = \frac{{}^{3}C_3\ {}^{7}C_1}{{}^{10}C_4} = \frac{1}{30}$$

Hence, the probability distribution of the random variable is

X	0	1	2	3
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

Example 10

Construct the distribution function of the discrete random variable X whose probability distribution is as given below:

X	1	2	3	4	5	6	7
$P(X = x)$	0.1	0.15	0.25	0.2	0.15	0.1	0.05

Solution

Distribution function of X

X	$P(X = x)$	$F(x)$
1	0.1	0.1
2	0.15	0.25
3	0.25	0.5
4	0.2	0.7
5	0.15	0.85
6	0.1	0.95
7	0.05	1

Example 11

A random variable X has the probability function given below:

X	0	1	2
$P(X = x)$	k	$2k$	$3k$

Find (i) k, (ii) $P(X < 2)$, $P(X \leq 2)$, $P(0 < X < 2)$, *and (iii) the distribution function.*

Solution:

(i) Since $P(X = x)$ is a probability density function,

$$\sum(P(X = x) = 1$$
$$k + 2k + 3k = 1$$
$$6k = 1$$
$$k = \frac{1}{6}$$

Hence, the probability distribution is

X	0	1	2
$P(X = x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

(ii) $P(X < 2) = P(X = 0) + P(X = 1) = \frac{1}{6} + \frac{2}{6} = \frac{1}{2}$

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} = 1$$

$$P(0 < X < 2) = P(X = 1) = \frac{1}{3}$$

(iii) Distribution function

X	$P(X = x)$	$F(x)$
0	$\frac{1}{6}$	$\frac{1}{6}$
1	$\frac{2}{6}$	$\frac{1}{2}$
2	$\frac{3}{6}$	1

Example 12

A random variable X takes the values –3, –2, –1, 0, 1, 2, 3, such that
$P(X = 0) = P(X > 0) = P(X < 0)$,
$P(X = -3) = P(X = -2) = P(X = -1) = P(X = 1) = P(X = 2) = P(X = 3)$.
Obtain the probability distribution and the distribution function of X.

Solution

Let $P(X = 0) = P(X > 0) = P(X < 0) = k_1$

Since $\sum P(X = x) = 1$

$$k_1 + k_1 + k_1 = 1$$

$$\therefore \quad k_1 = \frac{1}{3}$$

$$P(X = 0) = P(X > 0) = P(X < 0) = \frac{1}{3}$$

Let $\quad P(X = 1) = P(X = 2) = P(X = 3) = k_2$

$$P(X > 0) = P(X = 1) + P(X = 2) + P(X = 3)$$

$$\frac{1}{3} = k_2 + k_2 + k_2$$

$$\therefore \quad k_2 = \frac{1}{9}$$

$$P(X = 1) = P(X = 2) = P(X = 3) = \frac{1}{9}$$

Similarly, $P(X = -3) = P(X = -2) = P(X = -1) = \dfrac{1}{9}$

Probability distribution and distribution function

X	$P(X = x)$	$F(x)$
–3	$\frac{1}{9}$	$\frac{1}{9}$
–2	$\frac{1}{9}$	$\frac{2}{9}$
–1	$\frac{1}{9}$	$\frac{3}{9}$
0	$\frac{1}{3}$	$\frac{6}{9}$
1	$\frac{1}{9}$	$\frac{7}{9}$
2	$\frac{1}{9}$	$\frac{8}{9}$
3	$\frac{1}{9}$	1

Example 13

A discrete random variable X has the following distribution function:

$$F(x) = \begin{cases} 0 & x < 1 \\ \dfrac{1}{3} & 1 \le x < 4 \\ \dfrac{1}{2} & 4 \le x < 6 \\ \dfrac{5}{6} & 6 \le x < 10 \\ 1 & x \ge 10 \end{cases}$$

Find (i) $P(2 < X \le 6)$, *(ii)* $P(X = 5)$, *(iii)* $P(X = 4)$, *(iv)* $P(X \le 6)$, *and (v)* $P(X = 6)$.

Solution

(i) $P(2 < X \leq 6) = F(6) - F(2) = \dfrac{5}{6} - \dfrac{1}{3} = \dfrac{3}{6} = \dfrac{1}{2}$

(ii) $P(X = 5) = P(X \leq 5) - P(X < 5) = F(5) - P(X < 5) = \dfrac{1}{2} - \dfrac{1}{2} = 0$

(iii) $P(X = 4) = P(X \leq 4) - P(X < 4) = F(4) - P(X < 4) = \dfrac{1}{2} - \dfrac{1}{3} = \dfrac{1}{6}$

(iv) $P(X \leq 6) = F(6) = \dfrac{5}{6}$

(v) $P(X = 6) = P(X \leq 6) - P(X < 6) = F(6) - P(X < 6) = \dfrac{5}{6} - \dfrac{1}{2} = \dfrac{1}{3}$

EXERCISE 2.1

1. Verify whether the following functions can be considered as probability mass functions:

(i) $P(X = x) = \dfrac{x^2 + 1}{18}$, $x = 0, 1, 2, 3$ **[Ans.:** Yes)

(ii) $P(X = x) = \dfrac{x^2 - 2}{8}$, $x = 1, 2, 3$ **[Ans.:** No]

(iii) $P(X = x) = \dfrac{2x + 1}{18}$, $x = 0, 1, 2, 3$ **[Ans.:** No]

2. The probability density function of a random variable X is

X	0	1	2	3	4	5	6
$P(X = x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

Find $P(X < 4)$ and $P(3 < X \leq 6)$.

$\left[\textbf{Ans.:} \dfrac{16}{49}, \dfrac{33}{49}\right]$

3. A random variable X has the following probability distribution:

X	1	2	3	4	5	6	7
$P(X = x)$	k	$2k$	$3k$	k^2	$k^2 + k$	$2k^2$	$4k^2$

Find (i) k, (ii) $P(X < 5)$, (iii) $P(X > 5)$, and (iv) $P(0 \leq X \leq 5)$

$\left[\textbf{Ans.:} \dfrac{1}{8} \text{ (ii) } \dfrac{49}{64} \text{ (iii) } \dfrac{3}{32} \text{ (iv) } \dfrac{29}{32}\right]$

4. A discrete random variable X has the following probability distribution:

X	–2	–1	0	1	2	3
$P(X = x)$	0.1	k	0.2	$2k$	0.3	$3k$

Find (i) k, (ii) $P(X \geq 2)$, and (iii) $P(-2 < X < 2)$.

$$\left[\textbf{Ans.: } \frac{1}{15} \text{ (ii) } \frac{1}{2} \text{ (iii) } \frac{2}{5}\right]$$

5. Given the following probability function of a discrete random variable X:

X	0	1	2	3	4	5	6	7
$P(X = x)$	0	c	$2c$	$2c$	$3c$	c^2	$2c^2$	$7c^2 + c$

Find (i) c, (ii) $P(X \geq 6)$, (iii) $P(X < 6)$, and (iv) Find k if $P(X \leq k) > \frac{1}{2}$, where k is a positive integer.

[**Ans.:** (i) 0.1 (ii) 0.19 (iii) 0.81 (iv) 4]

6. A random variable X assumes four values with probabilities $\frac{1+3x}{4}, \frac{1-x}{4}, \frac{1+2x}{4}$ and $\frac{1-4x}{4}$. For what value of x do these values represent the probability distribution of X?

$$\left[\textbf{Ans.: } -\frac{1}{3} \leq X \leq \frac{1}{4}\right]$$

7. Let X denote the number of heads in a single toss of 4 fair coins.

Determine (i) $P(X < 2)$, and (ii) $P(1 < X \leq 3)$.

$$\left[\textbf{Ans.: } \text{(i) } \frac{5}{16} \text{ (ii) } \frac{5}{8}\right]$$

8. If 3 cars are selected from a lot of 6 cars containing 2 defective cars, find the probability distribution of the number of defective cars.

Ans.:

X	0	1	2
$P(X = x)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{2}{5}$

9. Five defective bolts are accidentally mixed with 20 good ones. Find the probability distribution of the number of defective bolts, if four bolts are drawn at random from this lot.

Ans.:

X	0	1	2	3	4
$P(X = x)$	$\frac{969}{2530}$	$\frac{1140}{2530}$	$\frac{380}{2530}$	$\frac{40}{2530}$	$\frac{1}{2530}$

10. Two dice are rolled at once. Find the probability distribution of the sum of the numbers on them.

Ans.:

X	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

11. A random variable X takes three values 0, 1, and 2 with probabilities $\frac{1}{3}, \frac{1}{6}$, and $\frac{1}{2}$ respectively. Obtain the distribution function of X.

$\left[\textbf{Ans.:}\ F(0) = \frac{1}{3}, F(1) = \frac{1}{2}, F(2) = 1\right]$

12. A random variable X has the following probability function:

x	0	1	2	3	4
$P(X = x)$	k	$3k$	$5k$	$7k$	$9k$

Find (i) the value of k, (ii) $P(X < 3)$, $P(X \geq 3)$, $P(0 < X < 4)$, and (iii) distribution function of X.

$\left[\begin{array}{l}\textbf{Ans.:}\ \text{(i)}\ \frac{1}{25}, \text{(ii)} \frac{9}{25}, \frac{16}{25}, \frac{3}{5} \\ \qquad \text{(iii)}\ F(0) = \frac{1}{25}, F(1) = \frac{4}{25}. F(2) = \frac{9}{25}, F(3) = \frac{16}{25}, F(4) = 1\end{array}\right]$

13. A random variable X has the probability function

X	–2	–1	0	1	2	3
$P(X = x)$	0.1	k	0.2	$2k$	0.3	k

Find (i) k, (ii) $P(X \leq 1)$, (iii) $P(-2 < X < 1)$, and (iv) obtain the distribution function of X.

[**Ans.:** (i) 0.1 (ii) 0.6 (iii) 0.3]

14. The following is the distribution function $F(x)$ of a discrete random variable X:

X	–3	–2	–1	0	1	2	3
$P(X = x)$	0.08	0.2	0.4	0.65	0.8	0.9	1

Find (i) the probability distribution of X, (ii) $P(-2 \le X \le 1)$, and (iii) $P(X \ge 1)$.

Ans.: (i)

X	–3	–2	–1	0	1	2	3
$P(X = x)$	0.08	0.12	0.2	0.25	0.15	0.1	0.1

(ii) 0.72 (ii) 0.35

2.5 MEASURES OF CENTRAL TENDENCY FOR DISCRETE PROBABILITY DISTRIBUTION

The behaviour of a random variable is completely characterized by the distribution function $F(x)$ or density function $p(x)$. Instead of a function, a more compact description can be made by a single numbers such as mean, median, mode, variance, and standard deviation known as measures of central tendency of the random variable X.

1. Mean The mean or average value (μ) of the probability distribution of a discrete random variable X is called as expectation and is denoted by $E(X)$.

$$\mu = E(X) = \sum_{i=1}^{\infty} x_i \; p(x_i) = \sum x \; p(x)$$

where $p(x)$ is the probability density function of the discrete random variable X. Expectation of any function $\phi(x)$ of a random variable X is given by

$$E\left[\phi(x)\right] = \sum_{i=1}^{\infty} \phi(x_i)\, p(x_i) = \sum \phi(x)\, p(x)$$

Some important results on expectation:

(i) $E(X + k) = E(X) + k$

(ii) $E(aX \pm b) = aE(X) \pm b$

(iii) $E(X + Y) = E(X) + E(Y)$ provided $E(X)$ and $E(Y)$ exists.

(iv) $E(XY) = E(X)\, E(Y)$ if X and Y are two independent random variables.

2. Variance Variance characterizes the variability in the distributions since two distributions with same mean can still have different dispersion of data about their means. Variance of the probability distribution of a discrete random variable X is given by

$$\begin{aligned}
\text{Var}(X) = \sigma^2 &= E(X - \mu)^2 \\
&= E(X^2 - 2X\mu + \mu^2) \\
&= E(X^2) - E(2X\mu) + E(\mu^2) \\
&= E(X^2) - 2\mu\, E(X) + \mu^2 && [\because E(\text{constant}) = (\text{constant})] \\
&= E(X^2) - 2\mu\mu + \mu^2
\end{aligned}$$

$$= E(X^2) - \mu^2$$
$$= E(X^2) - [E(X)]^2$$

Some important results on variance:

(i) $\text{Var}(k) = 0$

(ii) $\text{Var}(kX) = k^2 \text{ Var}(X)$

(iii) $\text{Var}(X + k) = \text{Var}(X)$

(iv) $\text{Var}(aX + b) = a^2 \text{ Var}(X)$

3. Standard Deviation Standard deviation is the positive square root of the variance.

$$\text{SD} = \sigma = \sqrt{\sum_{i=1}^{\infty} x_i^2 \; p(x_i) - \mu^2}$$
$$= \sqrt{E(X^2) - \mu^2}$$
$$= \sqrt{E(X^2) - [E(X)]^2}$$

Example 1

A random variable X has the following distribution:

X	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Find (i) mean, (ii) variance, and (iii) $P(1 < X < 6)$.

Solution

(i) $$\text{Mean} = \mu = \sum xp(x)$$
$$= 1\left(\frac{1}{36}\right) + 2\left(\frac{3}{36}\right) + 3\left(\frac{5}{36}\right) + 4\left(\frac{7}{36}\right) + 5\left(\frac{9}{36}\right) + 6\left(\frac{11}{36}\right)$$
$$= \frac{161}{36}$$
$$= 4.47$$

(ii) $$\text{Variance} = \sigma^2 = \sum x^2 p(x) - \mu^2$$
$$= 1\left(\frac{1}{36}\right) + 4\left(\frac{3}{36}\right) + 9\left(\frac{5}{36}\right) + 16\left(\frac{7}{36}\right) + 25\left(\frac{9}{36}\right) + 36\left(\frac{11}{36}\right) - (4.47)^2$$

$$= \frac{791}{36} - 19.98$$
$$= 1.99$$

(iii) $P(1 < X < 6) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$

$$= \frac{3}{36} + \frac{5}{36} + \frac{7}{36} + \frac{9}{36}$$
$$= \frac{24}{36}$$
$$= 0.67$$

Example 2

The probability distribution of a random variable X is given below. Find (i) $E(X)$, (ii) $\text{Var}(X)$, *(iii) $E(2X - 3)$, and (iv)* $\text{Var}\,(2X - 3)$

X	–2	–1	0	1	2
$P(X = x)$	0.2	0.1	0.3	0.3	0.1

Solution

(i) $E(X) = \Sigma x\, p(x)$

$= -2(0.2) - 1(0.1) + 0 + (0.3) + 2(0.1)$

$= 0$

(ii) $\text{Var}(X) = \Sigma x^2\, p(x) - [E(X)]^2$

$= 4(0.2) + 1(0.1) + 0 + 1(0.3) + 4(0.1) - 0$

$= 1.6$

(iii) $E(2X - 3) = 2E(X) - 3$

$= 2(0) - 3$

$= -3$

(iv) $\text{Var}\,(2X - 3) = (2)^2\, \text{Var}\,(X)$

$= 4(1.6)$

$= 6.4$

Example 3

Mean and standard deviation of a random variable X are 5 *and* 4 *respectively. Find $E(X^2)$ and standard deviation of* $(5 - 3X)$.

Solution

$$E(X) = \mu = 5$$
$$\text{SD} = \sigma = 4$$
$$\therefore \quad \text{Var}(X) = \sigma^2 = 16$$

$$
\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - [E(X)]^2 \\
16 &= E(X^2) - (5)^2 \\
\therefore \quad E(X^2) &= 41 \\
\mathrm{Var}\,(5 - 3X) &= \mathrm{Var}\,(5) - (-3)^2\, \mathrm{Var}\,(X) \\
&= 0 + 9(16) \\
&= 144 \\
\mathrm{SD}\,(5 - 3X) &= \sqrt{\mathrm{Var}\,(5 - 3X)} \\
&= \sqrt{144} \\
&= 12
\end{aligned}
$$

Example 4

A machine produces an average of 500 *items during the first week of the month and on average of* 400 *items during the last week of the month, the probability for these being* 0.68 *and* 0.32 *respectively. Determine the expected value of the production.*

[Summer 2015]

Solution

Let X be the random variable which denotes the items produced by the machine. The probability distribution is

X	500	400
$P(X = x)$	0.68	0.32

$$
\begin{aligned}
\text{Expected value of the production } E(X) &= \sum x\, p(x) \\
&= 500(0.68) + 400(0.32) \\
&= 468
\end{aligned}
$$

Example 5

The monthly demand for Allwyn watches is known to have the following probability distribution:

Demand (x)	1	2	3	4	5	6	7	8
Probability $p(x)$	0.08	0.12	0.19	0.24	0.16	0.10	0.07	0.04

Find the expected demand for watches. Also, compute the variance.

Solution

$$\begin{aligned}E(X) &= \Sigma\, x\, p(x)\\ &= 1(0.08)+2(0.12)+3(0.19)+4(0.24)+5(0.16)\\ &\quad +6(0.10)+7(0.07)+8(0.04)\\ &= 4.06\end{aligned}$$

$$\begin{aligned}\text{Var}(X) &= E(X^2)-[E(X)]^2\\ &= \Sigma\, x^2\, p(x)-[E(X)]^2\\ &= 1(0.08)+4(0.12)+9(0.19)+16(0.24)+25(0.16)\\ &\quad +36(0.10)+49(0.07)+64(0.04)-(4.06)^2\\ &= 19.7-16.48\\ &= 3.21\end{aligned}$$

Example 6

A discrete random variable has the probability mass function given below:

X	–2	–1	0	1	2	3
$P(X=x)$	0.2	k	0.1	$2k$	0.1	$2k$

Find k, mean, and variance.

Solution

Since $P(X=x)$ is a probability mass function,

$$\sum P(X=x)=1$$

$$\begin{aligned}0.2+k+0.1+2k+0.1+2k &= 1\\ 5k+0.4 &= 1\\ 5k &= 0.6\\ k &= \frac{0.6}{5}=\frac{3}{25}\end{aligned}$$

Hence, the probability distribution is

X	–2	–1	0	1	2	3
$P(X=x)$	$\frac{2}{10}$	$\frac{3}{25}$	$\frac{1}{10}$	$\frac{6}{25}$	$\frac{1}{10}$	$\frac{6}{25}$

Mean = $E(X)=\sum x\, p(x)$

$$= (-2)\left(\frac{2}{10}\right)+(-1)\left(\frac{3}{25}\right)+0+1\left(\frac{6}{25}\right)+2\left(\frac{1}{10}\right)+3\left(\frac{6}{25}\right)$$

$$=\frac{6}{25}$$

$$\text{Variance} = \text{Var}(X) = E(X^2) - [E(X)]^2$$
$$= \sum x^2\, p(x) - [E(X)]^2$$
$$= 4\left(\frac{2}{10}\right) + 1\left(\frac{3}{25}\right) + 0 + 1\left(\frac{6}{25}\right) + 4\left(\frac{1}{10}\right) + 9\left(\frac{6}{25}\right) - \left(\frac{6}{25}\right)^2$$
$$= \frac{73}{250} - \frac{36}{625}$$
$$= \frac{293}{625}$$

Example 7

A random variable X has the following probability function:

x	0	1	2	3	4	5	6	7
$p(x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2 + k$

(i) *Determine k.* (ii) *Evaluate* $P(X < 6)$, $P(X \geq 6)$, $P(0 < X < 5)$ and $P(0 \leq X \leq 4)$. (iii) *Determine the distribution function of X. (iv) Find the mean. (v) Find the variance.*

Solution

(i) Since $p(x)$ is a probability mass function,

$$\sum p(x) = 1$$
$$0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$
$$10k^2 + 9k - 1 = 1$$
$$(10k - 1)(k + 1) = 0$$
$$k = \frac{1}{10} \text{ or } k = -1$$
$$k = \frac{1}{10} = 0.1 \left[\because \quad p(x) \geq 0, k \neq -1\right]$$

Hence, the probability function is

X	0	1	2	3	4	5	6	7
$P(X = x)$	0	0.1	0.2	0.2	0.3	0.01	0.02	0.17

(ii) $$\begin{aligned} P(X<6) &= P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5) \\ &= 0+0.1+0.2+0.2+0.3+0.01 \\ &= 0.81 \\ P(X\ge 6) &= 1-P(X<6) \\ &= 1-0.81 \\ &= 0.19 \\ P(0<X<5) &= P(X=1)+P(X=2)+P(X=3)+P(X=4) \\ &= 0.1+0.2+0.2+0.3 \\ &= 0.8 \\ P(0\le X\le 4) &= P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4) \\ &= 0+0.1+0.2+0.2+0.3 \\ &= 0.8 \end{aligned}$$

(iii) Distribution function of X

x	$p(x)$	$F(x)$
0	0	0
1	0.1	0.1
2	0.2	0.3
3	0.2	0.5
4	0.3	0.8
5	0.01	0.81
6	0.02	0.83
7	0.17	1

(iv) $$\begin{aligned} \mu &= \sum xp(x) \\ &= 0+1(0.1)+2(0.2)+3(0.2)+4(0.3)+5(0.01)+6(0.02)+7(0.17) \\ &= 3.66 \end{aligned}$$

(v) $$\begin{aligned} \mathrm{Var}(X) = \sigma^2 &= \sum x^2\, p(x) - \mu^2 \\ &= 0+1(0.1)+4(0.2)+9(0.2)+16(0.3)+25(0.01)+36(0.02) \\ &\qquad +49(0.17)-(3.66)^2 \\ &= 3.4044 \end{aligned}$$

Example 8

A fair die is tossed. Let the random variable X denote the twice the number appearing on the die. Write the probability distribution of X. Calculate mean and variance.

Solution

Let X be the random variable which denotes twice the number appearing on the die.

(i) Probability distribution of X

x	2	4	6	8	10	12
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(ii) Mean $= \mu = \sum xp(x)$

$$= 2\left(\frac{1}{6}\right)+4\left(\frac{1}{6}\right)+6\left(\frac{1}{6}\right)+8\left(\frac{1}{6}\right)+10\left(\frac{1}{6}\right)+12\left(\frac{1}{6}\right)$$

$$= 7$$

(iii) Variance $= \sigma^2 = \sum x^2 p(x) - \mu^2$

$$= 4\left(\frac{1}{6}\right)+16\left(\frac{1}{6}\right)+36\left(\frac{1}{6}\right)+64\left(\frac{1}{6}\right)+100\left(\frac{1}{6}\right)+144\left(\frac{1}{6}\right)-(7)^2$$

$$= 11.67$$

Example 9

Two unbiased dice are thrown at random. Find the probability distribution of the sum of the numbers on them. Also, find mean and variance.

Solution

Let X be the random variable which denotes the sum of the numbers on two unbiased dice. The random variable X can take values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. The probability distribution is

X	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Mean $= \mu = \sum x\, p(x)$

$$= 2\left(\frac{1}{36}\right)+3\left(\frac{2}{36}\right)+4\left(\frac{3}{36}\right)+5\left(\frac{4}{36}\right)+6\left(\frac{5}{36}\right)+7\left(\frac{6}{36}\right)+8\left(\frac{5}{36}\right)$$

$$+9\left(\frac{4}{36}\right)+10\left(\frac{3}{36}\right)+11\left(\frac{2}{36}\right)+12\left(\frac{1}{36}\right)$$

$$= \frac{252}{36}$$

$$= 7$$

$$\begin{aligned}\text{Variance} = \sigma^2 &= \sum x^2\, p(x) - \mu^2 \\ &= 4\left(\frac{1}{36}\right) + 9\left(\frac{2}{36}\right) + 16\left(\frac{3}{36}\right) + 25\left(\frac{4}{36}\right) + 36\left(\frac{5}{36}\right) \\ &\qquad + 49\left(\frac{6}{36}\right) + 64\left(\frac{5}{36}\right) + 81\left(\frac{4}{36}\right) + 100\left(\frac{3}{36}\right) \\ &\qquad + 121\left(\frac{2}{36}\right) + 144\left(\frac{1}{36}\right) - (7)^2 \\ &= \frac{1974}{36} - 49 \\ &= 5.83\end{aligned}$$

Example 10

A sample of 3 *items is selected at random from a box containing* 10 *items of which* 4 *are defective. Find the expected number of defective items.*

Solution

Let X be the random variable which denotes the defective items.

Total number of items = 10

Number of good items = 6

Number of defective items = 4

$$P(X=0) = P(\text{no defective item}) = \frac{{}^6C_3}{{}^{10}C_3} = \frac{1}{6}$$

$$P(X=1) = P(\text{one defective item}) = \frac{{}^6C_2\ {}^4C_1}{{}^{10}C_3} = \frac{1}{2}$$

$$P(X=2) = P(\text{two defective items}) = \frac{{}^6C_1\ {}^4C_2}{{}^{10}C_3} = \frac{3}{10}$$

$$P(X=3) = P(\text{three defective items}) = \frac{{}^4C_3}{{}^{10}C_3} = \frac{1}{30}$$

Hence, the probability distribution is

X	0	1	2	3
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

Expected number of defective items $= E(X) = \sum x\, p(x)$

$$= 0 + 1\left(\frac{1}{2}\right) + 2\left(\frac{3}{10}\right) + 3\left(\frac{1}{30}\right)$$

$$= 1.2$$

Example 11

A player tosses two fair coins. He wins ₹ 100 if a head appears and ₹ 200 if two heads appear. On the other hand, he loses ₹ 500 if no head appears. Determine the expected value of the game. Is the game favourable to the players?

Solution

Let X be the random variable which denotes the number of heads appearing in tosses of two fair coins.

$$S = \{\text{HH, HT, TH, TT}\}$$

$$p(x_1) = P(X=0) = P(\text{no heads}) = \frac{1}{4}$$

$$p(x_2) = P(X=1) = P(\text{one head}) = \frac{2}{4} = \frac{1}{2}$$

$$p(x_3) = P(X=2) = P(\text{two heads}) = \frac{1}{4}$$

Amount to be lost if no head appears $= x_1 = -$ ₹ 500

Amount to be won if one head appears $= x_2 =$ ₹ 100

Amount to be won if two heads appear $= x_3 =$ ₹ 200

Expected value of the game $= \mu = \sum x\, p(x)$

$$= x_1\, p(x_1) + x_2\, p(x_2) + x_3\, p(x_3)$$

$$= -500\left(\frac{1}{4}\right) + 100\left(\frac{1}{2}\right) + 200\left(\frac{1}{4}\right)$$

$$= ₹\ -25$$

Hence, the game is not favourable to the player.

Example 12

Amit plays a game of tossing a die. If a number less than 3 *appears, he gets ₹ a, otherwise he has to pay ₹* 10. *If the game is fair, find a.*

Solution

Let X be the random variable which denotes tossing of a die.

Probability of getting a number less than 3, i.e., 1 or 2 = $p(x_1) = \frac{2}{6} = \frac{1}{3}$

Probability of getting number more than or equal to 3, i.e., 3, 4, 5, or 6 $= p(x_2) = \frac{4}{6} = \frac{2}{3}$

Amount to be received for number less than 3 = x_1 = ₹ a
Amount to be paid for numbers more than or equal to 3 = x_2 = ₹ –10

$$\begin{aligned} E(X) &= \sum x\, p(x) \\ &= x_1\, p(x_1) + x_2\, p(x_2) \\ &= a\left(\frac{1}{3}\right) + (-10)\left(\frac{2}{3}\right) \\ &= \frac{a}{3} - \frac{20}{3} \end{aligned}$$

For a pair game, $E(x) = 0$.

$$\frac{a}{3} - \frac{20}{3} = 0$$
$$a = 20$$

Example 13

A man draws 2 balls from a bag containing 3 white and 5 black balls. If he is to receive ₹ 14 for every white ball which he draws and ₹ 7 for every black ball, what is his expectation?

Solution

Let X be the random variable which denotes the balls drawn from a bag. 2 balls drawn may be either (i) both white, or (ii) both black, or (iii) one white and one black.

Probability of drawing 2 white balls = $p(x_1) = \frac{^3C_2}{^8C_2} = \frac{3}{28}$

Probability of drawing 2 black balls = $p(x_2) = \frac{^5C_2}{^8C_2} = \frac{10}{28}$

Probability of drawing 1 white and 1 black ball = $p(x_3) = \frac{^3C_1\ ^5C_1}{^8C_2} = \frac{15}{28}$

Amount to be received for 2 white balls = x_1 = ₹ 14 × 2 = ₹ 28
Amount to be received for 2 black balls = x_2 = ₹ 7 × 2 = ₹ 14
Amount to be received for 1 white and 1 black ball = x_3 = ₹ 14 + ₹ 7 = ₹ 21

$$\text{Expectation} = E(X) = \sum x\, p(x)$$
$$= x_1\, p(x_1) + x_2\, p(x_2) + x_3\, p(x_3)$$
$$= 28\left(\frac{3}{28}\right) + 14\left(\frac{10}{28}\right) + 21\left(\frac{15}{28}\right)$$
$$= ₹\ 19.25$$

Example 14

The probability that there is at least one error in an account statement prepared by A is 0.2 and for B and C, they are 0.25 and 0.4 respectively. A, B, and C prepared 10, 16, and 20 statements respectively. Find the expected number of correct statements in all.

Solution

Let $p(x_1)$, $p(x_2)$ and $p(x_3)$ be the probabilities of the events that there is no error in the account statements prepared by A, B, and C respectively.

$$p(x_1) = 1 - (\text{Probability of at least one error in the account statement prepared by } A)$$
$$= 1 - 0.2$$
$$= 0.8$$

Similarly, $p(x_2) = 1 - 0.25 = 0.75$

$p(x_3) = 1 - 0.4 = 0.6$

Also, $x_1 = 10$, $x_2 = 16$, $x_3 = 20$

$$\text{Expected number of correct statements} = E(X) = \sum x\, p(x)$$
$$= x_1\, p(x_1) + x_2\, p(x_2) + x_3\, p(x_3)$$
$$= 10(0.8) + 16(0.75) + 20(0.6)$$
$$= 32$$

Example 15

A man has the choice of running either a hot-snack stall or an ice-cream stall at a seaside resort during the summer season. If it is a fairly cool summer, he should make ₹ 5000 by running the hot-snack stall, but if the summer is quite hot, he can only expect to make ₹ 1000. On the other hand, if he operates the ice-cream stall, his profit is estimated at ₹ 6500, if the summer is hot, but only ₹ 1000 if it is cool. There is a 40 percent chance of the summer being hot. Should he opt for running the hot-snack stall or the ice-cream stall?

Solution

Let X and Y be the random variables which denote the income from the hot-snack and ice-cream stalls respectively.

Probability of hot summer $= p_1 = 40\% = 0.4$

Probability of cool summer $= p_2 = 1 - p_1 = 1 - 0.4 = 0.6$

$$x_1 = 1000, \qquad x_2 = 5000, \qquad y_1 = 6500, \qquad y_2 = 1000$$

$$\begin{aligned}\text{Expected income from hot-snack stall } &= E(X)\\ &= x_1 p_1 + x_2 p_2\\ &= 1000(0.4) + 5000(0.6)\\ &= ₹\ 3400\end{aligned}$$

$$\begin{aligned}\text{Expected income from ice-cream stall } &= E(Y)\\ &= y_1 p_1 + y_2 p_2\\ &= 6500\ (0.4) + 1000(0.6)\\ &= ₹\ 3200\end{aligned}$$

Hence, he should opt for running the hot-snack stall.

EXERCISE 2.2

1. The probability distribution of a random variable X is given by

X	–2	–1	0	1	2	3
$P(X = x)$	0.1	k	0.2	$2k$	0.3	k

Find k, the mean, and variance.

[**Ans.:** 0.1, 0.8, 2.16]

2. Find the mean and variance of the following distribution:

X	4	5	6	8
$P(X = x)$	0.1	0.3	0.4	0.2

[**Ans.:** 5.9, 1.49]

3. Find the value of k from the following data:

X	0	10	15
$P(X = x)$	$\frac{k-6}{5}$	$\frac{2}{k}$	$\frac{14}{5k}$

Also, find the distribution function and expectation of X.

Ans.: 8,

X	0	10	15
$F(X)$	$\frac{2}{5}$	$\frac{13}{20}$	1

, $\frac{31}{4}$

4. For the following distribution,

X	–3	–2	–1	0	1	2
$P(X = x)$	0.01	0.1	0.2	0.3	0.2	0.15

Find (i) $P(X \geq 1)$, (ii) $P(X < 0)$, (iii) $E(X)$, and (iv) Var(X)

[**Ans.:** (i) 0.35 (ii) 0.35 (iii) 0.05 (iv) 1.8475]

5. A random variable X has the following probability function:

X	0	1	2	3	4	5	6	7	8
$P(X = x)$	$\frac{k}{45}$	$\frac{k}{15}$	$\frac{k}{9}$	$\frac{k}{5}$	$\frac{2k}{45}$	$\frac{6k}{45}$	$\frac{7k}{45}$	$\frac{8k}{45}$	$\frac{4k}{45}$

Determine (i) k, (ii) mean, (iii) variance, and (iv) SD.

[**Ans.:** (i) 1 (ii) 0.4622 (iii) 4.9971 (iv) 2.24]

6. A fair coin is tossed until a head or five tails appear. Find (i) discrete probability distribution, and (ii) mean of the distribution.

Ans.: (i)

X	1	2	3	4	5
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$

(ii) 1.9

7. Let X denotes the minimum of two numbers that appear when a pair of fair dice is thrown once. Determine (i) probability distribution, (ii) expectation, and (iii) variance.

Ans.: (i)

X	1	2	3	4	5	6
$P(X = x)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

(ii) 2.5278 (iii) 1.9713

8. For the following probability distribution,

X	–3	–2	–1	0	1	2	3
$P(X = x)$	0.001	0.01	0.1	?	0.1	0.01	0.001

Find (i) missing probability, (ii) mean, and (iii) variance.

[**Ans.:** (i) 0.778 (ii) 0.2 (iii) 0.258]

9. A discrete random variable can take all integer values from 1 to k each with the probability of $\frac{1}{k}$. Show that its mean and variance are $\frac{k+1}{2}$ and $\frac{k^2+1}{2}$ respectively.

10. An urn contains 6 white and 4 black balls; 3 balls are drawn without replacement. What is the expected number of black balls that will be obtained?

$\left[\textbf{Ans.: } \frac{6}{5}\right]$

11. A six-faced die is tossed. If a prime number occurs, Anil wins that number of rupees but if a nonprime number occurs, he loses that number of rupees. Determine whether the game is favourable to the player.

[**Ans.:** The game is favourable to Anil]

12. A man runs an ice-cream parlour at a holiday resort. If the summer is mild, he can sell 2500 cups of ice cream; if it is hot, he can sell 4000 cups; if it is very hot, he can sell 5000 cups. It is known that for any year, the probability of summer to be mild is $\frac{1}{7}$ and to be hot is $\frac{4}{7}$. A cup of ice cream costs ₹ 2 and is sold for ₹ 3.50. What is his expected profit?

[**Ans.:** ₹ 6107.14]

13. A player tosses two fair coins. He wins ₹ 1 or ₹ 2 as 1 tail or 1 head appears. On the other hand, he loses ₹ 5 if no head appears. Find the expected gain or loss of the player.

[**Ans.:** Loss of ₹ 0.25]

14. 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 ₹ 20. Determine their expectations.

$\left[\textbf{Ans.: } ₹\ 8,\ ₹\ 6,\ ₹\ 4,\ ₹\ 2\right]$

2.6 CONTINUOUS PROBABILITY DISTRIBUTION

Let X be a continuous random variable such that the probability of the variable X falling in the small interval $x - \frac{1}{2}\mathrm{d}x$ to $x + \frac{1}{2}\mathrm{d}x$ is $f(x)\,\mathrm{d}x$, i.e.,

$$P\left(x - \frac{1}{2}\mathrm{d}x \le X \le x + \frac{1}{2}\mathrm{d}x\right) = f(x)\,\mathrm{d}x$$

The function $f(x)$ is called the probability density function of the random variable X and the continuous curve $y = f(x)$ is called the probability curve.

Properties of Probability Density Function

(i) $f(x) \ge 0, \qquad -\infty < x < \infty$

(ii) $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$

(iii) $P(a < x < b) = \int_a^b f(x)\,\mathrm{d}x$

2.7 CONTINUOUS DISTRIBUTION FUNCTION

If X is a continuous random variable having the probability density function $f(x)$ then the function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x, \quad -\infty < x < \infty$$

is called the distribution function or cumulative distribution function of the random variable X.

Properties of Distribution Function

(i) $F(-\infty) = 0$

(ii) $F(\infty) = 1$

(iii) $0 \le F(x) \le 1, \quad -\infty < x < \infty$

(iv) $P(a < X < b) = F(b) - F(a)$

(v) $F'(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x), \quad f(x) \ge 0$

Example 1

Show that the function $f(x)$ defined by

$$f(x) = \frac{1}{7} \qquad 1 < x < 8$$
$$= 0 \qquad \textit{otherwise}$$

is a probability density function for a random variable. Hence, find $P(3 < X < 10)$.

Solution

$$f(x) \geq 0 \qquad \text{in} \qquad 1 < x < 8$$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{8} f(x)\,dx + \int_{8}^{\infty} f(x)\,dx$$
$$= 0 + \int_{1}^{8} \frac{1}{7}\,dx + 0$$
$$= \frac{1}{7}\left|x\right|_1^8$$
$$= \frac{1}{7}(8-1)$$
$$= 1$$

Hence, $f(x)$ is a probability density function.

$$P(3 < X < 10) = \int_{3}^{10} f(x)\,dx$$
$$= \int_{3}^{8} f(x)\,dx + \int_{8}^{10} f(x)\,dx$$
$$= \int_{3}^{8} \frac{1}{7}\,dx + 0$$
$$= \frac{1}{7}\left|x\right|_3^8$$
$$= \frac{1}{7}(8-3)$$
$$= \frac{5}{7}$$

Example 2

Is the function $f(x)$ defined by

$$f(x) = e^{-x} \qquad x \geq 0$$
$$= 0 \qquad x < 0$$

is a probability density function. If so, find the probability that the variate having this density falls in the interval (1, 2).

Solution

$$f(x) \geq 0 \quad \text{in} \quad (0, \infty)$$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx$$

$$= 0 + \int_{0}^{\infty} e^{-x}\,dx$$

$$= \left| -e^{-x} \right|_{0}^{\infty}$$

$$= -e^{-\infty} + 1$$

$$= 1$$

Hence, $f(x)$ is a probability density function.

$$P(1 \leq X \leq 2) = \int_{1}^{2} f(x)\,dx$$

$$= \int_{1}^{2} e^{-x}\,dx$$

$$= \left| -e^{-x} \right|_{1}^{2}$$

$$= -e^{-2} + e^{-1}$$

$$= 0.233$$

Example 3

If a random variable has the probability density function $f(x)$ as

$$f(x) = 2e^{-2x} \qquad x > 0$$
$$= 0 \qquad x \leq 0$$

Find the probabilities that it will take on a value (i) between 1 *and* 3, *and (ii) greater than 0.5.*

Solution

(i) Probability that the variable will take a value between 1 and 3

$$P(1 < X < 3) = \int_1^3 f(x)\,dx$$

$$= \int_1^3 2e^{-2x}\,dx$$

$$= 2\left|\frac{e^{-2x}}{-2}\right|_1^3$$

$$= -(e^{-6} - e^{-2})$$

$$= e^{-2} - e^{-6}$$

(ii) Probability that the variable will take a value greater than 0.5

$$P(X > 0.5) = \int_{0.5}^{\infty} f(x)\,dx$$

$$= \int_{0.5}^{\infty} 2e^{-2x}\,dx$$

$$= 2\left|\frac{e^{-2x}}{-2}\right|_{0.5}^{\infty}$$

$$= -(e^{-\infty} - e^{-1})$$

$$= e^{-1}$$

Example 4

Find the constant k such that the function

$$\begin{aligned} f(x) &= kx^2 \qquad 0 < x < 3 \\ &= 0 \qquad \textit{otherwise} \end{aligned}$$

is a probability density function and compute (i) $P(1 < x < 2)$, *(ii)* $P(X < 2)$, *and (iii)* $P(X \geq 2)$.

Solution

Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_{-\infty}^{0} f(x)\,dx + \int_0^3 f(x)\,dx + \int_3^{\infty} f(x)\,dx = 1$$

$$0+\int_0^3 kx^2\,\mathrm{d}x+0=1$$

$$k\left|\frac{x^3}{3}\right|_0^3=1$$

$$\frac{k}{3}(27-0)=1$$

$$9k=1$$

$$k=\frac{1}{9}$$

Hence, $f(x)=\frac{1}{9}x^2 \qquad 0<x<3$

$= 0 \qquad$ otherwise

(i) $P(1<X<2)=\int_1^2 f(x)\,\mathrm{d}x$

$$=\int_1^2 \frac{1}{9}x^2\,\mathrm{d}x$$

$$=\frac{1}{9}\left|\frac{x^3}{3}\right|_1^2$$

$$=\frac{1}{27}(8-1)$$

$$=\frac{7}{27}$$

(ii) $P(X<2)=\int_{-\infty}^2 f(x)\,\mathrm{d}x$

$$=\int_{-\infty}^0 f(x)\,\mathrm{d}x+\int_0^2 f(x)\,\mathrm{d}x$$

$$=0+\int_0^2 \frac{1}{9}x^2\,\mathrm{d}x$$

$$=\frac{1}{9}\int_0^2 x^2\,\mathrm{d}x$$

$$=\frac{1}{9}\left|\frac{x^3}{3}\right|_0^2$$

$$=\frac{1}{27}(8-0)$$

$$=\frac{8}{27}$$

(iii) $P(X \ge 2) = 1 - P(X < 2)$

$$= 1 - \frac{8}{27}$$

$$= \frac{19}{27}$$

Example 5

If the probability density function of a random variable is given by

$$f(x) = k(1 - x^2) \quad 0 < x < 1$$
$$= 0 \quad \text{otherwise}$$

Find the value of k and the probabilities that a random variable having this probability density will take on a value (i) between 0.1 *and* 0.2*, and (ii) greater than* 0.5.

Solution

Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx = 1$$

$$0 + \int_{0}^{1} k(1 - x^2)\,dx + 0 = 1$$

$$k\left| x - \frac{x^3}{3} \right|_0^1 = 1$$

$$k\left(1 - \frac{1}{3}\right) = 1$$

$$k = \frac{3}{2}$$

Hence, $f(x) = \frac{3}{2}(1 - x^2) \quad 0 < x < 1$

$= 0$ otherwise

(i) Probability that the variable will take on a value between 0.1 and 0.2

$$P(0.1 < X < 0.2) = \int_{0.1}^{0.2} f(x)\,dx$$

$$= \int_{0.1}^{0.2} \frac{3}{2}(1 - x^2)\,dx$$

$$= \frac{3}{2}\left| x - \frac{x^3}{3} \right|_{0.1}^{0.2}$$

$$= \frac{3}{2}\left[\left(0.2 - \frac{0.008}{3}\right) - \left(0.1 - \frac{0.001}{3}\right)\right]$$

$$= 0.1465$$

(ii) Probability that the variable will take on a value greater than 0.5

$$P(X > 0.5) = \int_{0.5}^{\infty} f(x)\,\mathrm{d}x$$

$$= \int_{0.5}^{1} f(x)\,\mathrm{d}x + \int_{1}^{\infty} f(x)\,\mathrm{d}x$$

$$= \int_{0.5}^{1} \frac{3}{2}(1 - x^2)\,\mathrm{d}x + 0$$

$$= \frac{3}{2}\left| x - \frac{x^3}{3} \right|_{0.5}^{1}$$

$$= \frac{3}{2}\left[\left(1 - \frac{1}{3}\right) - \left(0.5 - \frac{0.125}{3}\right)\right]$$

$$= 0.3125$$

Example 6

If X is a continuous random variable with pdf

$$f(x) = x^2 \quad 0 \le x \le 1$$

$$= 0 \quad \textit{otherwise}$$

If $P(a \le X \le 1) = \dfrac{19}{81}$, *find the value of a.*

Solution

$$P(a \le X \le 1) = \frac{19}{81}$$

$$\int_a^1 f(x)\,\mathrm{d}x = \frac{19}{81}$$

$$\int_a^1 x^2\,\mathrm{d}x = \frac{19}{81}$$

$$\left|\frac{x^3}{3}\right|_a^1 = \frac{19}{81}$$

$$\frac{1}{3}(1-a) = \frac{19}{81}$$

$$1-a = \frac{19}{27}$$

$$a = \frac{46}{27}$$

Example 7

Let X be a continuous random variable with pdf

$$f(x) = kx\,(1-x),\ 0 \le x \le 1$$

Find k and determine a number b such that $P(X \le b) = P(X \ge b)$.

Solution

Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x) = 1$$

$$\int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_{0}^{1} f(x)\,\mathrm{d}x + \int_{1}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$0 + \int_{0}^{1} kx(1-x)\,\mathrm{d}x + 0 = 1$$

$$k\int_{0}^{1} (x - x^2)\,\mathrm{d}x = 1$$

$$k\left|\frac{x^2}{2} - \frac{x^3}{3}\right|_0^1 = 1$$

$$k\left[\left(\frac{1}{2} - \frac{1}{3}\right) - (0-0)\right] = 1$$

$$k\left(\frac{1}{6}\right) = 1$$

$$k = 6$$

Hence, $f(x) = 6(x - x^2) \qquad 0 \le x \le 1$

Since total probability is 1 and $P(X \le b) = P(X \ge b)$,

$$P(X \le b) = \frac{1}{2}$$

$$\int_0^b f(x)\,\mathrm{d}x = \frac{1}{2}$$

$$6\int_0^b (x - x^2)\,\mathrm{d}x = \frac{1}{2}$$

$$6\left|\frac{x^2}{2} - \frac{x^3}{3}\right|_0^b = \frac{1}{2}$$

$$\frac{b^2}{2} - \frac{b^3}{3} = \frac{1}{12}$$

$$6b^2 - 4b^3 = 1$$

$$4b^3 - 6b^2 + 1 = 0$$

$$(2b-1)(2b^2 - 2b - 1) = 0$$

$$b = \frac{1}{2} \text{ or } b = \frac{1 \pm \sqrt{3}}{2}$$

b lies in (0, 1).

$\therefore \quad b = \frac{1}{2}$

Example 8

The length of time (in minutes) that a certain lady speaks on the telephone is found to be a random phenomenon, with a probability function specified by the function

$$f(x) = A\,e^{-\frac{x}{5}} \qquad x \ge 0$$
$$= 0 \qquad \textit{otherwise}$$

(i) Find the value of A that makes $f(x)$ a probability density function.
(ii) What is the probability that the number of minutes that she will take over the phone is more than 10 minutes?

Solution

(i) For $f(x)$ to be a probability density function,

$$\int_{-\infty}^{\infty} f(x)\,x = 1$$

$$\int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_{0}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$0+\int_0^\infty A e^{-\frac{x}{5}}\mathrm{d}x = 1$$

$$A\left|\frac{e^{-\frac{x}{5}}}{-\frac{1}{5}}\right|_0^\infty = 1$$

$$-5A(e^{-\infty}-e^{-0}) = 1$$

$$-5A(0-1) = 1$$

$$5A = 1$$

$$A = \frac{1}{5}$$

Hence, $f(x) = \frac{1}{5}e^{-\frac{x}{5}} \quad x \ge 0$

$= 0 \quad$ otherwise

(ii) $P(X>10) = \int_{10}^\infty f(x)\,\mathrm{d}x$

$$= \int_{10}^\infty \frac{1}{5}e^{-\frac{x}{5}}\,\mathrm{d}x$$

$$= \frac{1}{5}\left|\frac{e^{-\frac{x}{5}}}{-\frac{1}{5}}\right|_{10}^\infty$$

$$= -\left(e^{-\infty}-e^{-2}\right)$$

$$= -\left(0-e^{-2}\right)$$

$$= e^{\frac{1}{2}}$$

Example 9

A continuous random variable X has a pdf $f(x)^2 = 3x^2$, $0 \le x \le 1$. *Find a and b such that*

(i) $P(X \le a) = P(X > a)$ and

(ii) $P(X > b) = 0.05$

Solution

Since total probability is 1 and $P(X \le a) = P(X > a)$,

$$P(X \le a) = \frac{1}{2}$$

$$\int_0^a f(x)\,\mathrm{d}x = \frac{1}{2}$$

$$\int_0^a 3x^2\mathrm{d}x = \frac{1}{2}$$

$$3\left|\frac{x^3}{3}\right|_0^a = \frac{1}{2}$$

$$a^3 = \frac{1}{2}$$

$$a = \left(\frac{1}{2}\right)^{\frac{1}{3}}$$

$$P(X > b) = 0.05$$

$$\int_b^1 f(x)\,\mathrm{d}x = 0.05$$

$$\int_b^1 3x^2\,\mathrm{d}x = 0.05$$

$$3\left|\frac{x^3}{3}\right|_b^1 = 0.05$$

$$1 - b^3 = 0.05$$

$$b^3 = \frac{19}{20}$$

$$b = \left(\frac{19}{20}\right)^{\frac{1}{3}}$$

Example 10

Let the continuous random variable X have the probability density function

$$f(x) = \frac{2}{x^3} \qquad 1 < x < \infty$$
$$= 0 \qquad \textit{otherwise}$$

Find $F(x)$.

Solution

$$F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x$$

$$= \int_{-\infty}^{1} f(x)\,\mathrm{d}x + \int_{1}^{x} f(x)\,\mathrm{d}x$$

$$= 0 + \int_{1}^{x} \frac{2}{x^3}\,\mathrm{d}x$$

$$= 2\left|\frac{x^{-2}}{-2}\right|_1^x$$

$$= -\left|\frac{1}{x^2}\right|_1^x$$

$$= -\left(\frac{1}{x^2} - 1\right)$$

$$= 1 - \frac{1}{x^2}$$

Hence, $F(x) = 1 - \dfrac{1}{x^2} \qquad 1 < x < \infty$

$\qquad = 0 \qquad$ otherwise

Example 11

Verify that the function F(x) is a distribution function.

$$F(x) = 0 \qquad x < 0$$

$$= 1 - e^{-\frac{x}{4}} \qquad x \ge 0$$

Also, find the probabilities $P(X \le 4), P(X \ge 8), P(4 \le X \le 8)$.

Solution

For the function $F(x)$,

(i) $F(-\infty) = 0$

(ii) $F(\infty) = 1 - e^{-\infty} = 1 - 0 = 1$

(iii) $0 \le F(x) \le 1 \qquad -\infty < x < \infty$

If $f(x)$ is the corresponding probability density function,

$$f(x) = F'(x) = 0 \qquad x < 0$$

$$= \frac{1}{4} e^{-\frac{x}{4}} \qquad x \ge 0$$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx$$

$$= 0 + \int_{0}^{\infty} \frac{1}{4} e^{-\frac{x}{4}}\,dx$$

$$= \frac{1}{4}\left|\frac{e^{-\frac{x}{4}}}{-\frac{1}{4}}\right|_{0}^{\infty}$$

$$= -\left|e^{-\frac{x}{4}}\right|_{0}^{\infty}$$

$$= -(0-1)$$

$$= 1$$

Hence, $F(x)$ is a distribution function.

$$P(X \le 4) = F(4)$$

$$= 1 - e^{-1}$$

$$= 1 - \frac{1}{e}$$

$$= \frac{e-1}{e}$$

$$P(X \ge 8) = 1 - P(X \le 8)$$

$$= 1 - F(8)$$

$$= 1 - (1 - e^{-2})$$

$$= e^{-2}$$

$$= \frac{1}{e^2}$$

$$P(4 \le X \le 8) = F(8) - F(4)$$

$$= (1 - e^{-2}) - (1 - e^{-1})$$

$$= e^{-1} - e^{-2}$$

$$= \frac{1}{e} - \frac{1}{e^2}$$

$$= \frac{e-1}{e^2}$$

Example 12

The troubleshooting capacity of an IC chip in a circuit is a random variable X whose distribution function is given by

$$
\begin{aligned}
F(x) &= 0 & x \le 3 \\
&= 1 - \frac{9}{x^2} & x > 3
\end{aligned}
$$

where x denotes the number of years. Find the probability that the IC chip will work properly (i) less than 8 years, (i) beyond 8 years, (iii) between 5 to 7 years, and (iv) anywhere from 2 to 5 years.

Solution

(i) $P(X \le 8) = F(8)$

$$
\begin{aligned}
&= 1 - \frac{9}{8^2} \\
&= 0.8594
\end{aligned}
$$

(ii) $P(X > 8) = 1 - P(X \le 8)$

$$
\begin{aligned}
&= 1 - F(8) \\
&= 1 - 0.8594 \\
&= 0.1406
\end{aligned}
$$

(iii) $P(5 \le X \le 7) = F(7) - F(5)$

$$
\begin{aligned}
&= \left(1 - \frac{9}{7^2}\right) - \left(1 - \frac{9}{5^2}\right) \\
&= 0.1763
\end{aligned}
$$

(iv) $P(2 \le X \le 5) = F(5) - F(2)$

$$
\begin{aligned}
&= \left(1 - \frac{9}{5^2}\right) - 0 \\
&= 0.64
\end{aligned}
$$

Example 13

The probability density function of a continuous random variable X is given by

$$
f(x) = \begin{cases} ax & 0 \le x \le 1 \\ a & 1 \le x \le 2 \\ 3a - ax & 2 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}
$$

(i) Find the value of a, and (ii) find the cdf of X.

Solution

(i) Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_{-\infty}^{\infty} f(x)\,dx + \int_0^1 f(x)\,dx + \int_1^2 f(x)\,dx + \int_2^3 f(x)\,dx = 1$$

$$0 + \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (3a - ax)\,dx = 1$$

$$a\left|\frac{x^2}{2}\right|_0^1 + a\left| x \right|_1^2 + \left|3ax - \frac{ax^2}{2}\right|_2^3 = 1$$

$$a\left(\frac{1}{2} - 0\right) + a(2-1) + \left[\left(9a - \frac{9a}{2}\right) - (6a - 2a)\right] = 1$$

$$\frac{1}{2}a + a + \frac{9a}{2} - 4a = 1$$

$$2a = 1$$

$$a = \frac{1}{2}$$

(ii) $F(x) = \int_{-\infty}^{x} f(x)\,dx$

For $0 \le x \le 1$,

$$F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_0^x f(x)\,dx$$

$$= 0 + \int_0^x ax\,dx$$

$$= a\left|\frac{x^2}{2}\right|_0^x$$

$$= \frac{ax^2}{2}$$

For $1 \le x \le 2$,

$$F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_0^1 f(x)\,dx + \int_1^x f(x)\,dx$$

$$= 0 + \int_0^1 ax\,dx + \int_1^x a\,dx$$

$$= a\left|\frac{x^2}{2}\right|_0^1 + a|x|_1^x$$

$$= a\left(\frac{1}{2}-0\right) + a(x-1)$$

$$= \frac{a}{2} + ax - a$$

$$= ax - \frac{a}{2}$$

For $2 \le x \le 3$,

$$F(x) = \int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_0^1 f(x)\,\mathrm{d}x + \int_1^2 f(x)\,\mathrm{d}x + \int_2^x f(x)\,\mathrm{d}x$$

$$= 0 + \int_0^1 ax\,\mathrm{d}x + \int_1^2 a\,\mathrm{d}x + \int_2^x (3a - ax)\,\mathrm{d}x$$

$$= a\left|\frac{x^2}{2}\right|_0^1 + a|x|_1^2 + \left|3ax - \frac{ax^2}{2}\right|_2^x$$

$$= a\left(\frac{1}{2}-0\right) + a(2-1) + \left[\left(3ax - \frac{ax^2}{2}\right) - (6a - 2a)\right]$$

$$= \frac{a}{2} + a + 3ax - \frac{ax^2}{2} - 4a$$

$$= 3ax - \frac{ax^2}{2} - \frac{5a}{2}$$

Hence,

$$\begin{aligned} F(x) &= \frac{ax^2}{2} && 0 \le x \le 1 \\ &= ax - \frac{a}{2} && 1 \le x \le 2 \\ &= 3ax - \frac{ax^2}{2} - \frac{5a}{2} && 2 \le x \le 3 \end{aligned}$$

Example 14

The pdf of a continuous random variable X is

$$f(x) = \frac{1}{2}e^{-|x|}$$

Find cdf F(x).

Solution

$$f(x) = \frac{1}{2}e^{x} \qquad -\infty < x < 0$$
$$= \frac{1}{2}e^{-x} \qquad 0 < x < \infty$$

$$F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x$$

For $x \le 0$,

$$F(x) = \int_{-\infty}^{x} \frac{1}{2}e^{x}\mathrm{d}x$$
$$= \frac{1}{2}\left|e^{x}\right|_{-\infty}^{x}$$
$$= \frac{1}{2}(e^{x} - e^{-\infty})$$
$$= \frac{1}{2}e^{x}$$

For $x > 0$,

$$F(x) = \int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_{0}^{x} f(x)\,\mathrm{d}x$$
$$= \int_{-\infty}^{0} \frac{1}{2}e^{x}\mathrm{d}x + \int_{0}^{x} \frac{1}{2}e^{-x}\,\mathrm{d}x$$
$$= \frac{1}{2}\left|e^{x}\right|_{-\infty}^{0} + \frac{1}{2}\left|-e^{-x}\right|_{0}^{x}$$
$$= \frac{1}{2}(1 - e^{-\infty}) + \frac{1}{2}(-e^{-x} + e^{0})$$
$$= \frac{1}{2} - \frac{1}{2}e^{-x} + \frac{1}{2}$$
$$= 1 - \frac{1}{2}e^{-x}$$

Hence, $F(x) = \frac{1}{2}e^{x} \qquad x \le 0$

$$= 1 - \frac{1}{2}e^{-x} \qquad x > 0$$

Example 15

Find the value of k and the distribution function F(x) given the probability density function of a random variable X as

$$f(x) = \frac{k}{x^2+1} \qquad -\infty < x < \infty$$

Solution

Since $f(x)$ is the probability density function,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$\int_{-\infty}^{\infty} \frac{k}{x^2+1}\,\mathrm{d}x = 1$$

$$k\int_{-\infty}^{\infty} \frac{1}{x^2+1}\,\mathrm{d}x = 1$$

$$k\left|\tan^{-1} x\right|_{-\infty}^{\infty} = 1$$

$$k\left[\tan^{-1}\infty - \tan^{-1}(-\infty)\right] = 1$$

$$k\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right] = 1$$

$$k\pi = 1$$

$$k = \frac{1}{\pi}$$

Hence, $f(x) = \dfrac{1}{\pi}\dfrac{1}{x^2+1} \qquad -\infty < x < \infty$

$$F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x$$

$$= \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{x^2+1}\,\mathrm{d}x$$

$$= \frac{1}{\pi}\left|\tan^{-1} x\right|_{-\infty}^{x}$$

$$= \frac{1}{\pi}\left[\tan^{-1} x - \tan^{-1}(-\infty)\right]$$

$$= \frac{1}{\pi}\left(\tan^{-1} x + \frac{\pi}{2}\right)$$

Example 16

Find the constant k such that

$$f(x) = kx^2 \quad 0 < x < 3$$
$$= 0 \quad \textit{otherwise}$$

is a probability function. Also, find the distribution function F(x) and $P(1 < X \le 2)$.

Solution

Since $f(x)$ is probability density function,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x + \int_{0}^{3} f(x)\,\mathrm{d}x + \int_{3}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$0 + \int_{0}^{3} kx^2\,\mathrm{d}x + 0 = 1$$

$$k\left|\frac{x^3}{3}\right|_0^3 = 1$$

$$k(9-0) = 1$$

$$k = \frac{1}{9}$$

Hence, $f(x) = \dfrac{1}{9}x^2 \quad 0 < x < 3$

$= 0 \quad$ otherwise

$$F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x$$

$$= \int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_{0}^{x} f(x)\,\mathrm{d}x$$

$$= 0 + \int_0^x \frac{1}{9}x^2\,\mathrm{d}x$$

$$= \frac{1}{9}\left|\frac{x^3}{3}\right|_0^x$$

$$= \frac{1}{27}x^3$$

Hence, $F(x) = \frac{1}{27}x^3 \qquad 0 < x < 3$

$= 0 \qquad$ otherwise

$$P(1 < x \le 2) = \int_1^2 f(x)\,\mathrm{d}x$$

$$= \int_1^2 \frac{1}{9}x^2\,\mathrm{d}x$$

$$= \frac{1}{9}\left|\frac{x^3}{3}\right|_1^2$$

$$= \frac{1}{27}(8-1)$$

$$= \frac{7}{27}$$

EXERCISE 2.3

1. Verify whether the following functions are probability density functions:

(i) $f(x) = k\,e^{-kx} \qquad x \ge 0,\ k > 0$

(ii) $f(x) = \frac{1}{2}\,e^{-|x|} \qquad -\infty < x < \infty$

(iii) $f(x) = \frac{2}{9}x\left(2 - \frac{x}{2}\right) \qquad 0 \le x \le 3$

$\left[\textbf{Ans.:} \text{ (i) Yes (ii) Yes (iii) Yes}\right]$

2. Find the value of k if the following are probability density functions:

(i) $f(x) = k(1+x) \qquad 2 \le x \le 5$

(ii) $f(x) = k(x - x^2) \qquad 0 \le x \le 1$

(iii) $f(x) = kx\,e^{-4x^2} \qquad 0 \le x \le \infty$

(iv) $f(x) = kx\,e^{-\frac{x^2}{4}} \qquad 0 \le x \le \infty$

$\left[\textbf{Ans.:} \text{ (i) } \frac{2}{27} \text{ (ii) } 6 \text{ (iii) } 8 \text{ (iv) } \frac{1}{2}\right]$

3. A function is defined as

$$f(x) = \begin{cases} 0 & x < 2 \\ \dfrac{2x+3}{18} & 2 \le x \le 4 \\ 0 & x > 4 \end{cases}$$

Show that $f(x)$ is a probability density function and find $P(2 < X < 3)$.

$$\left[\textbf{Ans.: } \frac{4}{9}\right]$$

4. Let X be a continuous random variable with probability distribution

$$f(x) = \begin{cases} \dfrac{x}{6} + k & 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

Find k, and $P(1 \le X \le 2)$.

$$\left[\textbf{Ans.: } 1, \frac{1}{3}\right]$$

5. Find the value of k such that $f(x)$ is a probability density function. Find also, $P(X \le 1.5)$.

$$f(x) = \begin{cases} kx & 0 \le x \le 1 \\ k & 1 \le x \le 2 \\ k(3-x) & 2 \le x \le 3 \end{cases}$$

$$\left[\textbf{Ans.: } \frac{1}{2}, \frac{1}{2}\right]$$

6. If X is a continuous random variable whose probability density function is given by

$$\begin{aligned} f(x) &= k(4x - 2x^2) && 0 < x < 2 \\ &= 0 && \text{otherwise} \end{aligned}$$

Find (i) the value of k, and (ii) $P(X > 1)$.

$$\left[\textbf{Ans.: } \text{(i) } \frac{3}{8} \text{ (ii) } \frac{1}{2}\right]$$

7. If a random variable has the probability density function

$$\begin{aligned} f(x) &= k(x^2 - 1) && -1 \le x \le 3 \\ &= 0 && \text{otherwise} \end{aligned}$$

Find (i) the value of k, and (ii) $P\left(\frac{1}{2} \le X \le \frac{5}{2}\right)$.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{3}{28} \text{ (ii) } \frac{19}{56}\right]$$

8. The probability density function is

$$\begin{aligned} f(x) &= k(3x^2 - 1) && -1 \le x \le 2 \\ &= 0 && \text{otherwise} \end{aligned}$$

Find (i) the value of k, and (ii) $P(-1 \le X \le 0)$.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{6} \text{ (ii) } 0\right]$$

9. Is the function defined by

$$\begin{aligned} f(x) &= 0 && x < 2 \\ &= \frac{1}{18}(2x+3) && 2 \le x \le 4 \\ &= 0 && x > 4 \end{aligned}$$

a probability density function? Find the probability that a variate having $f(x)$ as density function will fall in the interval $2 \le X \le 3$.

$$\left[\textbf{Ans.:} \text{ Yes, } \frac{4}{9}\right]$$

10. A random variable X gives measurements x between 0 and 1 with a probability function

$$\begin{aligned} f(x) &= 12x^3 - 21x^2 + 10x && 0 \le x \le 1 \\ &= 0 && \text{otherwise} \end{aligned}$$

(i) Find $P\left(X \le \frac{1}{2}\right)$ and $P\left(X > \frac{1}{2}\right)$.

(ii) Find a number k such that $P(X \le k) = \frac{1}{2}$.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{7}{16} \text{ (ii) } 0.452\right]$$

11. The distribution function of a random variable X is given by

$$F(x) = \begin{cases} 1 - e^{-x^2} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the probability density function.

$$\left[\begin{aligned} \textbf{Ans.:}\ f(x) &= 2xe^{-x^2} && x > 0 \\ &= 0 && \text{otherwise} \end{aligned}\right]$$

12. The cdf of a continuous random variable X is given by

$$F(x)=\begin{cases}0 & x<0\\ x^2 & 0\le x\le 1\\ 1 & x>1\end{cases}$$

Find the pdf and $P\left(\frac{1}{2}\le X\le\frac{4}{5}\right)$.

$\left[\textbf{Ans.:}\ 0.195\right]$

13. Find the distribution function corresponding to the following probability density functions:

(i) $f(x)=\begin{cases}\frac{1}{2}x^2e^{-x} & 0\le x<\infty\\ 0 & \text{otherwise}\end{cases}$

(ii) $f(x)=x \qquad 0\le x\le 1$
$\qquad =2-x \qquad 1\le x\le 2$
$\qquad =0 \qquad \text{otherwise}$

(iii) $f(x)=\lambda(x-1)^4 \qquad 1\le x\le 3,\ \lambda>0$
$\qquad =0 \qquad \text{otherwise}$

$$\left[\begin{array}{l}\textbf{Ans.:}\ \text{(i)}\quad F(x)=\begin{cases}1-e^{-x}\left(1+x+\dfrac{x^2}{2}\right) & x\ge 0\\ 0 & \text{otherwise}\end{cases}\\ \text{(ii)}\quad F(x)=\begin{cases}0 & x<0\\ \dfrac{x^2}{2} & 0\le x\le 1\\ 2x-0.5x^2-1 & 1\le x\le 2\\ 1 & x>2\end{cases}\\ \text{(iii)}\quad \lambda=\dfrac{5}{32},\quad F(x)=\begin{cases}0 & x\le 1\\ \dfrac{5}{32}(x-1)^4 & 1\le x\le 3\\ 1 & x\ge 3\end{cases}\end{array}\right]$$

14. A continuous random variable X has the following probability density function

$$f(x)=\frac{a}{x^5} \qquad 2\le x\le 10$$

Determine the constant a, distribution function of X, and find the probability of the event $4 \le x \le 7$.

$$\left[\textbf{Ans.: } \frac{2500}{39}, F(x) = \frac{625}{39}\left(\frac{1}{16} - \frac{1}{x^4}\right), 0.056\right]$$

2.8 MEASURES OF CENTRAL TENDENCY FOR CONTINUOUS PROBABILITY DISTRIBUTION

1. Mean The mean or average value (μ) of the probability distribution of a continuous random variable X is called the expectation and is denoted by $E(X)$.

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$$

where $f(x)$ is the probability density function of the continuous random variable. Expectation of any function $\phi(x)$ of a continuous random variable X is given by

$$E[\phi(x)] = \int_{-\infty}^{\infty} \phi(x)\, f(x)\, \mathrm{d}x$$

2. Median The median is the point which divides the entire distribution into two equal parts. In case of a continuous distribution, the median is the point which divides the total area into two equal parts. Thus, if a continuous random variable X is defined from a to b and M is the median,

$$\int_a^M f(x)\, \mathrm{d}x = \int_M^b f(x)\, \mathrm{d}x = \frac{1}{2}$$

By solving any one of this equation, the median is obtained.

3. Mode The mode is value of x for which $f(x)$ is maximum. Mode is given by

$$f'(x) = 0 \text{ and } f''(x) < 0 \quad \text{for} \quad a < x < b$$

4. Variance The variance of the probability distribution of a continuous random variable X is given by

$$\begin{aligned} \mathrm{Var}(X) = \sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, \mathrm{d}x \\ &= \int_{-\infty}^{\infty} x^2\, f(x)\, \mathrm{d}x - \mu^2 \end{aligned}$$

5. Standard Deviation The standard deviation of the probability distribution of a continuous random variable X is given by

$$\mathrm{SD} = \sqrt{\mathrm{Var}(X)} = \sigma$$

Example 1

For the continuous random variable having pdf

$$f(x) = 4x^3 \quad 0 \le x \le 1$$
$$= 0 \quad \textit{otherwise}$$

Find the mean and variance of X.

Solution

$$\text{Mean} = \mu = \int_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x$$

$$= \int_{-\infty}^{0} x\, f(x)\,\mathrm{d}x + \int_{0}^{1} x\, f(x)\,\mathrm{d}x + \int_{1}^{\infty} x\, f(x)\,\mathrm{d}x$$

$$= 0 + \int_{0}^{1} x(4x^3)\,\mathrm{d}x + 0$$

$$= 4\int_{0}^{1} x^4\,\mathrm{d}x$$

$$= 4\left|\frac{x^5}{5}\right|_0^1$$

$$= 4\left(\frac{1}{5} - 0\right)$$

$$= \frac{4}{5}$$

$$\text{Var}(X) = \int_{-\infty}^{\infty} x^2\, f(x)\,\mathrm{d}x - \mu^2$$

$$= \int_{-\infty}^{0} x^2 f(x)\,\mathrm{d}x + \int_{0}^{1} x^2 f(x)\,\mathrm{d}x + \int_{1}^{\infty} x^2 f(x)\,\mathrm{d}x - \mu^2$$

$$= 0 + \int_{0}^{1} x^2\,(4x^3)\,\mathrm{d}x + 0 - \left(\frac{4}{5}\right)^2$$

$$= 4\int_{0}^{1} x^5\,\mathrm{d}x - \frac{16}{25}$$

$$= 4\left|\frac{x^6}{6}\right|_0^1 - \frac{16}{25}$$

$$= \frac{4}{6} - \frac{16}{25}$$
$$= \frac{2}{75}$$

Example 2

For the triangular distribution

$$\begin{aligned} f(x) &= x && 0 < x \le 1 \\ &= 2 - x && 1 \le x \le 2 \\ &= 0 && \textit{otherwise} \end{aligned}$$

Find the mean and variance.

Solution

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x \\ &= \int_{-\infty}^{0} x\ f(x)\ \mathrm{d}x + \int_{0}^{1} x\, f(x)\ \mathrm{d}x + \int_{1}^{2} x\, f(x) + \int_{2}^{\infty} x\, f(x)\ \mathrm{d}x \\ &= 0 + \int_{0}^{1} x \cdot x\ \mathrm{d}x + \int_{1}^{2} x(2-x)\ \mathrm{d}x + 0 \\ &= \int_{0}^{1} x^2\ \mathrm{d}x + \int_{1}^{2} (2x - x^2)\ \mathrm{d}x \\ &= \left|\frac{x^3}{3}\right|_0^1 + \left|2\frac{x^2}{2} - \frac{x^3}{3}\right|_1^2 \\ &= \left(\frac{1}{3} - 0\right) + \left[\left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right)\right] \\ &= \frac{1}{3} + \frac{4}{3} - \frac{2}{3} \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{Var}\ (X) &= \int_{-\infty}^{\infty} x^2\ f(x)\, \mathrm{d}x - \mu^2 \\ &= \int_{-\infty}^{0} x^2 f(x)\ \mathrm{d}x + \int_{0}^{1} x^2\ f(x)\ \mathrm{d}x + \int_{1}^{2} x^2\ f(x)\ \mathrm{d}x + \int_{2}^{\infty} x^2 f(x)\ \mathrm{d}x - \mu^2 \end{aligned}$$

$$= 0 + \int_0^1 x^2 \cdot x \, dx + \int_1^2 x^2 (2-x) \, dx + 0 - 1$$

$$= \int_0^1 x^3 \, dx + \int_1^2 (2x^2 - x^3) \, dx - 1$$

$$= \left|\frac{x^4}{4}\right|_0^1 + \left|\frac{2x^3}{3} - \frac{x^4}{4}\right|_1^2 - 1$$

$$= \left(\frac{1}{4} - 0\right) + \left[\left(\frac{16}{3} - \frac{16}{4}\right) - \left(\frac{2}{3} - \frac{1}{4}\right)\right] - 1$$

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

$$= \frac{1}{6}$$

Example 3

If the probability density function of X is given by

$$f(x) = \begin{cases} \dfrac{x}{2} & 0 < x \le 1 \\ \dfrac{1}{2} & 1 < x \le 2 \\ \dfrac{3-x}{2} & 2 < x < 3 \\ 0 & \textit{otherwise} \end{cases}$$

Find the expected value of $f(x) = x^2 - 5x + 3$.

Solution

$$E\left[E\phi(x)\right] = \int_{-\infty}^{\infty} \phi(x) \, f(x) \, dx$$

$$E(x^2 - 5x + 3) = \int_{-\infty}^{\infty} (x^2 - 5x + 3) \, f(x) \, dx$$

$$= \int_0^1 (x^2 - 5x + 3)\frac{x}{2}\,\mathrm{d}x + \int_1^2 (x^2 - 5x + 3)\frac{1}{2}\,\mathrm{d}x + \int_2^3 (x^2 - 5x + 3)\left(\frac{3-x}{2}\right)\mathrm{d}x$$

$$= \frac{1}{2}\int_0^1 (x^3 - 5x^2 + 3x)\,\mathrm{d}x + \frac{1}{2}\int_1^2 (x^2 - 5x + 3)\,\mathrm{d}x + \frac{1}{2}\int_2^3 (-x^3 + 8x^2 - 18x + 9)\,\mathrm{d}x$$

$$= \frac{1}{2}\left|\frac{x^4}{4} - \frac{5x^3}{3} + \frac{3x^2}{2}\right|_0^1 + \frac{1}{2}\left|\frac{x^3}{3} - \frac{5x^2}{2} + 3x\right|_1^2 + \frac{1}{2}\left|-\frac{x^4}{4} + \frac{8x^3}{3} - \frac{18x^2}{2} + 9x\right|_2^3$$

$$= \frac{1}{2}\left(\frac{1}{4} - \frac{5}{3} + \frac{3}{2}\right) + \frac{1}{2}\left(\frac{8}{3} - 10 + 6 - \frac{1}{3} + \frac{5}{2} - 3\right) + \frac{1}{2}\left(-\frac{81}{4} + \frac{216}{3} - \frac{162}{2} + 27 + \frac{16}{4} - \frac{64}{3} + \frac{72}{2} - 18\right)$$

$$= \frac{1}{24} - \frac{13}{12} - \frac{19}{24}$$

$$= -\frac{11}{6}$$

Example 4

A continuous random variable has the probability density function

$$\begin{aligned} f(x) &= kxe^{-\lambda x} \qquad x \geq 0, \lambda > 0 \\ &= 0 \qquad \textit{otherwise} \end{aligned}$$

Determine (i) k, (ii) mean, and (iii) variance.

Solution

Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$\int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_0^{\infty} f(x)\,\mathrm{d}x = 1$$

$$0 + \int_0^{\infty} k\,x\,e^{-\lambda x}\,\mathrm{d}x = 1$$

$$k\int_0^{\infty} x e^{-\lambda x}\, dx = 1$$

$$k\left| x\frac{e^{-\lambda x}}{-\lambda} - 1\frac{e^{-\lambda x}}{\lambda^2}\right|_0^{\infty} = 1$$

$$k\left[(0-0)-\left(0-\frac{1}{\lambda^2}\right)\right] = 1$$

$$k = \lambda^2$$

Hence, $f(x) = \lambda^2 x e^{-\lambda x} \qquad x \ge 0, \lambda = 0$

$\qquad\qquad = 0 \qquad\qquad$ otherwise

(ii) Mean $= \mu = \int_{-\infty}^{\infty} x f(x)\, dx$

$$= \int_{-\infty}^{0} x\, f(x)\, dx + \int_0^{\infty} x\, f(x)\, dx$$

$$= 0 + \int_0^{\infty} x\, \lambda^2\, x e^{-\lambda x}\, dx$$

$$= \lambda^2 \int_0^{\infty} x^2\, e^{-\lambda x}\, dx$$

$$= \lambda^2 \left| x^2\left(\frac{e^{-\lambda x}}{-\lambda}\right) - 2x\left(\frac{e^{-\lambda x}}{\lambda^2}\right) + 2\left(\frac{e^{-\lambda x}}{-\lambda^3}\right)\right|_0^{\infty}$$

$$= \lambda^2\left[(0-0+0)-\left(0-0-\frac{2}{\lambda^3}\right)\right]$$

$$= \frac{2}{\lambda}$$

(iii) Variance $= \sigma^2 = \int_{-\infty}^{\infty} x^2\, f(x)\, dx - \mu^2$

$$= \int_{-\infty}^{0} x^2\, f(x)\, dx + \int_0^{\infty} x^2\, f(x)\, dx - \mu^2$$

$$= 0 + \int_0^{\infty} x^2\, \lambda^2\, x e^{-\lambda x}\, dx - \left(\frac{2}{\lambda}\right)^2$$

$$= \lambda^2 \int_0^{\infty} x^3\, e^{-\lambda x}\, dx - \frac{4}{\lambda^2}$$

$$= \lambda^2 \left| x^3 \left(\frac{e^{-\lambda x}}{-\lambda x} \right) - 3x^2 \left(\frac{e^{-\lambda x}}{\lambda^2} \right) + 6x \left(\frac{e^{-\lambda x}}{-\lambda^3} \right) - 6 \left(\frac{e^{-\lambda x}}{\lambda^4} \right) \right|_0^\infty - \frac{4}{\lambda^2}$$

$$= \lambda^2 \left[(0 - 0 + 0 - 0) - \left(0 - 0 + 0 - \frac{6}{\lambda^4} \right) \right] - \frac{4}{\lambda^2}$$

$$= \frac{6}{\lambda^2} - \frac{4}{\lambda^2}$$

$$= \frac{2}{\lambda^2}$$

Example 5

The probability density f(x) of a continuous random variable is given by $f(x) = k\, e^{-|x|}$, $-\infty < x < \infty$ *(i) show that* $k = \frac{1}{2}$, *and (ii) find the mean and variance of the distribution. (iii) Also, find the probability that the variate lies between* 0 *and* 4.

Solution

(i) Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x)\, \mathrm{d}x = 1$$

$$\int_{-\infty}^{\infty} k\, e^{-|x|}\, \mathrm{d}x = 1$$

$$k \int_{-\infty}^{\infty} e^{-|x|}\, \mathrm{d}x = 1$$

$$2k \int_0^{\infty} e^{-|x|}\, \mathrm{d}x = 1 \qquad \left[\because \ e^{-|x|} \text{ is an even function} \right]$$

$$2k \int_0^{\infty} e^{-x}\, \mathrm{d}x = 1 \qquad \left[\because \ |x| = x \quad 0 \le x \le \infty \right]$$

$$2k \left| -e^{-x} \right|_0^\infty = 1$$

$$-2k\, (0 - 1) = 1$$

$$k = \frac{1}{2}$$

Hence, $f(x) = \frac{1}{2} e^{-|x|} \qquad -\infty < x < \infty$

(ii) $$\mu = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} x\, e^{-|x|}\, \mathrm{d}x$$

$$= 0 \qquad [\because \text{ the integrand is an odd function}]$$

(iii) $$\mathrm{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} x^2\, f(x)\, \mathrm{d}x - \mu^2$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} x^2\, e^{-|x|}\, \mathrm{d}x - 0$$

$$= 2\left(\frac{1}{2}\right)\int_{0}^{\infty} x^2\, e^{-|x|}\, \mathrm{d}x \qquad [\because \text{ the integrand is an even function}]$$

$$= \int_{0}^{\infty} x^2\, e^{-|x|}\, \mathrm{d}x$$

$$= \left| x^2 \frac{e^{-x}}{-1} - 2x\frac{e^{-x}}{1} + 2\frac{e^{-x}}{-1} \right|_0^{\infty}$$

$$= 0 - (-2)$$

$$= 2$$

(iii) Probability that the variate lies between 0 and 4

$$P(0 < X < 4) = \int_0^4 f(x)\, \mathrm{d}x$$

$$= \frac{1}{2}\int_0^4 e^{-|x|}\, \mathrm{d}x$$

$$= \frac{1}{2}\int_0^4 e^{-x}\, \mathrm{d}x \qquad [\because \ |x| = x \quad 0 < x < 4]$$

$$= -\frac{1}{2}\left| e^{-x} \right|_0^4$$

$$= -\frac{1}{2}(e^{-4} - 1)$$

$$= 0.4908$$

Example 6

The daily consumption of electric power is a random variable X with probability density function

$$f(x) = k\,x\,e^{-\frac{x}{3}} \qquad x > 0$$
$$= 0 \qquad x \le 0$$

Find the value of k, the expectation of X, and the probability that on a given day, the electric consumption is more than the expected value.

Solution

Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$\int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_{0}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$0 + \int_{0}^{\infty} k\,x\,e^{-\frac{x}{3}}\,\mathrm{d}x = 1$$

$$k\left| x\left(\frac{e^{-\frac{x}{3}}}{-\frac{1}{3}}\right) - (1)\left(\frac{e^{-\frac{x}{3}}}{\frac{1}{9}}\right)\right|_0^{\infty} = 1$$

$$k\big[(0-0)-(0-9)\big] = 1$$

$$9k = 1$$

$$k = \frac{1}{9}$$

Hence, $f(x) = \frac{1}{9} x e^{-\frac{x}{3}} \qquad x > 0$

$$= 0 \qquad x \le 0$$

$$E(X) = \int_{-\infty}^{\infty} x\,f(x)\,\mathrm{d}x$$

$$= \int_{-\infty}^{0} x\,f(x)\,\mathrm{d}x + \int_{0}^{\infty} x\,f(x)\,\mathrm{d}x$$

$$= 0 + \int_{0}^{\infty} x \cdot \frac{1}{9}\,x\,e^{-\frac{x}{3}}\,\mathrm{d}x$$

$$= \frac{1}{9}\int_{0}^{\infty} x^2\,e^{-\frac{x}{3}}\,\mathrm{d}x$$

$$= \frac{1}{9}\left| x^2\left(\frac{e^{-\frac{x}{3}}}{-\frac{1}{3}}\right) - 2x\left(\frac{e^{-\frac{x}{3}}}{\frac{1}{9}}\right) + 2\left(\frac{e^{-\frac{x}{3}}}{-\frac{1}{27}}\right)\right|_0^\infty$$

$$= \frac{1}{9}(0-0+0+54)$$

$$= 6$$

$$P(X>6) = \int_0^6 f(x)\,\mathrm{d}x$$

$$= \int_0^6 \frac{1}{9} x e^{-\frac{x}{3}}\,\mathrm{d}x$$

$$= \frac{1}{9}\int_0^6 x e^{-\frac{x}{3}}\,\mathrm{d}x$$

$$= \frac{1}{9}\left| x\left(\frac{e^{-\frac{x}{3}}}{-\frac{1}{3}}\right) - 1\left(\frac{e^{-\frac{x}{3}}}{\frac{1}{9}}\right)\right|_0^\infty$$

$$= \frac{1}{9}\Big[(0-0) - \big(-18e^{-2} - 9e^{-2}\big)\Big]$$

$$= 3\,e^{-2}$$

$$= 0.406$$

Example 7

Let X be a random variable with E(X) = 10 *and Var(X) =* 25*. Find the positive values of a and b such that Y = aX – b has an expectation of* 0 *and a variance of* 1.

Solution

$$E(Y) = E(aX-b)$$

$$0 = aE(X) - b$$

$$= a(10) - b$$

$$10a - b = 0$$

$$\mathrm{Var}(Y) = \mathrm{Var}(aX-b)$$

$$1 = a^2\ \mathrm{Var}(X)$$

$$= a^2(25)$$

$$25a^2 = 1$$
$$a = \frac{1}{5}$$
$$b = 2$$

Example 8

A continuous random variable X is distributed over the interval [0, 1] *with pdf f(x) = ax*^{2} *+ bx, where a, b are constants. If the mean of X is* 0.5*, find the values of a and b.*

Solution

Since $f(x)$ is probability density function,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,\mathrm{d}x + \int_{1}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$0 + \int_{0}^{1} (ax^2 + bx)\,\mathrm{d}x + 0 = 1$$

$$\left|\frac{ax^3}{3} + \frac{bx^2}{2}\right|_0^1 = 1$$

$$\frac{a}{3} + \frac{b}{2} = 1$$

$$2a + 3b = 6 \qquad ...(1)$$

Also, $\mu = 0.5$

$$\int_{0}^{1} x\, f(x)\,\mathrm{d}x = 0.5$$

$$\int_{0}^{1} x\,(ax^2 + bx)\,\mathrm{d}x = 0.5$$

$$\int_{0}^{1} (ax^3 + bx^2)\,\mathrm{d}x = 0.5$$

$$\left|\frac{ax^4}{4} + \frac{bx^3}{3}\right|_0^1 = 0.5$$

$$\frac{a}{4} + \frac{b}{3} = 0.5$$

$$3a + 4b = 6 \qquad ...(2)$$

Solving Eqs (1) and (2),

$$a = -6, \qquad b = 6$$

Example 9

A continuous random variable X has the pdf defined by f (x) = A + Bx, $0 \le x \le 1$. *If the mean of the distribution is* $\frac{1}{3}$, *find A and B.*

Solution

Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$\int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_{0}^{1} f(x)\,\mathrm{d}x + \int_{1}^{\infty} f(x)\,\mathrm{d}x = 1$$

$$0 + \int_{0}^{1} (A + Bx)\,\mathrm{d}x + 0 = 1$$

$$\left| Ax + \frac{Bx^2}{2} \right|_0^1 = 1$$

$$A + \frac{B}{2} = 1 \qquad \text{...(1)}$$

Also,

$$\mu = \frac{1}{3}$$

$$\int^{\infty} x\, f(x)\,\mathrm{d}x = \frac{1}{3}$$

$$\int_{-\infty}^{0} x\, f(x)\,\mathrm{d}x + \int_{0}^{1} x\, f(x)\,\mathrm{d}x = \frac{1}{3}$$

$$0 + \int_{0}^{1} x\,(A + Bx)\,\mathrm{d}x = \frac{1}{3}$$

$$\int_{0}^{1} (Ax + Bx^2)\,\mathrm{d}x = \frac{1}{3}$$

$$\left| \frac{Ax^2}{2} + \frac{Bx^3}{3} \right|_0^1 = \frac{1}{3}$$

$$\frac{A}{2} + \frac{B}{3} = \frac{1}{3}$$

$$3A + 2B = 2 \qquad \text{...(2)}$$

Solving Eqs (1) and (2),

$$A = 2, \qquad B = -2$$

Example 10

A continuous random variable has probability density function $f(x) = 6(x - x^2) \qquad 0 \le x \le 1.$

Find the (i) mean, (ii) variance, (iii) median, and (iv) mode.

Solution

(i)
$$\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$$
$$= \int_{-\infty}^{0} x\, f(x)\, dx + \int_{0}^{1} x\, f(x)\, dx + \int_{1}^{\infty} x\, f(x)\, dx$$
$$= 0 + \int_{0}^{1} x\, 6(x - x^2)\, dx + 0$$
$$= 6\int_{0}^{1} (x^2 - x^3)\, dx$$
$$= 6\left|\frac{x^3}{3} - \frac{x^4}{4}\right|_0^1$$
$$= 6\left(\frac{1}{3} - \frac{1}{4}\right)$$
$$= \frac{1}{2}$$

(ii)
$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$$
$$= \int_{-\infty}^{0} x^2 f(x)\, dx + \int_{0}^{1} x^2\, f(x)\, dx + \int_{1}^{\infty} x^2 f(x)\, dx - \mu^2$$
$$= 0 + \int_0^1 x^2\, 6(x - x^2)\, dx + 0 - \frac{1}{4}$$
$$= 6\int_0^1 (x^3 - x^4)\, dx - \frac{1}{4}$$
$$= 6\left|\frac{x^4}{4} - \frac{x^5}{5}\right|_0^1 - \frac{1}{4}$$
$$= 6\left(\frac{1}{4} - \frac{1}{5}\right) - \frac{1}{4}$$

$$= \frac{6}{20} - \frac{1}{4}$$

$$= \frac{1}{20}$$

(iii) $$\int_a^M f(x)\,\mathrm{d}x = \int_M^b f(x)\,\mathrm{d}x = \frac{1}{2}$$

$$\int_0^M 6(x - x^2)\,\mathrm{d}x = \frac{1}{2}$$

$$6\left|\frac{x^2}{2} - \frac{x^3}{3}\right|_0^M = \frac{1}{2}$$

$$6\left(\frac{M^2}{2} - \frac{M^3}{3}\right) = \frac{1}{2}$$

$$3M^2 - 2M^3 = \frac{1}{2}$$

$$4M^3 - 6M^2 + 1 = 0$$

$$(2M - 1)\,(2M^2 - 2M - 1) = 0$$

$$M = \frac{1}{2} \quad \text{or} \quad M = \frac{1 \pm \sqrt{3}}{2}$$

$$M = \frac{1}{2} \text{ lies in } (0, 1)$$

Hence, median $M = \dfrac{1}{2}$

(iv) Mode is the value of x for which $f(x)$ is maximum. For $f(x)$ to be maximum, $f'(x) = 0$ and $f''(x) < 0$.

$$f'(x) = 0$$

$$6(1 - 2x) = 0$$

$$x = \frac{1}{2}$$

$$f''(x) = -12\,x$$

At $x = \dfrac{1}{2}$, $f''(x) = -12 < 0$

Hence, $f(x)$ is maximum at $x = \dfrac{1}{2}$.

Mode $= \dfrac{1}{2}$

Example 11

The probability density function of a random variable X is

$$f(x) = \frac{1}{2}\sin x \quad 0 \le x \le \pi$$
$$= 0 \qquad \textit{otherwise}$$

Find the mean, mode, and median of the distribution and also, find the probability between 0 and $\frac{\pi}{2}$.

Solution

(i) $\mu = \int_{-\infty}^{\infty} f(x)\,dx$

$$= \int_{-\infty}^{0} x\,f(x)\,dx + \int_{0}^{\pi} x\,f(x)\,dx + \int_{\pi}^{\infty} x\,f(x)\,dx$$

$$= 0 + \int_{0}^{\pi} x\left(\frac{1}{2}\sin x\right)dx + 0$$

$$= \frac{1}{2}\int_{0}^{\pi} x\sin x\,dx$$

$$= \frac{1}{2}\left|-x\cos x + \sin x\right|_{0}^{\pi}$$

$$= \frac{\pi}{2}$$

(ii) Mode is the value of x for which $f(x)$ is maximum. For $f(x)$ to be maximum, $f'(x) = 0$ and $f''(x) < 0$.

$$f'(x) = 0$$
$$\cos x = 0$$
$$x = \frac{\pi}{2}$$
$$f''(x) = -\frac{1}{2}\sin x$$

At $x = \frac{\pi}{2},\ f''(x) = -\frac{1}{2} < 0$

Hence, $f(x)$ is maximum of $x = \frac{\pi}{2}$.

Mode = $\frac{\pi}{2}$

(iii) $$\int_a^M f(x)\,dx = \int_M^b f(x)\,dx = \frac{1}{2}$$

$$\int_0^M \frac{1}{2}\sin x\,dx = \int_M^\pi \frac{1}{2}\sin x\,dx = \frac{1}{2}$$

$$\int_0^M \frac{1}{2}\sin x\,dx = \frac{1}{2}$$

$$-\frac{1}{2}|\cos x|_0^M = \frac{1}{2}$$

$$-\frac{1}{2}(\cos M - 1) = \frac{1}{2}$$

$$1 - \cos M = 0$$

$$\cos M = 0$$

$$M = \frac{\pi}{2}$$

Hence, median $M = \dfrac{\pi}{2}$

(iv) $$P\left(0 < X < \frac{\pi}{2}\right) = \int_0^{\frac{\pi}{2}} f(x)\,dx$$

$$= \int_0^{\frac{\pi}{2}} \frac{1}{2}\sin x\,dx$$

$$= -\frac{1}{2}|\cos x|_0^{\frac{\pi}{2}}$$

$$= -\frac{1}{2}(0-1)$$

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

Example 12

The cumulative distribution function of a continuous random variable X is

$$F(x) = 1 - e^{-2x} \qquad x \ge 0$$
$$= 0 \qquad x < 0$$

Find the (i) the probability density function, (ii) mean, and (iii) variance.

Solution

(i) $f(x) = \dfrac{\mathrm{d}}{\mathrm{d}x} F(x)$

$$f(x) = \frac{1}{2} e^{-2x} \qquad x \geq 0$$
$$= 0 \qquad x < 0$$

(ii) $\mu = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$

$$= \int_{-\infty}^{0} x\, f(x)\, \mathrm{d}x + \int_{0}^{\infty} x\, f(x)\, \mathrm{d}x$$
$$= 0 + \int_{0}^{\infty} x \cdot \frac{1}{2} e^{-2x}\, \mathrm{d}x$$
$$= \frac{1}{2} \int_{0}^{\infty} x\, e^{-2x}\, \mathrm{d}x$$
$$= \frac{1}{2} \left| x\left(\frac{e^{-2x}}{-2}\right) - 1\left(\frac{e^{-2x}}{4}\right) \right|_{0}^{\infty}$$
$$= \frac{1}{2}\left[(0-0) - \left(0 - \frac{1}{4}\right)\right]$$
$$= \frac{1}{8}$$

(iii) $\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2\, f(x)\, \mathrm{d}x - \mu^2$

$$= \int_{-\infty}^{0} x^2\, f(x)\, \mathrm{d}x + \int_{0}^{\infty} x^2\, f(x)\, \mathrm{d}x - \mu^2$$
$$= 0 + \int_{0}^{\infty} x^2 \cdot \frac{1}{2} e^{-2x}\, \mathrm{d}x - \left(\frac{1}{8}\right)^2$$
$$= \frac{1}{2} \int_{0}^{\infty} x^2\, e^{-2x} \mathrm{d}x - \frac{1}{64}$$
$$= \frac{1}{2} \left| x^2\left(\frac{e^{-2x}}{-2}\right) - 2x\left(\frac{e^{-2x}}{4}\right) + 2\left(\frac{e^{-2x}}{-8}\right) \right|_{0}^{\infty} - \frac{1}{64}$$
$$= \frac{1}{2}\left[(0-0-0) - \left(0 - 0 - \frac{1}{4}\right)\right] - \frac{1}{64}$$
$$= \frac{1}{8} - \frac{1}{64}$$
$$= \frac{7}{64}$$

Example 13

A continuous random variable X has the distribution function

$$F(x) = 0 \qquad x \le 1$$
$$= k(x-1)^4 \qquad 1 < x \le 3$$
$$= 1 \qquad x > 3$$

Determine (i) f(x), (ii) k, and (iii) mean.

Solution

(i) $f(x) = \dfrac{d}{dx} F(x)$

$$f(x) = 0 \qquad x \le 1$$
$$= 4k(x-1)^3 \qquad 1 < x \le 3$$
$$= 0 \qquad x > 3$$

(ii) Since $f(x)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$
$$\int_{-\infty}^{1} f(x)\,dx + \int_{1}^{3} f(x)\,dx + \int_{3}^{\infty} f(x)\,dx = 1$$
$$0 + \int_{1}^{3} 4k(x-1)^3\,dx + 0 = 1$$
$$4k\left|\frac{(x-1)^4}{4}\right|_1^3 = 1$$
$$k(16-0) = 1$$
$$k = \frac{1}{16}$$

Hence, $f(x) = 0 \qquad x \le 1$

$$= \frac{1}{4}(x-1)^3 \qquad 1 < x \le 3$$
$$= 0 \qquad x > 3$$

(iii) $\mu = \int_{-\infty}^{\infty} x\,f(x)\,dx$

$$= \int_{-\infty}^{1} x\,f(x)\,dx + \int_{1}^{3} x\,f(x)\,dx + \int_{3}^{\infty} x\,f(x)\,dx$$
$$= 0 + \int_{1}^{3} x \cdot \frac{1}{4}(x-1)^3\,dx + 0$$
$$= \frac{1}{4}\int_{1}^{3} x(x-1)^3\,dx$$

$$= \frac{1}{4}\int_0^2 (t+1)\, t^3 \, dt \qquad \begin{bmatrix} \text{Putting} \quad x-1=t \\ \text{When } x=1, t=0 \\ \text{When } x=3, t=2 \end{bmatrix}$$

$$= \frac{1}{4}\int_0^2 (t^4 + t^3)\, dt$$

$$= \frac{1}{4}\left| \frac{t^5}{5} + \frac{t^4}{4} \right|_0^2$$

$$= \frac{1}{4}\left[\left(\frac{2^5}{5} + \frac{2^4}{4}\right) - (0)\right]$$

$$= 2.6$$

EXERCISE 2.4

1. If the probability density function is given by

$$\begin{aligned} f(x) &= kx^2(1-x^3) \qquad && 0 \le x \le 1 \\ &= 0 && \text{otherwise} \end{aligned}$$

Find (i) k, (ii) $P\left(0 < X < \frac{1}{2}\right)$, (iii) $\overline{X}$, and (iv) σ^2.

$$\left[\textbf{Ans.:} \text{ (i) } 6 \text{ (ii) } \frac{15}{64} \text{ (iii) } \frac{9}{14} \text{ (iv) } \frac{9}{245}\right]$$

2. If the probability density function of a random variable is given by

$$\begin{aligned} f(x) &= kx \qquad && 0 \le x \le 2 \\ &= 2k && 2 \le x \le 4 \\ &= 6k - kx && 4 \le x \le 6 \end{aligned}$$

Find (i) k, (ii) $P(1 \le X \le 3)$, and (iii) $\overline{X}$.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{2} \text{ (ii) } \frac{1}{3} \text{ (iii) } \frac{383}{36}\right]$$

3. If the probability density of a random variable is given by

$$\begin{aligned} f(x) &= kxe^{-\frac{x}{3}} \qquad && x > 0 \\ &= 0 && x \le 0 \end{aligned}$$

Find (i) k, (ii) $\overline{X}$, and (iii) σ^2.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{9} \text{ (ii) } 6 \text{ (iii) } 18\right]$$

4. A continuous random variable has the probability density function

$$f(x) = 2e^{-2x} \qquad x > 0$$
$$\quad = 0 \qquad x \leq 0$$

Find (i) $E(X)$, (ii) $E(\overline{X})$, (iii) Var (X), and (iv) SD of X.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{2} \text{ (ii) } \frac{1}{2} \text{ (iii) } \frac{1}{4} \text{ (iv) } \frac{1}{2}\right]$$

5. A random variable X has the pdf

$$f(x) = \frac{k}{1+x^2}, -\infty < x < \infty$$

Determine (i) k, (ii) $P(X \geq 0)$, (iii) mean, and (iv) variance.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{\pi} \text{ (ii) } \frac{1}{2} \text{ (iii) 0 (iv) does not exist}\right]$$

6. The distribution function of a continuous random variable X is given by $F(x) = 1-(1+x)e^{-x}, x \geq 0.$ Find (i) pdf, (ii) mean, and (iii) variance.

$$\left[\textbf{Ans.:} \text{ (i) } f(x) = xe^{-x}, x \geq 0 \text{ (ii) 2 (iii) 2}\right]$$

7. If $f(x)$ is the probability density function of a continuous random variable, find k, mean, and variance.

$$f(x) = kx^2 \qquad 0 \leq x \leq 1$$
$$\quad = (2-x)^2 \qquad 1 \leq x \leq 2$$

$$\left[\textbf{Ans.:} \text{ 2}, \frac{11}{12}, 0.626\right]$$

8. A continuous random variable X has the probability density function given by

$$f(x) = 2ax + b \qquad 0 \leq x \leq 2$$
$$\quad = 0 \qquad \text{otherwise}$$

If the mean of the distribution is 3, find the constants a and b.

$$\left[\textbf{Ans.:} \frac{3}{2}, -\frac{5}{2}\right]$$

9. If X is a continuous random variable with probability density function given by

$$f(x) = k(x-x^3) \qquad 0 \leq x \leq 1$$
$$\quad = 0 \qquad \text{otherwise}$$

Find (i) k, (ii) mean, (iii) variance, and (iv) median.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{1}{2} \text{ (ii) } 0.06 \text{ (iii) } 0.04 \text{ (iv) } 2\right]$$

10. The probability density function of a random variable is given by

$$\begin{aligned} f(x) &= 0 && x < 2 \\ &= \frac{2x+3}{18} && 2 \le x \le 4 \\ &= 0 && x > 4 \end{aligned}$$

Find the mean and variance.

$$\left[\textbf{Ans.:} \text{ (i) } \frac{83}{27}, 0.33\right]$$

11. A continuous random variable X has the probability density function

$$\begin{aligned} f(x) &= x^3 && 0 \le x \le 1 \\ &= (2-x)^3 && 1 \le x \le 2 \\ &= 0 && \text{otherwise} \end{aligned}$$

Find $P(0.5 \le X \le 1.5)$ and mean of the distribution.

$$\left[\textbf{Ans.:} \frac{15}{32}, \frac{1}{2}\right]$$

12. The probability density function of a continuous random variable X is given by

$$f(x) = kx\,(2-x) \qquad 0 \le x \le 2$$

Find k, mean, and variance.

$$\left[\textbf{Ans.:} \frac{3}{4}, 1, \frac{1}{5}\right]$$

2.9 BINOMIAL DISTRIBUTION

Consider n independent trials of a random experiments which results in either success or failure. Let p be the probability of success remaining constant every time and $q = 1 - p$ be the probability of failure. The probability of x successes and $n - x$ failures is given by $p^x\, q^{n-x}$ (multiplication theorem of probability). But these x successes and $n - x$ failures can occur in any of the ${}^nC_x$ ways in each of which the probability is same. Hence, the probability of x successes is ${}^nC_x p^x\, q^{n-x}$.

$$P(X = x) = {}^nC_x\; p^x\; q^{n-x}, \quad x = 0, 1, 2, ..., n, \text{ where } p + q = 1$$

A random variable X is said to follow the binomial distribution if the probability of x is given by

$$P(X = x) = p(x) = {}^nC_x\ p^x\ q^{n-x}, \quad x = 0, 1, 2, \ldots, n \text{ and } q = 1 - p$$

The two constants n and p are called the parameters of the distribution.

2.9.1 Examples of Binomial Distribution

(i) Number of defective bolts in a box containing n bolts.
(ii) Number of post-graduates in a group of n people.
(iii) Number of oil wells yielding natural gas in a group of n wells test drilled.
(iv) Number of machines lying idle in a factory having n machines.

2.9.2 Conditions for Binomial Distribution

The binomial distribution holds under the following conditions:
(i) The number of trials n is finite.
(ii) There are only two possible outcomes, success or failure.
(iii) The trials are independent of each other.
(iv) The probability of success p is constant for each trial.

2.9.3 Constants of the Binomial Distribution

1. Mean of the Binomial Distribution

$$\begin{aligned}
E(X) &= \sum x\ p(x) \\
&= \sum_{x=0}^{n} x\ {}^nC_x\ p^x\ q^{n-x} \\
&= 0 \cdot {}^nC_0\ p^0\ q^n + 1 \cdot {}^nC_1\ p q^{n-1} + 2 \cdot {}^nC_2\ p^2\ q^{n-2} + \cdots + n\ p^n \\
&= np\,[q^{n-1} + {}^{(n-1)}C_1\ q^{n-2}\ p + {}^{(n-1)}C_2\ q^{n-3}\ p^2 + \cdots + p^{n-1}] \\
&= np\,(q+p)^{\,n-1} \\
&= np \qquad\qquad [\because\ p + q = 1]
\end{aligned}$$

2. Variance of the Binomial Distribution

$$\begin{aligned}
\text{Var}(X) &= E(X^2) - \mu^2 \\
&= \sum_{x=0}^{n} x^2\ p(x) - \mu^2 \\
&= \sum_{x=0}^{n} x^2\ {}^nC_x\ p^x\ q^{n-x} - \mu^2 \\
&= \sum_{x=0}^{n} \left[x + x(x-1)\right] {}^nC_x\ p^x\ q^{n-x} - \mu^2 \\
&= \sum_{x=0}^{n} x\ {}^nC_x\ p^x\ q^{n-x} + \sum_{x=0}^{n} x\,(x-1)\ {}^nC_x\ p^x\ q^{n-x} - \mu^2
\end{aligned}$$

$$= np + \sum_{x=0}^{n} x(x-1)\,\frac{n(n-1)}{x(x-1)}\cdot {}^{(n-2)}C_{x-2} p^{x}\, q^{n-x} - \mu^2$$

$$= np + \sum_{x=0}^{n} n(n-1)\cdot {}^{(n-2)}C_{x-2} p^2 p^{x-2}\, q^{n-x} - \mu^2$$

$$= np + n(n-1)\, p^2 \sum_{x=0}^{n} {}^{(n-2)}C_{x-2}\, p^{x-2}\, q^{n-x} - \mu^2$$

$$= np + n(n-1)p^2 \cdot (q+p)^{n-2} - \mu^2$$

$$= np + n(n-1)\,p^2 - \mu^2 \qquad [\because\ \ p+q=1]$$

$$= np\,[1+(n-1)p] - \mu^2$$

$$= np\,[1-p+np] - \mu^2$$

$$= np\,[q+np] - \mu^2 \qquad [\because\ \ 1-p=q]$$

$$= np\,(q+np) - (np)^2$$

$$= npq$$

3. Standard Deviation of the Binomial Distribution

$$\text{SD} = \sqrt{\text{Variance}} = \sqrt{npq}$$

4. Mode of the Binomial Distribution

Mode of the binomial distribution is the value of x at which $p(x)$ has maximum value.

Mode = integral part of $(n + 1)p$, if $(n + 1)p$ is not an integer

= $(n + 1)\, p$ and $(n + 1)\, p - 1$, if $(n + 1)\, p$ is an integer.

2.9.4 Recurrence Relation for the Binomial Distribution

For the binomial distribution,

$$P(X=x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$$

$$P(X=x+1) = {}^{n}C_{x+1}\, p^{x+1}\, q^{n-x-1}$$

$$\frac{P(X=x+1)}{P(X=x)} = \frac{{}^{n}C_{x+1}\, p^{x+1}\, q^{n-x-1}}{{}^{n}C_{x}\, p^{x}\, q^{n-x}}$$

$$= \frac{n!}{(x+1)!\,(n-x-1)!} \times \frac{x!\,(n-x)!}{n!} \cdot \frac{p}{q}$$

$$= \frac{(n-x)\,(n-x-1)!\,x!}{(x+1)\,x!\,(n-x-1)!} \cdot \frac{p}{q}$$

$$= \frac{n-x}{x+1} \cdot \frac{p}{q}$$

$$P(X = x+1) = \frac{n-x}{x+1} \cdot \frac{p}{q} \cdot P(X = x)$$

2.9.5 Binomial Frequency Distribution

If n independent trials constitute one experiment and this experiment is repeated N times, the frequency of x successes is $N\,P(X = x)$, i.e., $N\,{}^nC_x\, p^x\, q^{n-x}$. This is called expected or theoretical frequency $f(x)$ of a success.

$$\sum_{x=0}^{n} f(x) = N \sum_{x=0}^{n} P(X = x) = N \qquad \left[\because \sum_{x=0}^{n} P(X = x) = 1\right]$$

The expected or theoretical frequencies $f(0), f(1), f(2), \ldots, f(n)$ of $0, 1, 2, \ldots, n$, successes are respectively the first, second, third, ..., $(n+1)^{\text{th}}$ term in the expansion of $N(q+p)^n$. The possible number of successes and their frequencies is called a binomial frequency distribution. In practice, the expected frequencies differ from observed frequencies due to chance factor.

Example 1

The mean and standard deviation of a binomial distribution are 5 and 2. Determine the distribution.

Solution

$$\mu = np = 5$$

$$\text{SD} = \sqrt{npq} = 2$$

$$npq = 4$$

$$\frac{npq}{np} = \frac{4}{5}$$

$$\therefore \quad q = \frac{4}{5}$$

$$p = 1 - q = 1 - \frac{4}{5} = \frac{1}{5}$$

$$np = 5$$

$$n\left(\frac{1}{5}\right) = 5$$

$$\therefore \quad n = 25$$

Hence, the binomial distribution is

$$P(X = x) = {}^nC_x\, p^x\, q^{n-x}$$

$$= {}^{25}C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{25-x}, \qquad x = 0, 1, 2, \ldots, 25$$

Example 2

The mean and variance of a binomial variate are 8 *and* 6. *Find* $P(X \geq 2)$.

Solution

$$\mu = np = 8$$

$$\sigma^2 = npq = 6$$

$$\frac{npq}{np} = \frac{6}{8} = \frac{3}{4}$$

$$\therefore \quad q = \frac{3}{4}$$

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

$$np = 8$$

$$n\left(\frac{1}{4}\right) = 8$$

$$\therefore \quad n = 32$$

$$P(X = x) = {}^nC_x \; p^x \; q^{n-x}$$

$$= {}^{32}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{32-x}, \quad x = 0, 1, 2, ..., 32$$

$$P(X \geq 2) = 1 - P(X < 2)$$

$$= 1 - \left[P(X = 0) + P(X = 1)\right]$$

$$= 1 - \sum_{x=0}^{1} P(X = x)$$

$$= 1 - \sum_{x=0}^{1} {}^{32}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{32-x}$$

$$= 0.9988$$

Example 3

Suppose $P(X = 0) = 1 - P(X = 1)$. *If* $E(X) = 3$ *Var* (X), *find* $P(X = 0)$.

Solution

$$E(X) = 3\,\text{Var}\,(X)$$

$$np = 3\,npq$$

$$1 = 3\,q$$

$$\therefore \quad q = \frac{1}{3}$$

$$p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3}$$

Let $\quad P(\text{X} = 1) = p$

$$\begin{aligned} P(X = 0) &= 1 - P(X = 1) \\ &= 1 - p \\ &= 1 - \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

Example 4

The mean and variance of a binomial distribution are 4 *and* $\frac{4}{3}$ *respectively. Find* $P(X \geq 1)$.

Solution

$$\mu = np = 4$$

$$\sigma^2 = npq = \frac{4}{3}$$

$$\frac{npq}{np} = \frac{\frac{4}{3}}{4} = \frac{1}{3}$$

$$\therefore \quad q = \frac{1}{3}$$

$$p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$np = 4$$

$$n\left(\frac{2}{3}\right) = 4$$

$$\therefore \quad n = 6$$

$$\begin{aligned} P(X = x) &= {}^nC_x \ p^x \ q^{n-x} \\ &= {}^6C_x \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{6-x}, \quad x = 0, 1, 2, \ldots, 6 \end{aligned}$$

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) \\ &= 1 - P(X = 0) \end{aligned}$$

$$= 1 - {}^6C_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^6$$
$$= 0.9986$$

Example 5

A discrete random variable X has mean 6 *and variance* 2*. If it is assumed that the distribution is binomial, find the probability that* $5 \le X \le 7$.

Solution

$$\mu = np = 6$$
$$\sigma^2 = npq = 2$$
$$\frac{npq}{np} = \frac{2}{6} = \frac{1}{3}$$
$$\therefore \quad q = \frac{1}{3}$$
$$p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3}$$
$$np = 6$$
$$n\left(\frac{2}{3}\right) = 6$$
$$\therefore \quad n = 9$$
$$P(X = x) = {}^nC_x \; p^x \; q^{n-x}$$
$$= {}^9C_x \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{9-x}, \quad x = 0, 1, 2, ..., 9$$
$$P(5 \le X \le 7) = P(X = 5) + P(X = 6) + P(X = 7)$$
$$= \sum_{x=5}^{7} P(X = x)$$
$$= \sum_{x=5}^{7} {}^9C_x \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{9-x}$$
$$= \frac{4672}{6561}$$
$$= 0.7121$$

Example 6

With the usual notation, find p for a binomial distribution if $n = 6$ *and* $9P(X = 4) = P(X = 2)$.

Solution

For the binomial distribution,

$$P(X = x) = {}^nC_x \ p^x \ q^{n-x}, \ \ x = 0, 1, 2, ..., n$$

$$n = 6$$

$$9\,P(X = 4) = P(X = 2)$$

$$9\ {}^6C_4 \ p^4 \ q^2 = {}^6C_2 \ p^2 q^4$$

$$9\,p^2 = q^2 = (1 - p)^2$$

$$9\,p^2 = 1 - 2p + p^2$$

$$8p^2 + 2p - 1 = 0$$

$$p = \frac{-2 \pm \sqrt{4 + 32}}{2 \times 8} = \frac{-2 \pm 6}{16} = -\frac{1}{2}, \frac{1}{4}$$

Since probability connot be negative, $p = \dfrac{1}{4}$.

Example 7

In a binomial distribution consisting of 5 independent trials, the probability of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the parameter p of the distribution.

Solution

$$n = 5, \quad P(X = 1) = 0.4096, \quad P(X = 2) = 0.2048$$

Probability of getting x successes out of 5 trials

$$P(X = x) = {}^nC_x \ p^x \ q^{n-x} = {}^5C_x \ p^x \ q^{5-x}, x = 0, 1, 2, ..., 5$$

$$P(X = 1) = {}^5C_1 \ p\,q^4 = 0.4096 \qquad \text{...(1)}$$

$$P(X = 2) = {}^5C_2 \ p^2 \ q^3 = 0.2048 \qquad \text{...(2)}$$

Dividing Eq. (2) by Eq. (1),

$$\frac{{}^5C_2 \ p^2 \ q^3}{{}^5C_1 \ p\,q^4} = \frac{0.2048}{0.4096}$$

$$\frac{10\,p}{5\,q} = \frac{1}{2}$$

$$\frac{p}{q} = \frac{1}{4}$$

$$4\,p = q = 1 - p$$

$$5\,p = 1$$

$$p = \frac{1}{5}$$

Example 8

In a binomial distribution, the sum and product of the mean and variance are $\frac{25}{3}$ and $\frac{50}{3}$ respectively. Determine the distribution.

Solution

For the binomial distribution,

$$np + npq = \frac{25}{3}$$

$$np(1+q) = \frac{25}{3} \qquad \text{...(1)}$$

and

$$np(npq) = \frac{50}{3}$$

$$n^2\, p^2\, q = \frac{50}{3} \qquad \text{...(2)}$$

Squaring Eq. (1) and then dividing by Eq. (2),

$$\frac{n^2\, p^2\, (1+q)^2}{n^2\, p^2\, q} = \frac{\frac{625}{9}}{\frac{50}{3}}$$

$$\frac{1+2q+q^2}{q} = \frac{25}{6}$$

$$6(q^2+2q+1) = 25q$$

$$6q^2 - 13q + 6 = 0$$

$$(2q-3)(3q-2) = 0$$

$$q = \frac{3}{2} \text{ or } q = \frac{2}{3}$$

Since q can not be greater than 1,

$$q = \frac{2}{3}$$

$$p = 1 - q = 1 - \frac{2}{3} = \frac{1}{3}$$

From Eq. (1),

$$n\left(\frac{1}{3}\right)\left(1+\frac{2}{3}\right) = \frac{25}{3}$$

$$\therefore \qquad n = 15$$

Hence, the binomial distribution is

$$P(X = x) = {}^{15}C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{15-x}, \qquad x = 0, 1, 2, \ldots 15$$

Example 9

If the probability of a defective bolt is $\frac{1}{8}$*, find the (i) mean, and (ii) variance for the distribution of* 640 *defective bolts.*

Solution

$$p = \frac{1}{8}, \quad n = 640$$

$$\mu = np = \frac{640}{8} = 80$$

$$q = 1 - p = 1 - \frac{1}{8} = \frac{7}{8}$$

Variance of the distribution = $npq = 640\left(\frac{1}{8}\right)\left(\frac{7}{8}\right) = 70$

Example 10

In eight throws of a die, 5 *or* 6 *is considered as a success. Find the mean number of success and the standard deviation.*

Solution

Let p be the probability of success.

$$p = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$$

$$n = 8$$

$$\mu = np = 8\left(\frac{1}{3}\right) = \frac{8}{3}$$

$$\text{SD} = \sqrt{npq} = \sqrt{8\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)} = \frac{4}{3}$$

Example 11

4 coins are tossed simultaneously. What is the probability of getting (i) 2 heads? (ii) at least 2 heads? (iii) at most 2 heads?

Solution

Let p be the probability of getting a head in the toss of a coin.

$$p = \frac{1}{2}, \quad q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}, \quad n = 4$$

The probability of getting x heads when 4 coins are tossed

$$P(X = x) = {}^nC_x \, p^x q^{n-x} = {}^4C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}, \quad x = 0, 1, 2, 3, 4$$

(i) Probability of getting 2 heads when 4 coins are tossed

$$P(X = 2) = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

(ii) Probability of getting at least two heads when 4 coins are tossed

$$\begin{aligned} P(X \ge 2) &= P(X = 2) + P(X = 3) + P(X = 4) \\ &= \sum_{x=2}^{4} P(X = x) \\ &= \sum_{x=2}^{4} {}^4C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x} \\ &= \frac{11}{16} \end{aligned}$$

(iii) Probability getting at most 2 heads when 4 coins are tossed

$$\begin{aligned} P(X \le 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \sum_{x=0}^{2} P(X = x) \\ &= \sum_{x=0}^{2} {}^4C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x} \\ &= \frac{11}{16} \end{aligned}$$

Example 12

Two dice are thrown five times. Find the probability of getting the sum as 7 (i) at least once, (ii) two times, and (iii) $P(1 < X < 15)$.

Solution

In a single throw of two dice, a sum of 7 can occur in 6 ways out of 6 × 6 = 36 ways.

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

Let p be the probability of getting the sum as 7 in a single throw of a pair of dice.

$$p = \frac{6}{36} = \frac{1}{6}, \quad q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}, \quad n = 5$$

Probability of getting the sum x times in 5 throws of a pair of dice

$$P(X = x) = {}^nC_x\, p^x\, q^{n-x} = {}^5C_x \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{5-x}, \quad x = 0, 1, 2, \ldots, 5$$

(i) Probability of getting the sum as 7 at least once in 5 throws of two dice

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - {}^5C_0 \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^5 \\ &= 1 - \frac{3125}{7776} \\ &= \frac{4651}{7776} \end{aligned}$$

(ii) Probability of getting the sum as 7 two times in 5 throws of two dice

$$P(X = 2) = {}^5C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = \frac{625}{3888}$$

(iii) Probability of getting the sum as 7 for $P(1 < X < 5)$ in 5 throws of two dice

$$\begin{aligned} P(1 < X < 5) &= P(X = 2) + P(X = 3) + P(X = 4) \\ &= \sum_{x=2}^{4} P(X = x) \\ &= \sum_{x=2}^{4} {}^5C_x \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{5-x} \\ &= \frac{1525}{7776} \end{aligned}$$

Example 13

If 10% of the screws produced by a machine are defective, find the probability that out of 5 screws chosen at random, (i) none is defective, (ii) one is defective, and (iii) at most two are defective.

Solution

Let p be the probability of defective screws.

$$p = 0.1, \qquad q = 1 - p = 1 - 0.1 = 0.9, \qquad n = 5$$

Probability that x screws out of 5 screws are defective

$$P(X = x) = {}^nC_x\ p^x\ q^{n-x} = {}^5C_x\ (0.1)^x\ (0.9)^{5-x},\ x = 0, 1, 2, ..., 5$$

(i) Probability that none of the screws out of 5 screws is defective

$$P(X = 0) = {}^5C_0\ (0.1)^0\ (0.9)^5 = 0.5905$$

(ii) Probability that one screw out of 5 screws is defective

$$P(X = 1) = {}^5C_1\ (0.1)^1\ (0.9)^4 = 0.3281$$

(iii) Probability that at most 2 screws out of 5 screws are defective

$$\begin{aligned} P(X \le 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \sum_{x=0}^{2} P(X = x) \\ &= \sum_{x=0}^{2} {}^5C_x (0.1)^x (0.9)^{5-x} \\ &= 0.9914 \end{aligned}$$

Example 14

A multiple-choice test consists of 8 questions with 3 answers to each question (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 making, what is the probability that the student secures a distinction? **[Summer 2015]**

Solution

Let p be the probability of getting an answer to a question correctly. There are three answers to each question, out of which only one is correct.

$$p = \frac{1}{3}, \quad q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}, \quad n = 8$$

Probability of getting x correct answers in an 8 questions test

$$P(X = x) = {}^nC_x\ p^x\ q^{n-x} = {}^8C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{8-x},\ \ x = 0, 1, 2, ..., 8$$

Probability of securing a distinction, i.e., getting at least 6 correct answers out of the 8 questions

$$\begin{aligned} P(X \le 6) &= P(X = 6) + P(X = 7) + P(X = 8) \\ &= \sum_{x=6}^{8} P(X = x) \end{aligned}$$

$$= \sum_{x=6}^{8} {}^8C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{8-x}$$
$$= \frac{43}{2187}$$
$$= 0.0197$$

Example 15

A and B play a game in which their chances of winning are in the ratio 3:2. Find A's chance of winning at least three games out of the five games played.

Solution

Let p be the probability that A wins the game.

$$p = \frac{3}{3+2} = \frac{3}{5}, \quad q = 1 - p = 1 - \frac{3}{5} = \frac{2}{5}, \quad n = 5$$

Probability that A wins x games out of 5 games

$$P(X = x) = {}^nC_x\, p^x\, q^{n-x} = {}^5C_x \left(\frac{3}{5}\right)^x \left(\frac{2}{5}\right)^{5-x}, \quad x = 0, 1, 2, ..., 5$$

Probability that A wins at least 3 games

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$$
$$= \sum_{x=3}^{5} P(X = x)$$
$$= \sum_{x=3}^{5} {}^5C_x \left(\frac{3}{5}\right)^x \left(\frac{2}{5}\right)^{5-x}$$
$$= \frac{2133}{3125}$$
$$= 0.6826$$

Example 16

It has been claimed that in 60% of all solar heat installations the utility bill is reduced by at least one-third. Accordingly, what are the probabilities that the utility bill will be reduced by at least one third in (i) four of five installations? (ii) at least four of five installations?

Solution

Let p be the probability that the utility bill is reduced by one-third in the solar heat installations.

$$p = 60\% = 0.6, \qquad q = 1 - p = 1 - 0.6 = 0.4, \qquad n = 5$$

Probability that the utility bill is reduced by one-third in x installations out of 5 installations

$$P(X = x) = {}^nC_x\ p^x\ q^{n-x} = {}^5C_x(0.6)^x\ (0.4)^{5-x},\ x = 0, 1, 2, ..., 5$$

Probability that the utility bill is reduced by one-third in 4 of 5 installations

$$P(X = 5) = {}^5C_4\ (0.6)^4\ (0.4)^1 = \frac{162}{625}$$

Probability that the utility bill is reduced by one-third in at least 4 of 5 installations

$$\begin{aligned} P(X \geq 4) &= P(X = 4) + P(X = 5) \\ &= \sum_{x=4}^{5} P(X = x) \\ &= \sum_{x=4}^{5} {}^5C_x\ (0.6)^x (0.4)^{5-x} \\ &= \frac{1053}{3125} \\ &= 0.337 \end{aligned}$$

Example 17

The incidence of an occupational disease in an industry is such that the workers have a 20% chance of suffering from it. What is the probability that out of 6 workers chosen at random, four or more will suffer from the disease?

Solution

Let p be the probability of a worker suffering from the disease.

$$p = 0.2, \qquad q = 1 - p = 1 - 0.2 = 0.8, \qquad n = 6$$

Probability that x workers will suffer from the disease

$$P(X = x) = {}^nC_x\ p^x\ q^{n-x} = {}^6C_x\ (0.2)^x\ (0.8)^{6-x},\ \ x = 0, 1, 2, ..., 6$$

Probability that 4 or more workers will suffer from the disease

$$\begin{aligned} P(X \geq 4) &= P(X = 4) + P(X = 5) + P(X = 6) \\ &= \sum_{x=4}^{6} P(X = x) \\ &= \sum_{x=4}^{6} {}^6C_x\ (0.2)^x (0.8)^{6-x} \\ &= \frac{53}{3125} \\ &= 0.017 \end{aligned}$$

Example 18

The probability that a man aged 60 will live up to 70 is 0.65. What is the probability that out of 10 such men now at 60 at least 7 will live up to 70?

Solution

Let p be the probability that a man will live up to 70.

$$p = 0.65, \qquad q = 1 - p = 1 - 0.65 = 0.35, \qquad n = 10$$

Probability that x men out of 10 will live up to 70

$$P(X = x) = {}^nC_x \, p^x \, q^{n-x} = {}^{10}C_x (0.65)^x \, (0.35)^{10-x}, \; x = 0, 1, 2, ..., 10$$

Probability that at least 7 men out of 10 will live up to 70

$$\begin{aligned} P(X \ge 7) &= P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\ &= \sum_{x=7}^{10} P(X = x) \\ &= \sum_{x=7}^{10} {}^{10}C_x \, (0.65)^x \, (0.35)^{10-x} \\ &= 0.5138 \end{aligned}$$

Example 19

In a multiple-choice examination, there are 20 questions. Each question has 4 alternative answers following it and the student must select one correct answer. 4 marks are given for a correct answer and 1 mark is deducted for a wrong answer. A student must secure at least 50% of the maximum possible marks to pass the examination. Suppose a student has not studied at all, so that he answers the questions by guessing only. What is the probability that he will pass the examination?

Solution

Since there are 20 questions and each carries with 4 marks, the maximum marks are 80. If the student solves 12 questions correctly and 8 questions wrongly, he gets 48 – 8 = 40 marks required for passing. If he gets more than 12 correct answers, he gets more than 40 marks. Let p be the probability of getting a correct answer.

$$p = \frac{1}{4}, \; q = 1 - p = 1 - \frac{1}{4} = \frac{3}{4}, \; n = 20$$

Probability of getting x correct answers out of 20 answers

$$P(X = x) = {}^nC_x \, p^x \, q^{n-x} = {}^{20}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{20-x}, \; x = 0, 1, 2, ..., 20$$

Probability of passing the examination, i.e., probability of getting at least 12 correct answers out of 20 answers

$$
\begin{aligned}
P(X \ge 12) &= \sum_{x=12}^{20} P(X = x) \\
&= \sum_{x=12}^{20} {}^{20}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{20-x} \\
&= 9.3539 \times 10^{-4}
\end{aligned}
$$

Example 20

The probability of a man hitting a target is $\frac{1}{3}$. (i) If he fires 5 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 the target at least once is more than 90%?

Solution

Let p be probability of hitting a target.

$$
p = \frac{1}{3}, \quad q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}, \quad n = 5
$$

Probability of hitting the target x times out of 5 times

$$
P(X = x) = {}^nC_x \, p^x \, q^{n-x} = {}^5C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{5-x}, \quad x = 0, 1, 2, \ldots, 5
$$

(i) Probability of hitting the target at least twice out of 5 times

$$
\begin{aligned}
P(X \ge 2) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\
&= \sum_{x=2}^{5} P(X = x) \\
&= \sum_{x=2}^{5} {}^5C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{5-x} \\
&= \frac{131}{243} \\
&= 0.5391
\end{aligned}
$$

(ii) Probability of hitting the target at least once out of 5 times

$$
\begin{aligned}
P(X \ge 1) &> 0.9 \\
1 - P(X = 0) &> 0.9 \\
1 - {}^nC_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^n &> 0.9 \\
1 - \left(\frac{2}{3}\right)^n &> 0.9
\end{aligned}
$$

For $n = 6,\ 1-\left(\frac{2}{3}\right)^6 = 0.9122$

Hence, the man must fire 6 times so that the probability of hitting the target at lest once is more than 90%.

Example 21

In sampling a large number of parts manufactured by a machine, the mean number of defectives in a sample of 20 *is* 2. *Out of* 1000 *such samples, how many would be expected to contain exactly two defective parts?* **[Summer 2015]**

Solution

Let p be the probability of parts being defective.

$$\mu = np = 2, \qquad n = 20, \qquad N = 1000$$
$$np = 2$$
$$20(p) = 2$$
$$\therefore \quad p = 0.1$$
$$q = 1 - p = 1 - 0.1 = 0.9$$

Probability that the samples contain x defective parts out of 20 parts

$$P(X = x) = {}^nC_x p^x\ q^{n-x} = {}^{20}C_x (0.1)^x (0.9)^{20-x},\ x = 0, 1, 2, ..., 20$$

Probability that the samples contain exactly 2 defective parts

$$P(X = 2) = {}^{20}C_2 (0.1)^2\ (0.9)^{18} = 0.2852$$

Expected number of samples to contain exactly 2 defective parts $= N\,P(X = 2)$

$$= 1000\ (0.2852)$$
$$= 285.2$$
$$\approx 285$$

Example 22

An irregular 6*-faced die is thrown such that the probability that it gives* 3 *even numbers in* 5 *throws is twice the probability that it gives* 2 *even numbers in* 5 *throws. How many sets of exactly* 5 *trials can be expected to give no even number out of* 2500 *sets?*

Solution

Let p be the probability of getting an even number in a throw of a die.

$$n = 5, \qquad N = 2500$$

Probability of getting x even numbers in 5 throws of a die

$$P(X = x) = {}^nC_x\ p^x\ q^{n-x} = {}^5C_x\ p^x q^{5-x},\ x = 0, 1, 2, ..., 5$$

$$P(X = 3) = 2\ P(X = 2)$$

$${}^5C_3\ p^3\ q^2 = 2\left({}^5C_2\ p^2\ q^3\right)$$

$$10\ p^3\ q^2 = 20\ p^2\ q^3$$

$$p = 2q$$

$$p = 2(1 - p) = 2 - 2p$$

$$\therefore \quad p = \frac{2}{3}$$

$$q = 1 - p = 1 - \frac{2}{3} = \frac{1}{3}$$

Probability of getting no even number in 5 throws of a die

$$P(X = 0) = {}^5C_0\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^5 = \frac{1}{243}$$

Expected number of sets $= NP\ (X = 0)$

$$= \frac{2500}{243}$$

Example 23

Out of 800 *families with* 5 *children each, how many would you expect to have (i)* 3 *boys? (ii)* 5 *girls? (iii) either* 2 *or* 3 *boys? (iv) at least one boy? Assume equal probabilities for boys and girls.*

Solution

Let p be the probability of having a boy in each family.

$$p = \frac{1}{2}, \quad q = 1 - \frac{1}{2} = 1 - \frac{1}{2} = \frac{1}{2}, \quad n = 5, \quad N = 800$$

Probability of having x boys out of 5 children in each family

$$P(X = x) = {}^nC_x\ p^x\ q^{n-x} = {}^5C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x},\ x = 0, 1, 2, ..., 5$$

(i) Probability of having 3 boys out of 5 children in each family

$$P(X = 3) = {}^5C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5}{16}$$

Expected number of families having 3 boys out of 5 children $= N\ P(X = 3)$

$$= 800\left(\frac{5}{16}\right)$$

$$= 250$$

(ii) Probability of having 5 girls, i.e., no boys out of 5 children in each family

$$P(X=0) = {}^5C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

Expected number of families 5 girls out of 5 children $= NP(X=0)$

$$= 800\left(\frac{1}{32}\right)$$
$$= 25$$

(iii) Probability of having either 2 or 3 boys out of 5 children in each family

$$P(X=2)+P(X=3) = \sum_{x=2}^{3} P(X=x)$$
$$= \sum_{x=2}^{3} {}^5C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x}$$
$$= \frac{5}{8}$$

Expected number of families having either 2 of 3 boys out of 5 children

$$= N[P(X=2)+P(X=3)]$$
$$= 800\left(\frac{5}{8}\right)$$
$$= 500$$

(iv) Probability of having at least one boy out of 5 children in each family

$$P(X\ge 1) = P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)$$
$$= \sum_{x=1}^{5} P(X=x)$$
$$= \sum_{x=1}^{5} {}^5C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x}$$
$$= \frac{31}{32}$$

Expected number of families having at least-one boy out of 5 children

$$= NP(X\ge 1)$$
$$= 800\left(\frac{31}{32}\right)$$
$$= 775$$

Example 24

If hens of a certain breed lay eggs on 5 *days a week on an average, find how many days during a season of* 100 *days a will poultry keeper with* 5 *hens of this breed expect to receive at least* 4 *eggs.*

Solution

Let p be the probability of hen laying an egg on any day of a week.

$$p=\frac{5}{7}, \quad q=1-p=1-\frac{5}{7}=\frac{2}{7}, \quad n=5, \quad N=100$$

Probability of x hens laying eggs on any day of a week

$$P(X=x)={}^nC_x\, p^x\, q^{n-x}={}^5C_x\left(\frac{5}{7}\right)^x\left(\frac{2}{7}\right)^{5-x}, x=0,1,2,...,5$$

Probability of receiving at least 4 eggs on any day of a week

$$\begin{aligned}P(X\ge 4)&=P(X=4)+P(X=5)\\&=\sum_{x=4}^{5}P(X=x)\\&=\sum_{x=4}^{5}{}^5C_x\left(\frac{5}{7}\right)^x\left(\frac{2}{7}\right)^{5-x}\\&=0.5578\end{aligned}$$

Expected number of days during a season of 100 days, a poultry keeper with 5 hens of this breed will receive at least 4 eggs $= N\,P(X\ge 4)$

$$\begin{aligned}&=100\,(0.5578)\\&=55.78\\&\approx 56\end{aligned}$$

Example 25

Seven unbiased coins are tossed 128 *times and the number of heads obtained is noted as given below:*

No. of heads	0	1	2	3	4	5	6	7
Frequency	7	6	19	35	30	23	7	1

Fit a binomial distribution to the data.

Solution

Since the coin is unbiased,

$$p=\frac{1}{2}, \quad q=\frac{1}{2}, n=7, N=128$$

For binomial distribution,

$$P(X=x)={}^nC_x\, p^x\, q^{n-x}={}^7C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{7-x}, x=0,1,2,...,7$$

Theoretical or expected frequency $f(x) = N\,P(X = x)$

$$f(x) = 128\ {}^7C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{7-x} = 128\ {}^7C_x \left(\frac{1}{2}\right)^7$$

$$f(0) = 128\ {}^7C_0 \left(\frac{1}{2}\right)^7 = 1$$

$$f(1) = 128\ {}^7C_1 \left(\frac{1}{2}\right)^7 = 7$$

$$f(2) = 128\ {}^7C_2 \left(\frac{1}{2}\right)^7 = 21$$

$$f(3) = 128\ {}^7C_3 \left(\frac{1}{2}\right)^7 = 35$$

$$f(4) = 128\ {}^7C_4 \left(\frac{1}{2}\right)^7 = 35$$

$$f(5) = 128\ {}^7C_5 \left(\frac{1}{2}\right)^7 = 21$$

$$f(6) = 128\ {}^7C_6 \left(\frac{1}{2}\right)^7 = 7$$

$$f(7) = 128\ {}^7C_7 \left(\frac{1}{2}\right)^7 = 1$$

Binomial distribution

No. of heads x	0	1	2	3	4	5	6	7
Expected binomial frequency $f(x)$	1	7	21	35	35	21	7	1

Example 26

Fit a binomial distribution to the following data:

x	0	1	2	3	4	5
f	2	14	20	34	22	8

Solution

$$\text{Mean} = \frac{\sum fx}{\sum f}$$

$$= \frac{2(0) + 14(1) + 20(2) + 34(3) + 22(4) + 8(5)}{2 + 14 + 20 + 34 + 22 + 8}$$

$$= \frac{284}{100}$$
$$= 2.84$$

For binomial distribution,

$$n = 5$$
$$\mu = np = 2.84$$
$$5p = 2.84$$
$$\therefore \quad p = 0.568$$
$$q = 1 - p = 1 - 0.568 = 0.432$$

$$P(X = x) = {}^nC_x \; p^x \; q^{n-x} = {}^5C_x (0.568)^x (0.432)^{5-x}, \quad x = 0, 1, 2, ..., 5$$
$$N = \sum f = 100$$

Theoretical or expected frequency $f(x) = N\,P(X = x)$

$$f(x) = 100\;{}^5C_x\;(0.568)^x (0.432)^{5-x}$$

$$f(0) = 100\;{}^5C_0\,(0.568)^0\;(0.432)^5 = 1.505 \approx 1.5$$
$$f(1) = 100\;{}^5C_1\,(0.568)^1\;(0.432)^4 = 9.89 \approx 10$$
$$f(2) = 100\;{}^5C_2\,(0.568)^2\;(0.432)^3 = 26.01 \approx 26$$
$$f(3) = 100\;{}^5C_2\,(0.568)^3\;(0.432)^2 = 34.2 \approx 34$$
$$f(4) = 100\;{}^5C_2\,(0.568)^4\;(0.432)^1 = 22.48 \approx 22$$
$$f(5) = 100\;{}^5C_2\,(0.568)^5\;(0.432)^0 = 5.91 \approx 6$$

Binomial Distribution

x	0	1	2	3	4	5
Expected binomial frquency	1.5	10	26	34	22	6

EXERCISE 2.5

1. Find the fallacy if any in the following statements:
   (a) The mean of a binomial distribution is 6 and SD is 4.
   (b) The mean of a binomial distribution is 9 and its SD is 4.

[**Ans.:** (a) False, $q = \frac{8}{3}$ is impossible
(b) False, $q = \frac{19}{9}$ is impossible]

2. The mean and variance of a binomial distribution are 3 and 1.2 respectively. Find n, p, and $P(X < 4)$.

$$\left[\textbf{Ans.:}\ 5, 0.6, \frac{2068}{3125}\right]$$

3. Find the binomial distribution if the mean is 5 and the variance is $\frac{10}{3}$. Find $P(X = 2)$.

$$\left[\textbf{Ans.:}\ P(X=x) = {}^{25}C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{25-x}, 0.003\right]$$

4. In a binomial distribution, the mean and variance are 4 and 3 respectively. Find $P(X \geq 1)$.

$$\left[\textbf{Ans.:}\ 0.9899\right]$$

5. The odds in favour of X winning a game against Y are 4:3. Find the probability of Y winning 3 games out of 7 played.

$$\left[\textbf{Ans.:}\ 0.0929\right]$$

6. On an average, 3 out of 10 students fail in an examination. What is the probability that out of 10 students that appear for the examination none will fail?

$$\left[\textbf{Ans.:}\ 0.0282\right]$$

7. If on the average rain falls on 10 days in every thirty, find the probability (i) that the first three days of a week will be fine and remaining wet, and (ii) that rain will fall on just three days of a week.

$$\left[\textbf{Ans.:}\ \text{(i)}\ \frac{8}{2187}\ \text{(ii)}\ \frac{280}{2187}\right]$$

8. Two unbiased dice are thrown three times. Find the probability that the sum nine would be obtained (i) once, and (ii) twice.

$$\left[\textbf{Ans.:}\ \text{(i)}\ 0.26\ \text{(ii)}\ 0.03\right]$$

9. For special security in a certain protected area, it was decided to put three lightbulbs on each pole. If each bulb has probability p of burning out in the first 100 hours of service, calculate the probability that at least one of them is still good after 100 hours. If $p = 0.3$, how many bulbs would be needed on each pole to ensure with 99% safety that at least one is good after 100 hours?

$$\left[\textbf{Ans.:}\ \text{(i)}\ 1 - p^3\ \text{(ii)}\ 4\right]$$

10. It is known from past records that 80% of the students in a school do their homework. Find the probability that during a random check of 10 students, (i) all have done their homework, (ii) at the most two have not done their homework, and (iii) at least one has not done the homework.

[**Ans.:** (i) 0.1074 (ii) 0.6778 (iii) 0.8926]

11. An insurance salesman sells policies to 5 men, all of identical age and good health. According to the actuarial tables, the probability that a man of this particular age will be alive 30 years hence is $\frac{2}{3}$. Find the probability that 30 years hence (i) at least 1 man will be alive, (ii) at least 3 men will be alive, and (iii) all 5 men will be alive.

$$\left[\textbf{Ans.:} \text{(i) } \frac{242}{243} \text{ (ii) } \frac{64}{81} \text{ (iii) } \frac{32}{243}\right]$$

12. A company has appointed 10 new secretaries out of which 7 are trained. If a particular executive is to get three secretaries selected at random, what is the chance that at least one of them will be untrained?

[**Ans.:** 0.7083]

13. The overall pass rate in a university examination is 70%. Four candidates take up such an examination. What is the probability that (i) at least one of them will pass? (ii) all of them will pass the examination?

[**Ans.:** (i) 0.9919 (ii) 0.7599]

14. The normal rate of infection of a certain disease in animals is known to be 25%. In an experiment with a new vaccine, it was observed that none of the animals caught the infection. Calculate the probability of the observed result.

$$\left[\textbf{Ans.:} \frac{729}{4096}\right]$$

15. Suppose that weather records show that on the average, 5 out of 31 days in October are rainy days. Assuming a binomial distribution with each day of October as an independent trial, find the probability that the next October will have at most three rainy days.

[**Ans.:** 0.2403]

16. Assuming that half the population of a village is female and assuming that 100 samples each of 10 individuals are taken, how many samples would you expect to have 3 or less females?

[**Ans.:** 17]

17. Assuming that half the population of a town is vegetarian so that the chance of an individual being vegetarian is $\frac{1}{2}$, and assuming that 100 investigators can take a sample of 10 individuals to see whether they are vegetarians, how many investigators would you expect to report that three people or less in the sample were vegetarians?

[**Ans.:** 17]

18. The probability of failure in a physics practical examination is 20%. If 25 batches of 6 students each take the examination, in how many batches of 4 or more students would pass?

[**Ans.:** 23]

19. A lot contains 1% defective items. What should be the number of items in a lot so that the probability of finding at least one defective item in it is at least 0.95?

[**Ans.:** 299]

20. The probability that a bomb will hit the target is 0.2. Two bombs are required to destroy the target. If six bombs are used, find the probability that the target will be destroyed.

[**Ans.:** 0.3447]

21. Out of 1000 families with 4 children each, how many would you expect to have (i) 2 boys and 2 girls? (ii) at least one boy? (iii) no girl? (iv) at most 2 girls?

[**Ans.:** (i) 375 (ii) 938 (iii) 63 (iv) 69]

22. In a sampling of a large number of parts produced by a machine, the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples, how many samples would you expect to contain at least 3 defectives?

[**Ans.:** 323]

23. Five pair coins are tossed 3200 times, find the frequency distribution of the number of heads obtained. Also, find the mean and SD.

[**Ans.:** (i) 100, 500, 1000, 1000, 500, 100 (ii) 1600 (iii) 28.28]

24. Fit a binomial distribution to the following data:

x	0	1	2	3	4
f	12	66	109	59	10

[**Ans.:** 17, 67, 96, 61, 15]

2.10 POISSON DISTRIBUTION

Poisson distribution is a limiting case of binomial distribution under the following conditions:

(i) The number of trials should be infinitely large, i.e., $n \to \infty$.

(ii) The probability of successes p for each trial should be very small, i.e., $p \to 0$.

(iii) $np = \lambda$ should be finite where λ is a constant.

The binomial distribution is

$$P(X = x) = {}^nC_x\ p^x\ q^{n-x}$$
$$= {}^nC_x \left(\frac{p}{q}\right)^x q^n$$
$$= {}^nC_x \left(\frac{p}{1-p}\right)^x (1-p)^n$$

Putting $p = \dfrac{\lambda}{n}$,

$$P(X = x) = \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!}\left(\frac{\frac{\lambda}{n}}{1-\frac{\lambda}{n}}\right)^x \left(1-\frac{\lambda}{x}\right)^n$$
$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!}\frac{\lambda^x}{n^x}\frac{1}{\left(1-\frac{\lambda}{n}\right)^x}\left(1-\frac{\lambda}{x}\right)^n$$
$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!}\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$
$$= \frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left[1-\left(\frac{x-1}{n}\right)\right]}{x!}\lambda^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

Since $\displaystyle\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-x} = e^{-\lambda}$

and $\displaystyle\lim_{n\to\infty}\left(1-\frac{1}{n}\right) = \lim_{n\to\infty}\left(1-\frac{2}{n}\right) = 1$

Taking the limits of both the sides as $n \to \infty$,

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, ..., \infty$$

A random variable X is said to follow poisson distribution if the probability of x is given by

$$P(X = x) = p(x) = \frac{e^{-\lambda}\,\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

where λ is called the *parameter of the distribution*.

2.10.1 Examples of Poisson Distribution

(i) Number of defective bulbs produced by a reputed company
(ii) Number of telephone calls per minute at a switchboard
(iii) Number of cars passing a certain point in one minute
(iv) Number of printing mistakes per page in a large text
(v) Number of persons born blind per year in a large city

2.10.2 Conditions of Poisson Distribution

The Poisson distribution holds under the following conditions:

(i) The random variable X should be discrete.
(ii) The numbers of trials n is very large.
(iii) The probability of success p is very small (very close to zero).
(iv) $\lambda = np$ is finite.
(v) The occurrences are rare.

2.10.3 Constants of the Poisson Distribution

1. Mean of the Poisson Distribution

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x\, p(x) \\
&= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\,\lambda^x}{x!} \\
&= \sum_{x=0}^{\infty} \frac{x e^{-\lambda}\,\lambda\,\lambda^{x-1}}{x!} \\
&= e^{-\lambda} \cdot \lambda \sum_{x=1}^{\infty} \frac{x\,\lambda^{x-1}}{x!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} && \left[\because \quad \frac{x}{x!} = \frac{1}{(x-1)!}\right] \\
&= \lambda e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \cdots\right) \\
&= \lambda e^{-\lambda}\, e^{\lambda} \\
&= \lambda
\end{aligned}$$

2. Variance of the Poisson Distribution

$$
\begin{aligned}
\operatorname{Var}(X) &= E(X^2) - \mu^2 \\
&= \sum_{x=0}^{\infty} x^2 \; p(x) - \mu^2 \\
&= \sum_{x=0}^{\infty} x^2 \, \frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2 \\
&= \sum_{x=0}^{\infty} x\left[(x-1)+x\right]\frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2 \\
&= \sum_{x=0}^{\infty} \frac{x(x-1)\,e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{x\,e^{-\lambda}\lambda^x}{x!} - \lambda^2 \\
&= \sum_{x=0}^{\infty} \frac{x(x-1)\,e^{-\lambda}\,\lambda^{x-2}\,\lambda^2}{x(x-1)(x-2)\cdots 1} + \lambda - \lambda^2 \\
&= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda - \lambda^2 \\
&= e^{-\lambda}\,\lambda^2\left(1 + \lambda + \frac{\lambda^2}{2!} + \cdots\right) + \lambda - \lambda^2 \\
&= -e^{\lambda}\, e^{-\lambda}\, \lambda^2 + \lambda - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}
$$

3. Standard Deviation of the Poisson Distribution

$$
\text{SD} = \sqrt{\text{Variance}} = \sqrt{\lambda}
$$

4. Mode of the Poisson Distribution

Mode is the value of x for which the probability $p(x)$ is maximum.

$$
p(x) \ge p(x+1) \;\text{ and }\; p(x) \ge p(x-1)
$$

When $p(x) \ge p(x + 1)$,

$$
\begin{aligned}
\frac{e^{-\lambda}\,\lambda^x}{x!} &\ge \frac{e^{-\lambda}\,\lambda^{x+1}}{(x+1)!} \\
1 &\ge \frac{\lambda}{x+1} \\
(x+1) &\ge \lambda \\
x &\ge \lambda - 1 \qquad \text{...(2.1)}
\end{aligned}
$$

Similarly, for $p(x) \ge p(x - 1)$,

$$
x \le \lambda \qquad \text{...(2.2)}
$$

Combining Eqs (2.1) and (2.2),

$$\lambda - 1 \le x \le \lambda$$

Hence, the mode of the Poisson distribution lies between $\lambda - 1$ and λ.

Case I If λ is an integer then $\lambda - 1$ is also an integer. The distribution is bimodal and the two modes are $\lambda - 1$ and λ.

Case II If λ is not an integer, the distribution is unimodal and the mode of the Poisson distribution is an integral part of λ. The mode is the integer between $\lambda - 1$ and λ.

2.10.4 Recurrence Relation for the Poisson Distribution

For the Poisson distribution,

$$p(x) = \frac{e^{-\lambda}\ \lambda^x}{x!}$$

$$p(x+1) = \frac{e^{-\lambda}\ \lambda^{x+1}}{(x+1)!}$$

$$\frac{p(x+1)}{p(x)} = \frac{e^{-\lambda}\ \lambda^{x+1}}{(x+1)!} \cdot \frac{x!}{e^{-\lambda}\ \lambda^x}$$

$$= \frac{\lambda}{x+1}$$

$$p(x+1) = \frac{\lambda}{x+1}\, p(x)$$

Example 1

Find out the fallacy if any in the statement. "The mean of a Poisson distribution is 2 and the variance is 3."

Solution

In a Poisson distribution, the mean and variance are same. Hence, the above statement is false.

Example 2

If the mean of the Poisson distribution is 4, find

$$P(\lambda - 2\sigma < X < \lambda + 2\sigma).$$

Solution

For a Poisson distribution,

$$\text{Variance} = \lambda$$

$$\text{Mean} = \lambda = 4, \qquad \sigma = 2$$

$$P(X = x) = \frac{e^{-\lambda}\ \lambda^x}{x!} = \frac{e^{-4}4^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

$$\begin{aligned} P(\lambda - 2\sigma < X < \lambda + 2\sigma) &= P(0 < X < 8) \\ &= \sum_{x=1}^{7} P(X = x) \\ &= \sum_{x=1}^{7} \frac{e^{-4}4^x}{x!} \\ &= 0.9306 \end{aligned}$$

Example 3

If the mean of a Poisson variable is 1.8*, find (i) P(X >* 1)*, (ii) P(X =* 5)*,* and *(iii)* $P(0 < X < 5)$.

Solution

For a Poisson distribution,

$$\lambda = 1.8$$

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-1.8}1.8^x}{x!}, \quad x = 0, 1, 2, \ldots$$

(i)
$$\begin{aligned} P(X > 1) &= 1 - P(X \le 1) \\ &= 1 - \left[P(X = 0) + P(X = 1)\right] \\ &= 1 - \sum_{x=0}^{1} P(X = x) \\ &= 1 - \sum_{x=0}^{1} \frac{e^{-1.8}1.8^x}{x!} \\ &= 0.5372 \end{aligned}$$

(ii)
$$P(X = 5) = \frac{e^{-1.8}1.8^5}{5!} = 0.026$$

(iii)
$$\begin{aligned} P(0 < X < 5) &= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\ &= \sum_{x=1}^{4} P(X = x) \\ &= \sum_{x=1}^{4} \frac{e^{-1.8}1.8^x}{x!} \\ &= 0.7983 \end{aligned}$$

Example 4

If a random variable has a Poisson distribution such that P(X = 1) = P(X = 2), find (i) the mean of the distribution, (ii) P(X = 4), (iii) P(X ≥ 1), and (iv) P(1 < X < 4).

Solution

For a Poisson distribution,

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}, \quad x=0,1,2,\ldots$$

(i) $P(X=1)=P(X=2)$

$$\frac{e^{-\lambda}\lambda^1}{1!}=\frac{e^{-\lambda}\lambda^2}{2!}$$

$$\lambda^2=2\lambda$$

$$\lambda^2-2\lambda=0$$

$$\lambda(\lambda-2)=0$$

$$\lambda=0 \text{ or } \lambda=2$$

Since $\lambda \neq 0$, $\lambda=2$

Hence, $P(X=x)=\dfrac{e^{-\lambda}\lambda^x}{x!}=\dfrac{e^{-2}\,2^x}{x!}, \qquad x=0,1,2,\ldots$

(ii) $P(X=4)=\dfrac{e^{-2}2^4}{4!}=0.9022$

(iii)
$$\begin{aligned}P(X\geq 1)&=1-P(X<1)\\&=1-P(X=0)\\&=1-\frac{e^{-2}2^0}{0!}\\&=0.8647\end{aligned}$$

(iv)
$$\begin{aligned}P(1<X<4)&=P(X=2)+P(X=3)\\&=\sum_{x=2}^{3}P(X=x)\\&=\sum_{x=2}^{3}\frac{e^{-2}2^x}{x!}\\&=0.4511\end{aligned}$$

Example 5

If X is a Poisson variate such that P(X = 0) = P(X = 1), find P(X = 0) and using recurrence relation formula, find the probabilities at x = 1, 2, 3, 4, *and* 5.

Solution

For a Poisson distribution,

$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x=0,1,2,...$$

$$P(X=0)=P(X=1)$$

$$\frac{e^{-\lambda}\lambda^{0}}{0!}=\frac{e^{-\lambda}\lambda^{1}}{1!}$$

$$\lambda=1$$

Hence, $P(X=x)=\dfrac{e^{-\lambda}1^{x}}{x!}, \qquad x=0,1,2,...$

(i) $P(X=0)=\dfrac{e^{-\lambda}\lambda^{0}}{0!}=0.3678$

(ii) By recurrence relation,

$$p(x+1)=\frac{\lambda}{x+1}p(x)$$

$$p(x+1)=\frac{1}{x+1}p(x) \qquad [\because \quad \lambda=1]$$

$$p(1)=p(0)=0.3678$$

$$p(2)=\frac{1}{2}p(1)=\frac{1}{2}(0.3678)=0.1839$$

$$p(3)=\frac{1}{3}p(2)=\frac{1}{3}(0.1839)=0.0613$$

$$p(4)=\frac{1}{4}p(3)=\frac{1}{4}(0.0613)=0.015325$$

$$p(5)=\frac{1}{5}p(4)=\frac{1}{5}(0.015325)=0.003065$$

Example 6

If the variance of a Poisson variate is 3, *find the probability that (i)* $X=0$, *(ii)* $0<X\le 3$, *and (iii)* $1\le X<4$.

Solution

For a Poisson distribution,

Variance = Mean = $\lambda = 3$

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-3}3^x}{x!}, \quad x = 0, 1, 2, ...$$

(i) $P(X = 0) = \dfrac{e^{-3}3^0}{0!} = 0.0498$

(ii)
$$\begin{aligned} P(0 < X \le 3) &= P(X = 1) + P(X = 2) + P(X = 3) \\ &= \sum_{x=1}^{3} P(X = x) \\ &= \sum_{x=1}^{3} \frac{e^{-3}3^x}{x!} \\ &= 0.5974 \end{aligned}$$

(iii)
$$\begin{aligned} P(1 \le X < 4) &= P(X = 1) + P(X = 2) + P(X = 3) \\ &= \sum_{x=1}^{3} P(X = x) \\ &= \sum_{x=1}^{3} \frac{e^{-3}3^x}{x!} \\ &= 0.5974 \end{aligned}$$

Example 7

If a Poisson distribution is such that $\dfrac{3}{2}P(X = 1) = P(X = 3)$, *find (i)* $P(X \ge 1)$, *(ii)* $P(X \le 3)$, *and (iii)* $P(2 \le X \le 5)$.

Solution

For a Poisson distribution,

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, ...$$

$$\frac{3}{2}P(X = 1) = P(X = 3)$$

$$\frac{3}{2}\frac{e^{-\lambda}\lambda^1}{1!} = \frac{e^{-\lambda}\lambda^3}{3!}$$

$$\frac{3}{2}\lambda = \frac{\lambda^3}{6}$$

$$\lambda^3 - 9\lambda = 0$$

$$\lambda(\lambda^2 - 9) = 0$$
$$\lambda = 0, 3, -3$$

Since $\lambda > 0$, $\lambda = 3$

Hence, $P(x = x) = \dfrac{e^{-3} 3^x}{x!}, \qquad x = 0, 1, 2, \ldots$

(i) $$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) \\ &= 1 - P(X = 0) \\ &= 1 - \frac{e^{-3} 3^0}{0!} \\ &= 0.9502 \end{aligned}$$

(ii) $$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \sum_{x=0}^{3} P(X = x) \\ &= \sum_{x=0}^{3} \frac{e^{-3} 3^x}{x!} \\ &= 0.6472 \end{aligned}$$

(iii) $$\begin{aligned} P(2 \leq X \leq 5) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\ &= \sum_{x=2}^{5} P(X = x) \\ &= \sum_{x=2}^{5} \frac{e^{-3} 3^x}{x!} \\ &= 0.7169 \end{aligned}$$

Example 8

If X is a Poisson variate such that

$$P(X = 2) = 9\, P(X = 4) + 90\, P(X = 6)$$

Find (i) the mean of X, (ii) the variance of X, (iii) P(X < 2), (iv) P(X > 4), and (v) P(X ≥ 1).

Solution

For a Poisson distribution,

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

$$P(X = 2) = 9P(X = 4) + 90P(X = 6)$$

$$\frac{e^{-\lambda}\lambda^2}{2!} = 9\frac{e^{-\lambda}\lambda^4}{4!} + 90\frac{e^{-\lambda}\lambda^6}{6!}$$

$$= e^{-\lambda}\lambda^2\left(\frac{9\lambda^2}{4!} + \frac{90\lambda^4}{6!}\right)$$

$$\frac{1}{2} = \frac{9\lambda^2}{4!} + \frac{90\lambda^4}{6!}$$

$$\frac{1}{2} = \frac{3\lambda^2}{8} + \frac{\lambda^4}{8}$$

$$\lambda^4 + 3\lambda^2 - 4 = 0$$

$$\lambda^2 = -\frac{3 \pm \sqrt{9+16}}{2} = \frac{-3 \pm 5}{2} = 1, -4$$

Since $\lambda > 0$, $\lambda^2 = 1$

(i) Mean $= \lambda = 1$

(ii) Variance $= \lambda = 1$

$$P(X = x) = \frac{e^{-1}1^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

(iii) $P(X < 2) = P(X = 0) + P(X = 1)$

$$= \sum_{x=0}^{1} \frac{e^{-1}1^x}{x!}$$

$$= 0.7358$$

(iv) $P(X > 4) = 1 - P(X \le 4)$

$$= 1 - [P(X = 0) + (X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$$

$$= 1 - \sum_{x=0}^{4} \frac{e^{-1}1^x}{x!}$$

$$= 0.00366$$

(v) $P(X \ge 1) = 1 - P(X = 0)$

$$= 1 - \frac{e^{-1}1^0}{1!}$$

$$= 0.6321$$

Example 9

If a Poisson distribution is such that $\frac{3}{2}P(X = 1) = P(X = 3)$, *find (i)* $P(X \ge 1)$, *(ii)* $P(X \le 3)$, *and (iii)* $P(2 \le X \le 5)$.

Solution

$$\frac{3}{2}P(X=1)=P(X=3)$$

$$\frac{3}{2}\frac{e^{-\lambda}\lambda^1}{1!}=\frac{e^{-\lambda}\lambda^3}{3!}$$

$$\frac{3}{2}=\frac{\lambda^2}{6}$$

$$\lambda^2=9$$

$$\lambda=\pm 3$$

Since $\lambda>0$, $\quad \lambda=3$

$$P(X=x)=\frac{e^{-3}3^x}{x!}, \qquad x=0,1,2,\ldots$$

(i) $$\begin{aligned}P(X\geq 1)&=1-P(X<1)\\&=1-P(X=0)\\&=1-\frac{e^{-3}3^0}{0!}\\&=0.9502\end{aligned}$$

(ii) $$\begin{aligned}P(X\leq 3)&=P(X=0)+P(X=1)+P(X=2)+P(X=3)\\&=\sum_{x=0}^{3}P(X=x)\\&=\sum_{x=0}^{3}\frac{e^{-3}3^x}{x!}\\&=0.6472\end{aligned}$$

(iii) $$\begin{aligned}P(2\leq X\leq 5)&=P(X=2)+P(X=3)+P(X=4)+P(X=5)\\&=\sum_{x=2}^{5}P(X=x)\\&=\sum_{x=2}^{5}\frac{e^{-3}3^x}{x!}\\&=0.7169\end{aligned}$$

Example 10

If X is a Poisson variate such that

$$3P(X=4)=\frac{1}{2}P(X=2)+P(X=0)$$

Find (i) the mean of X, and (ii) $P(X\leq 2)$.

Solution

(i) For a Poisson distribution,

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

$$3P(X=4) = \frac{1}{2}P(X=2) + P(X=0)$$

$$3\frac{e^{-\lambda}\lambda^4}{4!} = \frac{1}{2}\frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^0}{0!}$$

$$\lambda^4 - 2\lambda^2 - 8 = 0$$

$$(\lambda^2 - 4)(\lambda^2 + 2) = 0$$

$$\lambda = \pm 2 \qquad (\because \quad \lambda \text{ is real})$$

$$\lambda = 2 \qquad (\because \quad \lambda > 0)$$

Mean $= \lambda = 2$

Hence, $\quad P(X=x) = \dfrac{e^{-2}2^x}{x!}, \qquad x = 0, 1, 2, \ldots$

(ii) $P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$

$$= \sum_{x=0}^{2} P(X=x)$$

$$= \sum_{x=0}^{2} \frac{e^{-2}2^x}{x!}$$

$$= 0.6766$$

Example 11

A manufacturer of cotterpins knows that 5% *of his products are defective. If he sells cotterpins in boxes of* 100 *and guarantees that not more than* 10 *pins will be defective, what is the approximate probability that a box will fail to meet the guaranteed quality?*

Solution

Let p be the probability of a pin being defective.

$$p = 5\% = 0.05, \qquad n = 100$$

Since p is very small and n is large, Poisson distribution is used.

$$\lambda = np = 100 \times 0.05 = 5$$

Let X be the random variable which denotes the number of defective pins in a box of 100.

Probability of x defective pins in a box of 100

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-5}5^x}{x!}, \quad x = 0, 1, 2, ...$$

Probability that a box will fail to meet the guaranteed quality

$$\begin{aligned} P(X > 10) &= 1 - P(X \le 10) \\ &= 1 - \sum_{x=0}^{10} P(X = x) \\ &= 1 - \sum_{x=0}^{10} \frac{e^{-5}5^x}{x!} \\ &= 0.0137 \end{aligned}$$

Example 12

A car-hire firm has two cars, which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson distribution with a mean of 1.5. Calculate the proportion of days on which (i) neither car is used, and (ii) the proportion of days on which some demand is refused.

Solution

$$\lambda = 1.5$$

Let X be the random variable which denotes the number of demands for a car on each day.

Probability of days on which there are x demands for a car

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-1.5}1.5^x}{x!}, \quad x = 0, 1, 2, ...$$

(i) Proportion or probability of days on which neither car is used

$$P(X = 0) = \frac{e^{-1.5}1.5^0}{0!} = 0.2231$$

(ii) Proportion or probability of days on which some demand is refused

$$\begin{aligned} P(X > 2) &= 1 - P(X \le 2) \\ &= 1 - \sum_{x=0}^{2} P(X = x) \\ &= 1 - \sum_{x=0}^{2} \frac{e^{-1.5}1.5^x}{x!} \\ &= 0.1912 \end{aligned}$$

Example 13

Six coins are tossed 6400 times. Using the Poisson distribution, what is the approximate probability of getting six heads 10 times?

Solution

Let p be the probability of getting one head with one coin.

$$p = \frac{1}{2}$$

Probability of getting 6 heads with 6 coins = $\left(\frac{1}{2}\right)^6 = \frac{1}{64}$

$$n = 6400$$

$$\lambda = np = 6400\left(\frac{1}{64}\right) = 100$$

Probability of getting x heads

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-100}100^x}{x!}, \qquad x = 0, 1, 2, ...$$

Probability of getting 6 heads 10 times

$$P(X = 10) = \frac{e^{-100}100^{10}}{10!} = 1.025 \times 10^{-30}$$

Example 14

If 2% of lightbulbs are defective, find the probability that (i) at least one is defective, and (ii) exactly 7 are defective. Also, find $P(1 < X < 8)$ in a sample of 100.

Solution

Let p be the probability of defective bulb.

$$p = 2\% = 0.02$$

$$n = 100$$

Since p is very small and n is large, Poisson distribution is used.

$$\lambda = np = 100(0.02) = 2$$

Let X be the random variable which denotes the number of defective bulbs in a sample of 100.

Probability of x defective bulb in a sample of 100

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-2}2^x}{x!}, \quad x = 0, 1, 2, ...$$

(i) Probability that at least one bulb is defective

$$P(X \geq 1) = 1 - P(X = 0)$$
$$= 1 - \frac{e^{-2} 2^0}{0!}$$
$$= 0.8647$$

(ii) Probability that exactly 7 bulbs are defective

$$P(X = 7) = \frac{e^{-2} 2^7}{7!} = 0.0034$$

(iii) $$P(1 < X < 8) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)$$
$$= \sum_{x=2}^{7} P(X = x)$$
$$= \sum_{x=2}^{7} \frac{e^{-2} 2^x}{x!}$$
$$= 0.5929$$

Example 15

An insurance company insured 4000 people against loss of both eyes in a car accident. Based on previous data, the rates were computed on the assumption that on the average, 10 persons in 100000 will have car accidents each year that result in this type of injury. What is the probability that more than 3 of the insured will collect on their policy in a given year?

Solution

Let p be the probability of loss of both eyes in a car accident.

$$p = \frac{10}{100000} = 0.0001$$
$$n = 4000$$

Since p is very small and n is large, Poisson distribution is used.

$$\lambda = np = 4000(0.0001) = 0.4$$

Let X be the random variable which denotes the number of car accidents in a group of 4000 people.

Probability of x car accidents in a group of 4000 people

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-0.4} 0.4^x}{x!}, \quad x = 0, 1, 2, ...$$

Probability that more than 3 of the insured will collect on their policy, i.e., probability of more than 3 car accidents in a group of 4000 people

$$P(X > 3) = 1 - P(X \le 3)$$
$$= 1 - \left[P(X = 0) + (X = 1) + P(X = 2) + P(X = 3)\right]$$
$$= 1 - \sum_{x=0}^{3} P(X = x)$$
$$= 1 - \sum_{x=0}^{3} \frac{e^{-0.4} 0.4^x}{x!}$$
$$= 0.00077$$

Example 16

If two cards are drawn from a pack of 52 cards which are diamonds, using Poisson distribution, find the probability of getting two diamonds at least 3 times in 51 consecutive trials of two cards drawing each time.

Solution

Let p be the probability of getting two diamonds from a pack of 52 cards.

$$p = \frac{^{13}C_2}{^{52}C_2} = \frac{3}{51}, \quad n = 51$$

Since p is very small and n is large, Poisson distribution is used.

$$\lambda = np = 51\left(\frac{3}{51}\right) = 3$$

Let X be the random variable which denotes the drawing of two diamond cards.

Probability of x trials of drawing two diamond cards in 51 trials

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-3}3^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

Probability of getting two diamond cards at least 3 times in 51 trials

$$P(X \ge 3) = 1 - P(X < 3)$$
$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$
$$= 1 - \sum_{x=0}^{2} \frac{e^{-3}3^x}{x!}$$
$$= 0.5768$$

Example 17

Suppose a book of 585 pages contains 43 typographical errors. If these errors are randomly distributed throughout the book, what is the probability that 10 pages, selected at random, will be free from errors?

Solution

Let p be the probability of errors in a page.

$$p = \frac{43}{585} = 0.0735, \quad n = 10$$

Since p is very small and n is large, Poisson distribution is used.

$$\lambda = np = 10(0.0735) = 0.735$$

Let X be the random variable which denotes the errors in the pages.
Probability of x errors in a page in a book of 585 pages

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-0.735}0.735^x}{x!}, \quad x = 0, 1, 2, ...$$

Probability that a random sample of 10 pages will contain no error.

$$P(X = 0) = \frac{e^{-0.735}0.735^0}{0!} = 0.4795$$

Example 18

A hospital switchboard receives an average of 4 *emergency calls in a* 10-*minute interval. What is the probability that (i) there are at most* 2 *emergency calls? (ii) there are exactly* 3 *emergency calls in an interval of* 10 *minutes?*

Solution

Let p be the probability of receiving emergency calls per minute.

$$p = \frac{4}{10} = 0.4, \quad n = 10$$

$$\lambda = np = 10(0.4) = 4$$

Let X be the random variable which denotes the number of emergency calls per minute.
Probability of x emergency calls per minute

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-4}4^x}{x!}, \quad x = 0, 1, 2, ...$$

Probability that there are at most 2 emergency calls

$$\begin{aligned} P(X \le 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \sum_{x=0}^{2} P(X = x) \\ &= \sum_{x=0}^{2} \frac{e^{-4}4^x}{x!} \\ &= 0.238 \end{aligned}$$

Probability that there are exactly 3 emergency calls

$$P(X=3)=\frac{e^{-4}4^3}{3!}=0.1954$$

Example 19

A manufacturer, who produces medicine bottles, finds that 0.1% *of the bottles are defective. The bottles are packed in boxes containing* 500 *bottles. A drug manufacturer buys* 100 *boxes from the producer of bottles. Using Poisson distribution, find how many boxes will contain (i) no defective bottles and (ii) at least* 2 *defective bottles.*

Solution

Let p be the probability of deflective bottles.

$$p=0.1\%=0.001$$
$$n=500$$
$$\lambda=np=500(0.001)=0.5$$

Let X be the random variable which denotes the number of defective bottles in a box of 500.

Probability of x defective bottles in a box of 500

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}=\frac{e^{-0.5}\,0.5^x}{x!},\quad x=0,1,2,\ldots$$

(i) Probability of no defective bottles in a box

$$P(X=0)=\frac{e^{-0.5}\,0.5^0}{0!}=0.6065$$

Number of boxes containing no defective bottles

$f(x)=N\,P(x=0)=100(0.6065)\approx 61$

(ii) Probability of at least 2 defective bottles

$$\begin{aligned}P(X\geq 2)&=1-P(X<2)\\&=1-\left[P(X=0)+P(X=1)\right]\\&=1\sum_{x=0}^{1}P(X=x)\\&=1-\sum_{x=0}^{1}\frac{e^{-0.5}\,0.5^x}{x!}\\&=0.0902\end{aligned}$$

Number of boxes containing at least 2 defective bottles

$f(x)=N\,P(X\geq 2)=100\,(0.0902)\approx 9$

Example 20

In a certain factory turning out blades, there is a small chance of $\frac{1}{500}$ *for any blade to be defective. The blades are supplied in packets of* 10. *Use the Poisson distribution to calculate the approximate number of packets containing no defective, one defective, and two defective blades in a consignment of* 10000 *packets.*

Solution

Let p be the probability of defective blades in a packet.

$$p = \frac{1}{500}, \qquad n = 10, \qquad N = 10000$$

$$\lambda = np = 10\left(\frac{1}{500}\right) = 0.02$$

Let X be the random variable which denotes the number of defective blades in a packet.

Probability of x defective blades in a packet

$$P(X = x) = \frac{e^{-\lambda}\,\lambda^x}{x!} = \frac{e^{-0.02}\,0.02^x}{x!}, \quad x = 0, 1, 2, \ldots$$

(i) Probability of no defective blades in a packet

$$P(X = 0) = \frac{e^{-0.02}\,0.02^0}{0!} = 0.9802$$

Number of packets with no defective blades

$f(x) = N\,P(X = 0) = 10000(0.9802) = 9802$

(ii) Probability of one defective blade in a packet

$$P(X = 1) = \frac{e^{-0.02}\,0.02^1}{1!} = 0.0196$$

Number of packets with one defective blade

$f(x) = N\,P(X = 1) = 10000\,(0.0196) = 196$

(iii) Probability of two defective blades in a packet

$$P(X = 2) = \frac{e^{-0.02}\,0.02^2}{2!} = 1.96 \times 10^{-4}$$

Number of packets with 2 defective blades

$f(x) = N\,P(X = 2) = 10000\,(1.96 \times 10^{-4}) = 1.96 \approx 2$

Example 21

The number of accidents in a year attributed to taxi drivers in a city follows Poisson distribution with a mean of 3. *Out of* 1000 *taxi drivers,* *find approximately the number of drivers with (i) no accidents in a year, and (ii) more than 3 accidents in a year.*

Solution

For a Poisson distribution,

$$\lambda = 3, N = 1000$$

Probably of x accidents in year

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-3}3^x}{x!}, \quad x = 0, 1, 2, ...$$

(i) Probability of no accidents in a year

$$P(X = 0) = \frac{e^{-3}3^0}{0!} = 0.0498$$

Number of drivers with no accidents

$$f(x) = N\,P(X = 0) = 1000(0.0498) = 49.8 \approx 50$$

(ii) Probability of more than 3 accidents in a year

$$P(X > 3) = 1 - P(X \le 3)$$
$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$
$$= 1 - \sum_{x=0}^{3} P(X = x)$$
$$= 1 - \sum_{x=0}^{3} \frac{e^{-3}3^x}{x!}$$
$$= 0.3528$$

Number of drivers with more than 3 accidents

$$f(x) = N\,P(X > 3) = 1000\,(0.3528) = 3528 \approx 353$$

Example 22

Fit a Poisson distribution to the following data:

Number of deaths *(x)*	0	1	2	3	4
Frequency *(f)*	122	60	15	2	1

Solution

$$\text{Mean} = \frac{\sum fx}{\sum f}$$
$$= \frac{122(0) + 60(1) + 15(2) + 2(3) + 1(4)}{122 + 60 + 15 + 2 + 1}$$
$$= \frac{100}{200}$$
$$= 0.5$$

For a Poisson distribution,

$$\lambda = 0.5$$

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-0.5}\,0.5^x}{x!}, \quad x = 0, 1, 2, 3, 4$$

$$N = \sum f = 100$$

Theoretical or expected frequency $f(x) = N\,P(X = x)$

$$f(x) = \frac{200\,e^{-0.5}\,0.5^x}{x!}$$

$$f(0) = \frac{200\,e^{-0.5}\,0.5^0}{0!} = 121.31 \approx 121$$

$$f(1) = \frac{200\,e^{-0.5}\,0.5^1}{1!} = 60.65 \approx 61$$

$$f(2) = \frac{200\,e^{-0.5}\,0.5^2}{2!} = 15.16 \approx 15$$

$$f(3) = \frac{200\,e^{-0.5}\,0.5^3}{3!} = 2.53 \approx 3$$

$$f(4) = \frac{200\,e^{-0.5}\,0.5^4}{4!} = 0.32 \approx 0$$

Poisson Distribution

Number of deaths *(x)*	0	1	2	3	4
Expected Poisson frequency $f(x)$	121	61	15	3	0

Example 23

Assuming that the typing mistakes per page committed by a typist follows a Poisson distribution, find the expected frequencies for the following distribution of typing mistakes:

Number of mistakes per page	0	1	2	3	4	5
Number of pages	40	30	20	15	10	5

Solution

$$\text{Mean} = \frac{\sum fx}{\sum f}$$

$$= \frac{40(0)+30(1)+20(2)+15(3)+10(4)+5(5)}{40+30+20+15+10+5}$$

$$= \frac{180}{120}$$
$$= 1.5$$

For a Poisson distribution,

$$\lambda = 1.5$$
$$P(X = x) = \frac{e^{-\lambda}\,\lambda^x}{x!} = \frac{e^{-1.5}\,1.5^x}{x!}, \quad x = 0, 1, 2, 3, 4, 5$$
$$N = \sum f = 120$$

Expected frequency $f(x) = N\,P(X = x)$

$$f(x) = \frac{120\,e^{-1.5}\,1.5^x}{x!}$$
$$f(0) = \frac{120\,e^{-1.5}\,1.5^0}{0!} = 26.78 \approx 27$$
$$f(1) = \frac{120\,e^{-1.5}\,1.5^1}{1!} = 40.16 \approx 40$$
$$f(2) = \frac{120\,e^{-1.5}\,1.5^2}{2!} = 30.12 \approx 30$$
$$f(3) = \frac{120\,e^{-1.5}\,1.5^3}{3!} = 15.06 \approx 15$$
$$f(4) = \frac{120\,e^{-1.5}\,1.5^4}{4!} = 5.65 \approx 6$$
$$f(5) = \frac{120\,e^{-1.5}\,1.5^5}{5!} = 1.69 \approx 2$$

EXERCISE 2.6

1. The mean and variance of a probability distribution is 2. Write down the distribution.

$$\left[\textbf{Ans.: } P(X = x) = \frac{e^{-2}\,2^x}{x!}, \quad x = 0, 1, 2, \ldots\right]$$

2. In a Poisson distribution, the probability $P(X = 0)$ is 20 per cent. Find the mean of the distribution.

[**Ans.:** 2.9957]

3. If X is a Poisson variate and $P(X = 0) = 6\,P(X = 3)$, find $P(X = 2)$.

[**Ans.:** 0.1839]

4. The standard deviation of a Poisson distribution is 3. Find the probability of getting 3 successes.

[**Ans.:** 0.0149]

5. The probability that a Poisson variable X takes a positive value is $1 - e^{-1.5}$. Find the variance and the probability that X lies between -1.5 and 1.5.

[**Ans.:** 1.5, 0.5578]

6. If 2 per cent bulbs are known to be defective bulbs, find the probability that in a lot of 300 bulbs, there will be 2 or 3 defective bulbs using Poisson distribution.

[**Ans.:** 0.1338]

7. In a certain manufacturing process, 5% of the tools produced turn out to be defective. Find the probability that in a sample of 40 tools, at most 2 will be defective.

[**Ans.:** 0.675]

8. If the probability that an individual suffers a bad reaction from a particular injection is 0.001, determine the probability that out of 2000 individuals (i) exactly three, and (ii) more than two individuals suffer a bad reaction.

[**Ans.:** (i) 0.1804 (ii) 0.3233]

9. It is known from past experience that in a certain plant, there are on the average 4 industrial accidents per year. Find the probability that in a given year, there will be less than 4 accidents. Assume Poisson distribution.

[**Ans.:** 0.43]

10. Find the probability that at most 5 defective fuses will be found in a box of 200 fuses, if experience shows that 2% of such fuses are defective.

[**Ans.:** 0.7851]

11. Assume that the probability of an individual coal minor being killed in a mine accident during a year is $\frac{1}{2400}$. Use appropriate statistical distribution to calculate the probability that in a mine employing 200 miners, there will be at least one fatal accident every year.

[**Ans.:** 0.07]

12. Between the hours of 2 and 4 p.m., the average number of phone calls per minute coming into the switchboard of a company is 2.5. Find the probability that during a particular minute, there will be (i) no phone call at all, (ii) 4 or less calls, and (iii) more than 6 calls.

[**Ans.:** (i) 0.0821 (ii) 0.8909 (iii) 0.0145]

13. Suppose that a local appliances shop has found from experience that the demand for tubelights is roughly distributed as Poisson with a mean of 4 tubelights per week. If the shop keeps 6 tubelights during a particular week, what is the probability that the demand will exceed the supply during that week?

[**Ans.:** 0.1106]

14. The distribution of the number of road accidents per day in a city is Poisson with a mean of 4. Find the number of days out of 100 days when there will be (i) no accident, (i) at least 2 accidents, and (iii) at most 3 accidents.

[**Ans.:** (i) 2 (ii) 91 (iii) 44]

15. A manufacturer of electric bulbs sends out 500 lots each consisting of 100 bulbs. If 5% bulbs are defective, in how many lot can we expect (i) 97 or more good bulbs? (ii) less than 96 good bulbs?

[**Ans.:** (i) 62 (ii) 132]

16. A firm produces articles, 0.1 per cent of which are defective. It packs them in cases containing 500 articles. If a wholesaler purchases 100 such cases, how many cases can be expectecd (i) to be free from defects? (ii) to have one defective article?

[**Ans.:** (i) 16 (ii) 30]

17. In a certain factory producing certain articles, the probability that an article is defective is $\frac{1}{500}$. The articles are supplied in packets of 20. Find approximately the number of packets containing no defective, one defective, two defectives in a consignment of 20000 packets.

[**Ans.:** 19200, 768, 15]

18. In a certain factory manufacturing razor blades, there is a small chance, $\frac{1}{50}$ for any blade to be defective. The blades are placed in packets, each containing 10 blades. Using the Poisson distribution, calculate the approximate number of packets containing not more than 2 defective blades in a consignment of 10000 packets.

[**Ans.:** 9988]

19. It is known that 0.5% of ballpen refills produced by a factory are defective. These refills are dispatched in packaging of equal numbers. Using a Poisson distribution, determine the number of refills in a packing to be sure that at least 95% of them contain no defective refills.

[**Ans.:** 10]

20. A manufacturer finds that the average demand per day for the mechanics to repair his new product is 1.5 over a period of one year and the demand per day is distributed as a Poisson variate. He employs two mechanics. On how many days in one year (i) would both mechanics would be free? (ii) some demand is refused?

[**Ans.:** (i) 81.4 days (ii) 69.8 days]

21. Fit a Poisson distribution to the following data:

X	0	1	2	3	4
f	211	90	19	5	0

[**Ans.:** $\lambda = 0.44$, Frequencies : 209, 92, 20, 3, 1]

22. Fit a Poisson distribution to the following data:

No. of defects per piece	0	1	2	3	4
No. of pieces	43	40	25	10	2

[**Ans.:** Frequencies: 42, 44, 24, 8, 2]

23. Fit a Poisson distribution to the following data:

X	0	1	2	3	4	5
f	142	156	69	27	5	1

[**Ans.:** Frequencies: 147, 147, 74, 24, 6, 2]

24. Fit a Poisson distribution to the following data:

X	0	1	2	3	4	5	6	7	8
f	56	156	132	92	37	22	4	0	1

[**Ans.:** Frequency : 70, 137, 135, 89, 44, 17, 6, 2, 0]

2.11 NORMAL DISTRIBUTION

A continuous random variable X is said to follow normal distribution with mean μ and variance σ^2, if its probability function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$-\infty < X < \infty, -\infty < \mu < \infty, \sigma > 0$$

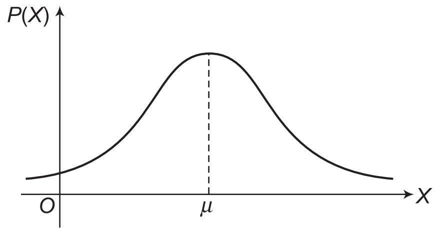


Fig. 2.1

where μ and σ are called parameters of the normal distribution. The curve representing the normal distribution is called the normal curve (Fig. 2.1).

2.11.1 Properties of the Normal Distribution

A normal probability curve, or normal curve, has the following properties:

(i) It is a bell-shaped symmetrical curve about the ordinate $X = \mu$. The ordinate is maximum at $X = \mu$.

(ii) It is a unimodal curve and its tails extend infinitely in both the directions, i.e., the curve is asymptotic to X-axis in both the directions.

(iii) All the three measures of central tendency coincide, i.e.,

mean = median = mode

(iv) The total area under the curve gives the total probability of the random variable X taking values between $-\infty$ to ∞. Mathematically,

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x = 1$$

(v) The ordinate at $X = \mu$ divides the area under the normal curve into two equal parts, i.e.,

$$\int_{-\infty}^{\mu} f(x)\,\mathrm{d}x = \int_{\mu}^{\infty} f(x)\,\mathrm{d}x = \frac{1}{2}$$

(vi) The value of $f(x)$ is always nonnegative for all values of X, i.e., the whole curve lies above the X-axis.

(vii) The points of inflexion (the point at which curvature changes) of the curve are at $X = \mu + \sigma$ and the curve changes from concave to convex at $X = \mu + \sigma$ to $X = \mu - \sigma$.

(viii) The area under the normal curve (Fig. 2.2) is distributed as follows:

(a) The area between the ordinates at $\mu - \sigma$ and $\mu + \sigma$ is 68.27%

(b) The area between the ordinates at $\mu - 2\sigma$ and $\mu + 2\sigma$ is 95.45%

(c) The area between the ordinates at $\mu - 3\sigma$ and $\mu + 3\sigma$ is 99.74%

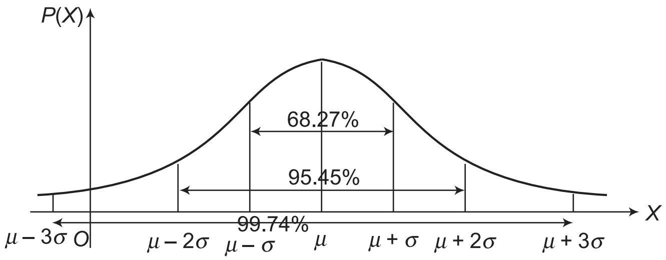


Fig. 2.2

2.11.2 Constants of the Normal Distribution

1. Mean of the Normal Distribution

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x$$

$$= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$$

Putting $\dfrac{x-\mu}{\sigma} = t,\ \mathrm{d}x = \sigma\,\mathrm{d}t$

$$E(X) = \int_{-\infty}^{\infty} (\mu + \sigma \mathrm{t}) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\,\mathrm{d}t$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\,\mathrm{d}t + \int_{-\infty}^{\infty} \sigma \frac{t}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\,\mathrm{d}t$$

Putting $t^2 = u$ in the second integral,

$2t\,\mathrm{d}t = \mathrm{d}u$

When $t \to \infty, \quad u \to \infty$

When $t \to -\infty, \quad u \to \infty$

$$E(X) = \mu \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} + \int_{\infty}^{\infty} \sigma \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u} \frac{\mathrm{d}u}{2} \qquad \left[\because \quad \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}\,\mathrm{d}t = \sqrt{2\pi}\right]$$

$$= \mu + 0 \qquad [\because \text{ the limits of integration are same}]$$

$$= \mu$$

2. Variance of the Normal Distribution

$$\mathrm{Var}(X) = E(X - \mu)^2$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \; f(x)\, \mathrm{d}x$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\;\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$$

Putting $\dfrac{x-\mu}{\sigma} = t, \mathrm{d}x = \sigma\ \mathrm{d}t$

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} \sigma^2 t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \mathrm{d}t$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2\, e^{-\frac{1}{2}t^2} \mathrm{d}t$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_{0}^{\infty} t^2\; e^{-\frac{1}{2}t^2} \mathrm{d}t \qquad [\because \quad \text{integral is an even function}]$$

Putting $\dfrac{t^2}{2} = u,$

$$t = \sqrt{2u}$$

$$\mathrm{d}t = \sqrt{2}\frac{1}{2\sqrt{u}}\, \mathrm{d}u = \frac{1}{\sqrt{2u}}\, \mathrm{d}u$$

When $t = 0,\quad u = 0$

When $t = \infty,\quad u = \infty$

$$\mathrm{Var}(X) = \frac{2\sigma^2}{\sqrt{2\pi}} \int_{0}^{\infty} 2u\, e^{-u} \frac{1}{\sqrt{2u}}\, \mathrm{d}u$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} e^{-u}\; u^{\frac{1}{2}}\, \mathrm{d}u$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \left\lceil\frac{3}{2}\right. \qquad \left[\because \quad \int_{0}^{\infty} e^{-x} x^{n-1}\, \mathrm{d}x = \lceil n \right]$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2} \left\lceil\frac{1}{2}\right.$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2} \sqrt{\pi}$$

$$= \sigma^2$$

3. Standard Deviation of the Normal Distribution

$$\mathrm{SD} = \sigma$$

4. Mode of the Normal Distribution

Mode is the value of x for which $f(x)$ is maximum. Mode is given by

$$f'(x) = 0 \text{ and } f''(x) < 0$$

For normal distribution,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Differentiating w.r.t. x,

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left[-\left(\frac{x-\mu}{\sigma^2}\right)\right]$$

$$= -\frac{x-\mu}{\sigma^2} f(x)$$

When $\quad f'(x) = 0, \quad x - \mu = 0$

$$x = \mu$$

$$f''(x) = -\frac{1}{\sigma^2}\left[(x-\mu)f'(x) + f(x)\right]$$

$$= -\frac{1}{\sigma^2}\left[(x-\mu)\left\{-\frac{(x-\mu)}{\sigma^2} f(x)\right\} + f(x)\right]$$

$$= -\frac{1}{\sigma^2} f(x)\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$$

At $x = \mu$,

$$f''(x) = \frac{f(x)}{\sigma^2} = -\frac{1}{\sigma^3\sqrt{2\pi}} < 0$$

Hence, $x = \mu$ is the mode of the normal distribution.

5. Median of the Normal Distribution

If M is median of the normal distribution,

$$\int_{-\infty}^{M} f(x)\,\mathrm{d}x = \frac{1}{2}$$

$$\int_{-\infty}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x = \frac{1}{2}$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x = \frac{1}{2} \qquad \text{...(2.3)}$$

Putting $\dfrac{x-\mu}{\sigma} = t$ in the first integral,

$$\mathrm{d}x = \sigma\,\mathrm{d}t$$

When $x = -\infty$, $t = -\infty$

When $x = \mu$, $t = 0$

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{1}{2}t^2}\sigma\,\mathrm{d}t$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{1}{2}t^2}\,\mathrm{d}t$$

$$= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{1}{2}t^2}\,\mathrm{d}t \qquad \text{[By symmetry]}$$

$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\pi}{2}}$$

$$= \frac{1}{2} \qquad \text{...(2.4)}$$

From Eqs (2.3) and (2.4),

$$\frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x = \frac{1}{2}$$

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x = 0$$

$$\int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x = 0$$

$$\mu = M\left[\because \quad \text{if} \int_a^b f(x)\,\mathrm{d}x = 0 \text{ then } a = b \text{ where } f(x) > 0\right]$$

Hence, mean = median for the normal distribution.

Note For normal distribution,

mean = median = mode = μ

Hence, the normal distribution is symmetrical.

2.11.3 Probability of a Normal Random Variable in an Interval

Let X be a normal random variable with mean μ and standard deviation σ. The probability of X lying in the interval (x_1, x_2) (Fig. 2.3) is given by

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$$

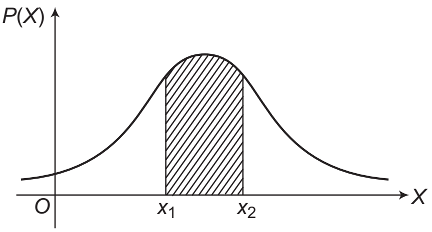


Fig. 2.3

Hence, the probability is equal to the area under the normal curve between the ordinates $X = x_1$ and $X = x_2$ respectively. $P(x_1 < X < x_2)$ can be evaluated easily by converting a normal random variable into another random variable.

Let $\quad Z = \dfrac{X-\mu}{\sigma}$ be a new random variable.

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}\left[E(X) - \mu\right] = 0$$

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathrm{Var}\,(X-\mu) = \frac{1}{\sigma^2}\mathrm{Var}(X) = 1$$

The distribution of Z is also normal. Thus, if X is a normal random variable with mean μ and standard deviation σ then $Z = \dfrac{X-\mu}{\sigma}$ is a normal random variable with mean 0 and standard deviation 1. Since the parameters of the distribution of Z are fixed, it is a known distribution and is termed *standard normal distribution*. Further, Z is termed as a *standard normal variate*. Thus, the distribution of any normal variate X can always be transformed into the distribution of the standard normal variate Z.

$$P(x_1 \le X \le x_2) = P\left[\left(\frac{x_1-\mu}{\sigma}\right) \le \left(\frac{X-\mu}{\sigma}\right) \le \left(\frac{x_2-\mu}{\sigma}\right)\right]$$

$$= P(z_1 \le Z \le z_2)$$

where $\quad z_1 = \dfrac{x_1-\mu}{\sigma}$ and $z_2 = \dfrac{x_2-\mu}{\sigma}$

This probability is equal to the area under the standard normal curve between the ordinates at $Z = z_1$ and $Z = z_2$.

Case I If both z_1 and z_2 are positive (or both negative) (Fig. 2.4),

$$\begin{aligned}P(x_1 \le X \le x_2) &= P(z_1 \le Z \le z_2)\\ &= P(0 \le Z \le z_2) - P(0 \le Z \le z_1)\\ &= (\text{Area under the normal curve from 0 to } z_2)\\ &\quad - (\text{Area under the normal curve from 0 to } z_1)\end{aligned}$$

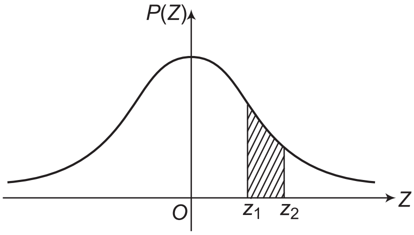


Fig. 2.4

Case II If $z_1 < 0$ and $z_2 > 0$ (Fig. 2.5),

$$\begin{aligned}P(x_1 \le X \le x_2) &= P(-z_1 \le Z \le z_2)\\ &= P(-z_1 \le Z \le 0) + P(0 \le Z \le z_2)\\ &= P(0 \le Z \le z_1) + P(0 \le Z \le z_2) \quad \text{[By symmetry]}\\ &= (\text{Area under the normal curve from 0 to } z_1)\\ &\quad + (\text{Area under the normal curve from 0 to } z_2)\end{aligned}$$

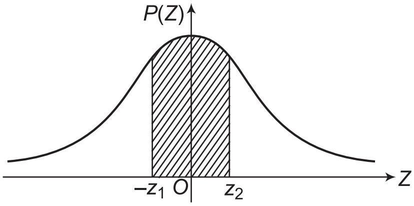


Fig. 2.5

When $X > x_1$, $Z > z_1$, the probability $P(Z > z_1)$ can be found for two cases as follows:

Case I If $z_1 > 0$ (Fig. 2.6),

$$\begin{aligned}P(X > x_1) &= P(Z > z_1)\\ &= 0.5 - P(0 \le Z \le z_1)\\ &= 0.5 - (\text{Area under the curve from 0 to } z_1)\end{aligned}$$

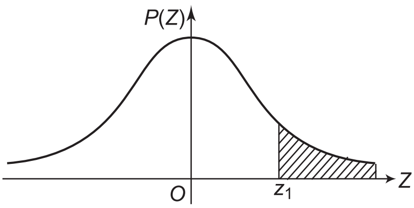


Fig. 2.6

Case II If $z_1 < 0$ (Fig. 2.7),

$$\begin{aligned}P(X > x_1) &= P(Z > -z_1)\\ &= 0.5 + P(-z_1 < Z < 0)\\ &= 0.5 + P(0 < Z < z_1) \quad \text{[By symmetry)}\\ &= 0.5 + (\text{Area under the curve from 0 to } z_1)\end{aligned}$$

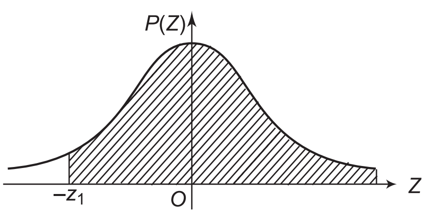


Fig. 2.7

When $X < x_1$, $Z < z_1$, the probability $P(Z < z_1)$ can be found for two cases as follows:

Case I If $z_1 > 0$ (Fig. 2.8),

$$\begin{aligned} P(X < x_1) &= P(Z < z_1) \\ &= 1 - P(Z \ge z_1) \\ &= 1 - [0.5 - P(0 < Z < z_1)] \\ &= 0.5 + P(0 < Z < z_1) \\ &= 0.5 + (\text{Area under the curve from 0 to } z_1) \end{aligned}$$

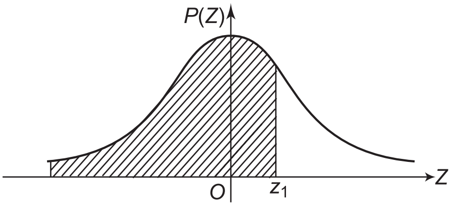


Fig. 2.8

Case II If $z_1 < 0$ (Fig. 2.9),

$$\begin{aligned} P(X < x_1) &= P(Z < -z_1) \\ &= 1 - P(Z \ge -z_1) \\ &= 1 - [0.5 + P(-z_1 \le Z \le 0)] \\ &= 1 - [0.5 + P(0 \le Z \le z_1)] \\ &\qquad\qquad [\text{By symmetry}] \\ &= 0.5 - P(0 \le Z \le z_1) \\ &= 0.5 - (\text{Area under the curve from 0 to } z_1) \end{aligned}$$

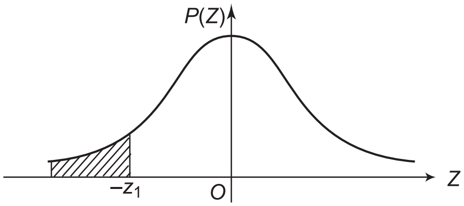


Fig. 2.9

Note

(i) $P(X < x_1) = F(x_1) = \int_{-\infty}^{x_1} f(x)\,dx$

Hence, $P(X < x_1)$ represents the area under the curve from $X = -\infty$ to $X = x_1$.

(ii) If $P(X < x_1) < 0.5$, the point x_1 lies to the left of $X = \mu$ and the corresponding value of standard normal variate will be negative (Fig. 2.10).

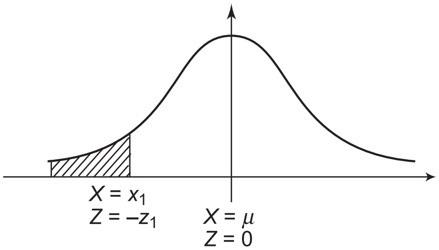


Fig. 2.10

(iii) If $P(X < x_1) > 0.5$, the point x_1 lies to the right of $x = \mu$ and the corresponding value of standard normal variate will be positive (Fig. 2.11).

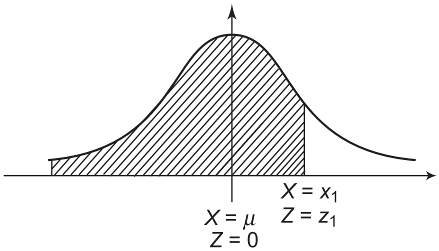
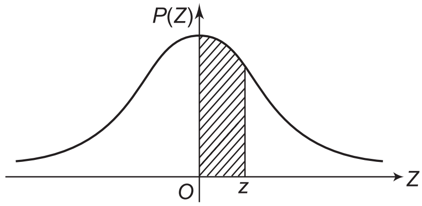


Fig. 2.11

Standard Normal (*Z*) Table, Area between 0 and *z*



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3990	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4115	0.4131	0.4147	0.4162
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

2.11.4 Uses of Normal Distribution

(i) The normal distribution can be used to approximate binomial and Poisson distributions.

(ii) It is used extensively in sampling theory. It helps to estimate parameters from statistics and to find confidence limits of the parameter.

(iii) It is widely used in testing statistical hypothesis and tests of significance in which it is always assumed that the population from which the samples have been drawn should have normal distribution.

(iv) It serves as a guiding instrument in the analysis and interpretation of statistical data.

(v) It can be used for smoothing and graduating a distribution which is not normal simply by contracting a normal curve.

Example 1

What is the probability that a standard normal variate Z will be (i) greater than 1.09? (ii) less than –1.65? (iii) lying between –1 and 1.96? (iv) lying between 1.25 and 2.75?

Solution

(i) $Z > 1.09$ (Fig. 2.12)

$$P(Z > 1.09) = 0.5 - P(0 \le Z \le 1.09)$$
$$= 0.5 - 0.3621$$
$$= 0.1379$$

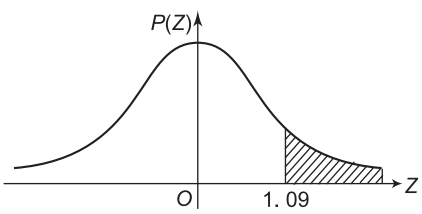


Fig. 2.12

(ii) $Z \le -1.65$ (Fig. 2.13)

$$P(Z \le -1.65) = 1 - P(Z > -1.65)$$
$$= 1 - [0.5 + P(-1.65 < Z < 0)]$$
$$= 1 - [0.5 + P(0 < Z < 1.65)] \quad \text{[By symmetry]}$$
$$= 0.5 - P(0 < Z < 1.65)$$
$$= 0.5 - 0.4505$$
$$= 0.0495$$

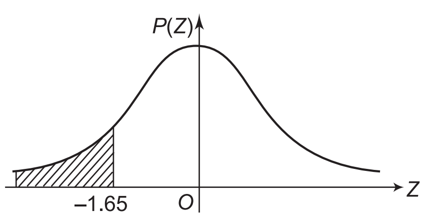


Fig. 2.13

(iii) $-1 < Z < 1.96$ (Fig. 2.14)

$$P(-1 < Z < 1.96)$$
$$= P(-1 < Z < 0) + P(0 < Z < 1.96)$$
$$= P(0 < Z < 1) + P(0 < Z < 1.96) \quad \text{[By symmetry]}$$
$$= 0.3413 + 0.4750$$
$$= 0.8163$$

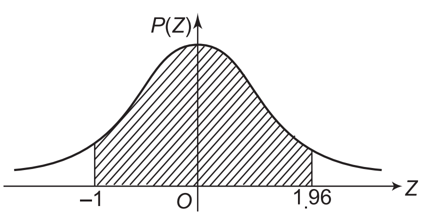


Fig. 2.14

(iv) $1.25 < Z < 2.75$ (Fig. 2.15)

$$P(1.25 < Z < 275)$$
$$= P(0 < Z < 2.75) - P(0 < Z < 1.25)$$
$$= 0.4970 - 0.3944$$
$$= 0.1026$$

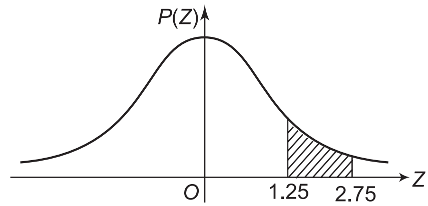


Fig. 2.15

Example 2

If X is a normal variate with a mean of 30 *and an SD of* 5, *find the probabilities that (i)* $26 \le X \le 40$, *and (ii)* $X \ge 45$.

Solution

$$\mu = 30, \qquad \sigma = 5$$

$$Z = \frac{X - \mu}{\sigma}$$

(i) When $X = 26, Z = \dfrac{26 - 30}{5} = -0.8$

When $X = 40, Z = \dfrac{40 - 30}{5} = 2$

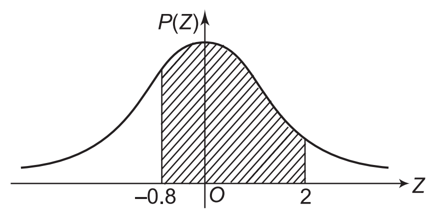


Fig. 2.16

$$P(26 \le X \le 40) = P(-0.8 \le Z \le 2) \text{ (Fig. 2.16)}$$
$$= P(-0.8 \le Z \le 0) + P(0 \le Z \le 2)$$
$$= P(0 \le Z \le 0.8) + P(0 \le Z \le 2) \qquad \text{[By symmetry]}$$
$$= 0.2881 + 0.4772$$
$$= 0.7653$$

(ii) When $X = 45, Z = \dfrac{45 - 30}{5} = 3$

$$P(X \ge 45) = P(Z \ge 3) \text{ (Fig. 2.17)}$$
$$= 0.5 - P(0 < Z < 3)$$
$$= 0.5 - 0.4987$$
$$= 0.0013$$

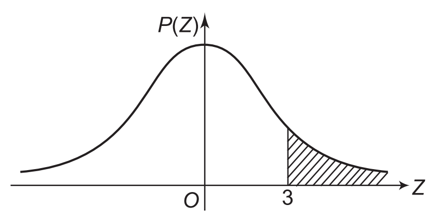


Fig. 2.17

Example 3

X is normally distributed and the mean of X is 12 *and the SD is* 4. *Find out the probability of the following:*

(i) $X \ge 20$ *(ii)* $X \le 20$ *(iii)* $0 \le X \le 12$.

Solution

$$\mu = 12, \qquad \sigma = 4$$

$$Z = \frac{X-\mu}{\sigma}$$

(i) When $X = 20, Z = \dfrac{20-12}{4} = 2$

$$P(X \geq 20) = P(Z \geq 2) \text{ (Fig. 2.18)}$$
$$= 0.5 - P(0 < Z < 2)$$
$$= 0.5 - 0.4772$$
$$= 0.0228$$

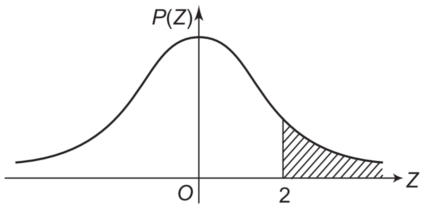


Fig. 2.18

(ii) $P(X \leq 20) = 1 - P(X > 20)$
$$= 1 - 0.0228$$
$$= 0.9772$$

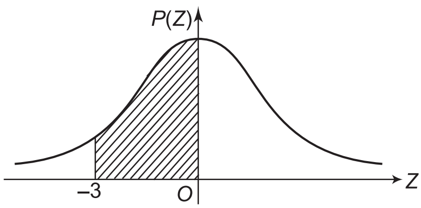


Fig. 2.19

(iii) When $X = 0, Z = \dfrac{0-12}{4} = -3$

When $X = 12, Z = \dfrac{12-12}{4} = 0$

$$P(0 \leq X \leq 12) = P(-3 \leq Z \leq 0) \text{ (Fig. 2.19)}$$
$$= P(0 \leq Z \leq 3) \qquad \text{[By symmetry]}$$
$$= 0.4987$$

Example 4

If X is normally distributed with a mean of 2 *and an SD of* 0.1, *find* $P(|X-2|) \geq 0.01)$?

Solution:

$$\mu = 2, \qquad \sigma = 0.1$$

$$Z = \frac{X-\mu}{\sigma}$$

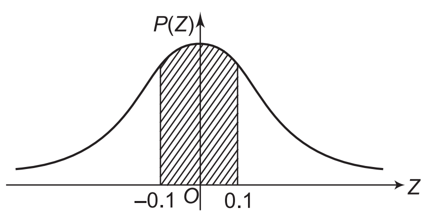


Fig. 2.20

When $X = 1.99, Z = \dfrac{1.99-2}{0.1} = -0.1$

When $X = 2.01, Z = \dfrac{2.01-2}{0.1} = 0.1$

$$
\begin{aligned}
P(|X-2| \le 0.01) &= P(1.99 \le X \le 2.01) \text{ (Fig. 2.20)} \\
&= P(-0.1 \le Z \le 0.1) \\
&= P(-0.1 \le Z \le 0) + P(0 \le Z \le 0.1) \\
&= P(0 \le Z \le 0.1) + P(0 \le Z \le 0.1) \qquad \text{[By symmetry]} \\
&= 2P(0 < Z \le 0.1) \\
&= 2(0.0398) \\
&= 0.0796
\end{aligned}
$$

$$
\begin{aligned}
P(|X-2| \ge 0.01) &= 1 - P(|X-2| < 0.01) \\
&= 1 - 0.0796 \\
&= 0.9204
\end{aligned}
$$

Example 5

If X is a normal variate with a mean of 120 *and a standard deviation of* 10*, find c such that (i) P*(*X* > *c*) = 0.02, *and (ii) P*(*X*< *c*) = 0.05.

Solution

For normal variate *X*,

$$\mu = 120, \quad \sigma = 10$$

$$Z = \frac{X-\mu}{\sigma}$$

(i) $P(X > c) = 0.02$

$$
\begin{aligned}
P(X < c) &= 1 - P(X \ge c) \\
&= 1 - 0.02 \\
&= 0.98
\end{aligned}
$$

Since $P(X < c) > 0.5$, the corresponding value of Z will be positive.

$$
\begin{aligned}
P(X > c) &= P(Z > z_1) \text{ (Fig. 2.21)} \\
0.02 &= 0.5 - P(0 \le Z \le z_1) \\
P(0 \le Z \le z_1) &= 0.48
\end{aligned}
$$

$\therefore$ $z_1 = 2.05$ [From normal table]

$$Z = \frac{c-120}{10} = z_1 = 2.05$$

$$c = 2.05(10) + 120 = 140.05$$

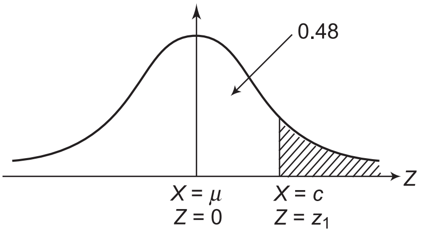


Fig. 2.21

(ii) Since $P(X<c)<0.5$, the corresponding value of Z will be negative.

$$
\begin{aligned}
P(X < c) &= P(Z < -z_1) \text{ (Fig. 2.22)} \\
0.05 &= 1 - P(Z \ge -z_1) \\
0.05 &= 1 - \left[0.5 + P(-z_1 \le Z \le 0)\right]
\end{aligned}
$$

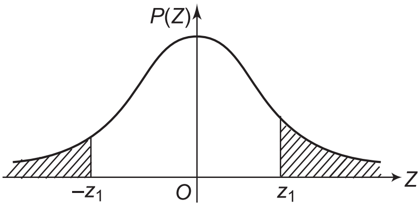


Fig. 2.22

$$0.05 = 1 - \left[0.5 + P(0 \le Z \le z_1)\right] \qquad \text{[By symmetry]}$$
$$0.05 = 0.5 - P(0 \le Z \le z_1)$$
$$P(0 \le Z \le z_1) = 0.5 - 0.05 = 0.45$$
$$\therefore \qquad z_1 = -1.64 \quad \text{[From normal table]}$$
$$Z = \frac{c - 120}{10} = z_1 = -1.64$$
$$c = 10(-1.64) + 120 = 103.6$$

Example 6

A manufacturer knows from his experience that the resistances of resistors he produces is normal with $\mu = 100$ *ohms and SD* $= \sigma = 2$ *ohms. What percentage of resistors will have resistances between* 98 *ohms and* 102 *ohms?*

Solution

Let X be the random variable which denotes the resistances of the resistors.

$$\mu = 100, \qquad \sigma = 2$$
$$Z = \frac{X - \mu}{\sigma}$$

When $X = 98, \quad Z = \dfrac{98 - 100}{2} = -1$

When $X = 102, \quad Z = \dfrac{102 - 100}{2} = 1$

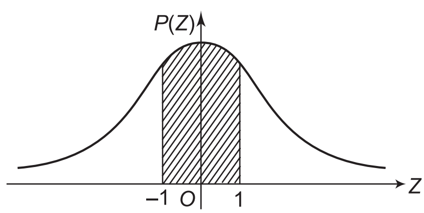


Fig. 2.23

$$\begin{aligned} P(98 \le X \le 102) &= P(-1 \le Z \le 1) \quad \text{(Fig. 2.23)} \\ &= P(-1 \le Z \le 0) + P(0 \le Z \le 1) \\ &= P(0 \le Z \le 1) + P(0 \le Z \le 1) \qquad \text{[By symmetry]} \\ &= 2P(0 \le Z \le 1) \\ &= 2(0.3413) \\ &= 0.6826 \end{aligned}$$

Hence, the percentage of resistors have resistances between 98 ohms and 102 ohms = 68.26%.

Example 7

The average seasonal rainfall in a place is 16 *inches with an SD of* 4 *inches. What is the probability that the rainfall in that place will be between* 20 *and* 24 *inches in a year?*

Solution

Let X be the random variable which denotes the seasonal rainfall in a year.

$$\mu = 16, \qquad \sigma = 4$$

$$Z = \frac{X - \mu}{\sigma}$$

When $X = 20$, $\quad Z = \dfrac{20 - 16}{4} = 1$

When $X = 24$, $\quad Z = \dfrac{24 - 16}{4} = 2$

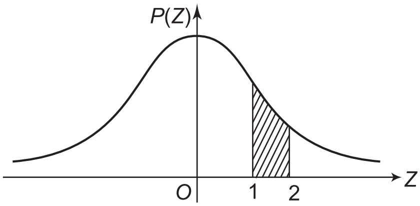


Fig. 2.24

$$\begin{aligned} P(20 < X < 24) &= P(1 < Z < 2) \quad \text{(Fig. 2.24)} \\ &= P(0 < Z < 2) - P(0 < Z < 1) \\ &= 0.4772 - 0.3413 \\ &= 0.1359 \end{aligned}$$

Example 8

The lifetime of a certain kind of batteries has a mean life of 400 *hours and the standard deviation as* 45 *hours. Assuming the distribution of lifetime to be normal, find (i) the percentage of batteries with a lifetime of at least* 470 *hours, (ii) the proportion of batteries with a lifetime between* 385 *and* 415 *hours, and (iii) the minimum life of the best* 5% *of batteries.*

Solution

Let X be the random variable which denotes the lifetime of a certain kind of batteries.

$$\mu = 400, \qquad \sigma = 45$$

$$Z = \frac{X - \mu}{\sigma}$$

(i) When $X = 470$,

$$Z = \frac{470 - 400}{45} = 1.56$$

$$\begin{aligned} P(X \geq 470) &= P(Z \geq 1.56) \quad \text{(Fig. 2.25)} \\ &= 0.5 - P(0 < Z < 1.56) \\ &= 0.5 - 0.4406 \\ &= 0.0594 \end{aligned}$$

Hence, the percentage of batteries with a lifetime of at least 470 hours $= 5.94\%$.

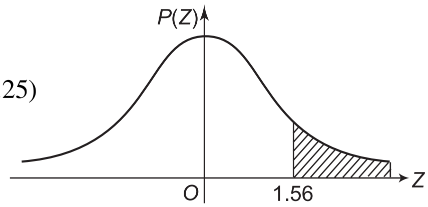


Fig. 2.25

(ii) When $X = 385$,

$$Z = \frac{385-400}{45} = -0.33$$

When $X = 415$,

$$Z = \frac{415-400}{45} = 0.33$$

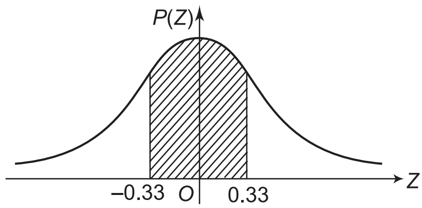


Fig. 2.26

$$\begin{aligned} P(385 < X < 415) &= P(-0.33 < Z < 0.33) \quad \text{(Fig. 2.26)} \\ &= P(-0.33 < Z < 0) + P(0 < Z < 0.33) \\ &= P(0 < Z < 0.33) + P(0 < Z < 0.33) \qquad \text{[By symmetry]} \\ &= 2\,P(0 < Z < 0.33) \\ &= 2(0.1293) \\ &= 0.2586 \end{aligned}$$

Hence, the proportion of batteries with a lifetime between 385 and 415 hours = 25.86%.

(iii) $P(X > x_1) = 0.05$ (Fig. 2.27)

$P(X > x_1) = P(Z > z_1)$

$0.05 = 0.5 - P(0 \le Z \le z_1)$

$P(0 \le Z \le z_1) = 0.5 - 0.05 = 0.45$

$\therefore$ $z_1 = 1.65$ [From normal table]

$$Z = \frac{x_1 - 400}{45} = z_1 = 1.65$$

$\therefore$ $x_1 = 1.65(45) + 400 = 474.25$ hours

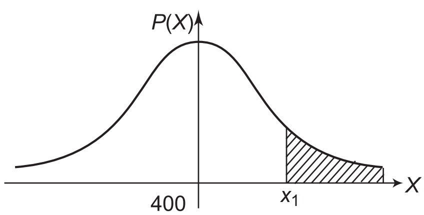


Fig. 2.27

Example 9

If the weights of 300 *students are normally distributed with a mean of* 68 *kg and a standard deviation of* 3 *kg, how many students have weights (i) greater than* 72 *kg? (ii) less than or equal to* 64 *kg? (iii) between* 65 *kg and* 71 *kg inclusive?*

Solution

Let X be the random variable which denotes the weight of a student.

$$\mu = 68, \quad \sigma = 3, \quad N = 300$$

$$Z = \frac{X-\mu}{\sigma}$$

(i) When $X = 72$, $\quad Z = \dfrac{72-68}{3} = 1.33$

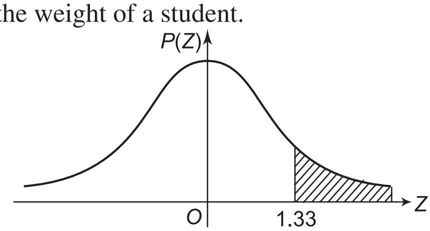


Fig. 2.28

$$
\begin{aligned}
P(X > 72) &= P(Z > 1.33) \quad \text{(Fig. 2.28)}\\
&= 0.5 - P(0 \le Z \le 1.33)\\
&= 0.5 - 0.4082\\
&= 0.0918
\end{aligned}
$$

Number of students with weights more than 72 kg $= N\ P(X > 72)$

$$
\begin{aligned}
&= 300(0.0918)\\
&= 27.54\\
&\approx 28
\end{aligned}
$$

(ii) When $X = 64$, $\quad Z = \dfrac{64-68}{3} = -1.33$

$$
\begin{aligned}
P(X \le 64) &= P(Z \le -1.33) \quad \text{(Fig. 2.29)}\\
&= P(Z \ge 1.33) \qquad \text{[By symmetry]}\\
&= 0.5 - P(0 < Z < 1.33)\\
&= 0.5 - 0.4082\\
&= 0.0918
\end{aligned}
$$

Number of students with weights less than or equal to 64 kg

$$
\begin{aligned}
&= N\ P(X \le 64)\\
&= 300(0.0918)\\
&= 27.54\\
&\approx 28
\end{aligned}
$$

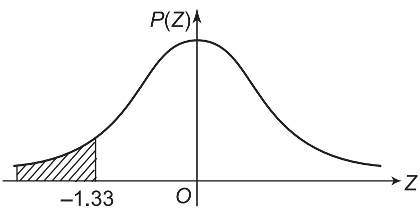


Fig. 2.29

(iii) When $X = 65$, $\quad Z = \dfrac{65-68}{3} = -1$

When $X = 71$, $\quad Z = \dfrac{71-68}{3} = 1$

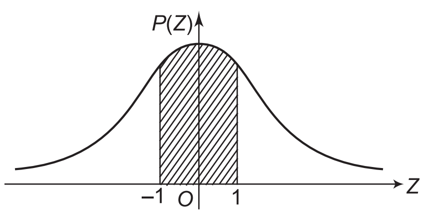


Fig. 2.30

$$
\begin{aligned}
P(65 \le X \le 71) &= P(-1 \le Z \le 1) \quad \text{(Fig. 2.30)}\\
&= P(-1 \le Z \le 0) + P(0 \le Z \le 1)\\
&= P(0 \le Z \le 1) + P(0 \le Z \le 1) \qquad \text{[By symmetry]}\\
&= 2P(0 \le Z \le 1)\\
&= 2(0.3413)\\
&= 0.6826
\end{aligned}
$$

Number of students with weights between 65 and 71 kg $= N\ P(65 \le X \le 71)$

$$
\begin{aligned}
&= 300(0.6826)\\
&= 204.78\\
&\approx 205
\end{aligned}
$$

Example 10

The mean yield for a one-acre plot is 662 kg with an SD of 32 kg. Assuming normal distribution, how many one-acre plots in a batch of 1000 plots would you expect to have yields (i) over 700 kg? (ii) below 650 kg? (iii) What is the lowest yield of the best 100 plots?

Solution

Let X be the random variable which denotes the yield for the one-acre plot.

$$\mu = 662, \qquad \sigma = 32, \qquad N = 1000$$

$$Z = \frac{X-\mu}{\sigma}$$

(i) When $X = 700$, $\quad Z = \dfrac{700-662}{32} = 1.19$

$$\begin{aligned} P(X > 700) &= P(Z > 1.19) \quad (\text{Fig. } 2.31) \\ &= 0.5 - P(0 \le Z \le 1.19) \\ &= 0.5 - 0.3830 \\ &= 0.1170 \end{aligned}$$

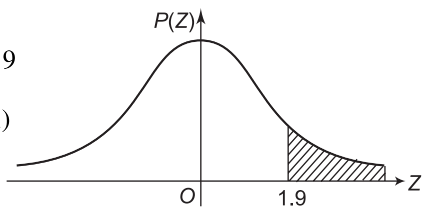


Fig. 2.31

Expected number of plots with yields over 700 kg $= N\, P(X > 700)$
$= 1000(0.1170)$
$= 117$

(ii) When $X = 650$,

$$Z = \frac{650-662}{32} = -0.38$$

$$\begin{aligned} P(X < 650) &= P(Z < -0.38) \quad (\text{Fig. } 2.32) \\ &= P(Z > 0.38) \\ &\qquad\qquad [\text{By symmetry}] \\ &= 0.5 - P(0 \le Z \le 0.38) \\ &= 0.5 - 0.1480 \\ &= 0.352 \end{aligned}$$

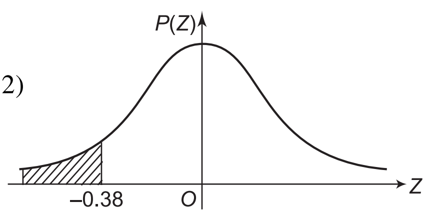


Fig. 2.32

Expected number of plots with yields below 650 kg $= N\, P(X < 650)$
$= 1000(0.352)$
$= 352$

(iii) The lowest yield, say, x_1 of the best 100 plots is given by

$$P(X > x_1) = \frac{100}{1000} = 0.1$$

When $X = x_1, \quad Z = \dfrac{x_1 - 662}{32} = z_1$

$$P(X > x_1) = P(Z > z_1)$$
$$0.1 = 0.5 - P(0 \le Z \le z_1)$$
$$P(0 \le Z \le z_1) = 0.4$$
$$\therefore \quad z_1 = 1.2 \text{ (approx.) [From normal table]}$$
$$\frac{x_1 - 662}{32} = 1.28$$
$$x_1 = 702.96$$

Hence, the best 100 plots have yields over 702.96 kg.

Example 11

Assume that the mean height of Indian soldiers is 68.22 *inches with a variance of* 10.8 *inches. How many soldiers in a regiment of* 1000 *would you expect to be over* 6 *feet tall?*

Solution

Let X be the continuous random variable which denotes the heights of Indian soldiers.

$$\mu = 68.22, \quad \sigma^2 = 10.8, \quad N = 1000$$
$$\sigma = 3.29$$
$$Z = \frac{X - \mu}{\sigma}$$

When X = 6 feet = 72 inches,

$$Z = \frac{72 - 68.22}{3.29} = 1.15$$

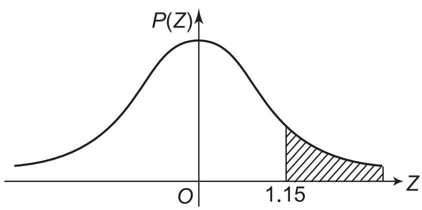


Fig. 2.33

$$P(X > 72) = P(Z > 1.15) \quad \text{(Fig. 2.33)}$$
$$= 0.5 - P(0 \le Z \le 1.15)$$
$$= 0.5 - 0.3749$$
$$= 0.1251$$

Expected number of Indian soldiers having heights over 6 feet (72 inches)

$$= N\ P(X > 72)$$
$$= 1000(0.1251)$$
$$= 125.1$$
$$\approx 125$$

Example 12

The marks obtained by students in a college are normally distributed with a mean of 65 and a variance of 25. If 3 students are selected at random from this college, what is the probability that at least one of them would have scored more than 75 marks?

Solution

Let X be the continuous random variable which denotes the marks of a student.

$$\mu = 65, \quad \sigma^2 = 25$$
$$\sigma = 5$$
$$Z = \frac{X-\mu}{\sigma}$$

When $X = 75, \quad Z = \dfrac{75-65}{5} = 2$

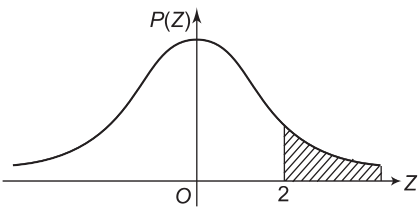


Fig. 2.34

$$P(X > 75) = P(Z > 2) \text{ (Fig. 2.34)}$$
$$= 0.5 - P(0 \le Z \le 2)$$
$$= 0.5 - 0.4772$$
$$= 0.0228$$

If p is the probability of scoring more than 75 marks,

$p = 0.0228, \; q = 1 - p = 1 - 0.0228 = 0.9772$

P(at least one student would have scored more than 75 marks)

$$= \sum_{x=1}^{3} {}^3C_x \, p^x \, q^{n-x}$$
$$= \sum_{x=1}^{3} {}^3C_x (0.0228)^x (0.9772)^{3-x}$$
$$= 0.0668$$

Example 13

Find the mean and standard deviation in which 7% of items are under 35 and 89% are under 63.

Solution

Let μ be the mean and σ be standard deviation of the normal curve.

$$P(X < 35) = 0.07$$
$$P(X < 63) = 0.89$$

$$P(X > 63) = 1 - P(X < 63) = 1 - 0.89 = 0.11$$

$$Z = \frac{X - \mu}{\sigma}$$

Since $P(X < 35) < 0.5$, the corresponding value of Z will be negative.

When $X = 35$, $Z = \dfrac{35 - \mu}{\sigma} = -z_1$ (say)

Since $P(X < 63) > 0.5$, the corresponding value of Z will be positive.

When $X = 63$, $Z = \dfrac{63 - \mu}{\sigma} = z_2$ (say)

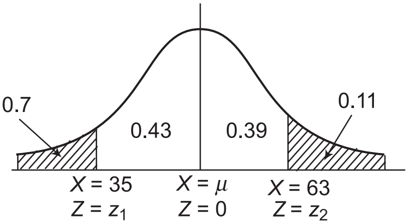


Fig. 2.35

From Fig. 2.35,

$$P(Z < -z_1) = 0.07$$

$$P(Z > z_2) = 0.11$$

$$\begin{aligned} P(0 < Z < z_1) &= P(-z_1 < Z < 0) \\ &= 0.5 - P(Z \le -z_1) \\ &= 0.5 - 0.07 \\ &= 0.43 \\ z_1 &= 1.48 \qquad \text{[From normal table]} \\ P(0 < Z < z_2) &= 0.5 - P(Z \ge z_2) \\ &= 0.5 - 0.11 \\ &= 0.39 \\ z_2 &= 1.23 \qquad \text{[From normal table]} \end{aligned}$$

Hence, $\dfrac{35 - \mu}{\sigma} = -1.48$

$$-1.48\,\sigma + \mu = 35 \qquad \text{...(1)}$$

and $\dfrac{63 - \mu}{\sigma} = 1.23$

$$1.23\,\sigma + \mu = 63 \qquad \text{...(2)}$$

Solving Eqs (1) and (2),

$$\mu = 50.29, \qquad \sigma = 10.33$$

Example 14

In an examination, it is laid down that a student passes if he secures 40 % *or more. He is placed in the first, second, and third division according to whether he secures* 60% *or more marks, between* 50% *and* 60% *marks and between* 40% *and* 50% *marks respectively. He gets a distinction in case he secures* 75% *or more. It is noticed from the result that* 10% *of* *the students failed in the examination, whereas 5% of them obtained distinction. Calculate the percentage of students placed in the second division. (Assume normal distribution of marks.)*

Solution

Let X be the random variable which denotes the marks of students in the examination. Let μ be the mean and σ be the standard deviation of the normal distribution of marks.

$$P(X < 40) = 0.10$$

$$P(X \geq 75) = 0.05$$

$$P(X < 75) = 1 - P(X \geq 75) = 1 - 0.05 = 0.95$$

$$Z = \frac{X - \mu}{\sigma}$$

Since $P(X < 40) < 0.5$, the corresponding value of Z will be negative.

When $X = 40$, $\quad Z = \dfrac{40 - \mu}{\sigma} = -z_1$ (say)

Since $P(X < 75) < 0.5$, the corresponding value of Z will be positive.

When $X = 75$, $\quad Z = \dfrac{75 - \mu}{\sigma} = z_2$ (say)

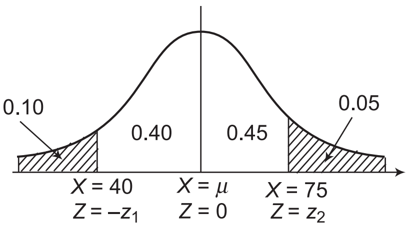


Fig. 2.36

From Fig. 2.36,

$$P(Z < -z_1) = 0.10$$

$$P(Z > z_2) = 0.05$$

$$\begin{aligned} P(0 < Z < z_1) &= P(-z_1 < Z < 0) \\ &= 0.5 - P(Z \leq -z_1) \\ &= 0.5 - 0.10 \\ &= 0.40 \end{aligned}$$

$$z_1 = 1.28 \quad \text{[From normal table]}$$

$$\begin{aligned} P(0 < Z < z_2) &= 0.5 - P(Z \geq z_2) \\ &= 0.5 - 0.05 \\ &= 0.45 \end{aligned}$$

$$z_2 = 1.64 \quad \text{[From normal table]}$$

Hence, $\quad \dfrac{40 - \mu}{\sigma} = -1.28$

$$\mu - 1.28\,\sigma = 40 \qquad \text{...(1)}$$

and $\quad \dfrac{75 - \mu}{\sigma} = 1.64$

$$\mu + 1.64\,\sigma = 75 \qquad \text{...(2)}$$

Solving Eqs (1) and (2),

$$\mu = 55.34 \approx 55$$
$$\sigma = 11.98 \approx 12$$

Probability that a student is placed in the second division is equal to the probability that his score lies between 50 and 60

When $X = 50, \quad Z = \dfrac{50-55}{12} = -0.42$

When $X = 60, \quad Z = \dfrac{60-55}{12} = 0.42$

$$\begin{aligned} P(50 < X < 60) &= P(-0.42 < Z < 0.42) \\ &= P(-0.42 < Z < 0) + P(0 < Z < 0.42) \\ &= P(0 < Z < 0.42) + P(0 < Z < 0.42) \qquad \text{[By symmetry]} \\ &= 2\,P(0 < Z < 0.42) \\ &= 2\,(0.1628) \\ &= 0.3256 \\ &\approx 0.32 \end{aligned}$$

Hence, the percentage of students placed in the second division = 32%.

2.11.5 Fitting a Normal Distribution

Fitting a normal distribution or a normal curve to the data means to find the equation of the curve in the form $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ which will be as close as possible to the points given. There are two purposes of fitting a normal curve:

(i) To judge the whether the normal curve is the best fit to the sample data.

(ii) To use the normal curve to estimate the characteristics of a population.

The area method for fitting a normal curve is given by the following steps:

(i) Find the mean μ and standard deviation σ for the given data if not given.

(ii) Write the class intervals and lower limits X of class intervals in two columns.

(iii) Find $Z = \dfrac{X-\mu}{\sigma}$ for each class interval.

(iv) Find the area corresponding to each Z from the normal table.

(v) Find the area under the normal curve between the successive values of Z. These are obtained by subtracting the successive areas when the corresponding Z's have the same sign and adding them when the corresponding Z's have the same sign and adding them when the corresponding Z's opposite sign.

(vi) Find the expected frequencies by multiplying the relative frequencies by the number of observations.

Example 1

Fit a normal curve from the following distribution. It is given that the mean of the distribution is 43.7 *and its standard distribution is* 14.8.

Class interval	11–20	21–30	31–40	41–50	51–60	61–70	71–80
Frequency	20	28	40	60	32	20	8

Solution

$\mu = 43.7, \quad \sigma = 14.8 \quad N = \Sigma f = 200$

The series is converted into an inclusive series.

Class Interval	Lower class	$Z = \dfrac{X-\mu}{\sigma}$	Area from 0 to Z	Area in class Interval	Expected Frequencies
10.5–20.5	10.5	–2.24	0.4875	0.0457	9.14 ≈ 9
20.5–30.5	20.5	–1.57	0.4418	0.1285	25.7 ≈ 26
30.5–40.5	30.5	–0.89	0.3133	0.2262	45.24 ≈ 45
40.5–50.5	40.5	–0.22	0.0871	0.2643	52.86 ≈ 53
50.5–60.5	50.5	0.46	0.1772	0.1957	39.14 ≈ 39
60.5–70.5	60.5	1.14	0.3729	0.092	18.4 ≈ 18
70.5–80.5	70.5	1.81	0.4649	0.0287	5.74 ≈ 65
	80.5	2.49	0.4936		

Example 2

Fit a normal distribution to the following data:

X	125	135	145	155	165	175	185	195	205
Y	1	1	14	22	25	19	13	3	2

It is given that $\mu = 165.5$ *and* $\sigma = 15.26$.

Solution

$\mu = 165.5, \quad \sigma = 15.26 \quad N = \Sigma f = 100$

The data is first converted into class intervals with inclusive series.

Class Interval	Lower class	$Z = \frac{X-\mu}{\sigma}$	Area from 0 to Z	Area in class Interval	Expected Frequencies
120–130	120	–2.98	0.4986	0.0085	$0.85 \approx 1$
130–140	130	–2.33	0.4901	0.0376	$3.74 \approx 4$
140–150	140	–1.67	0.4525	0.1064	$10.64 \approx 11$
150–160	150	–1.02	0.3461	0.2055	$20.55 \approx 21$
160–170	160	–0.36	0.1406	0.2547	$25.47 \approx 25$
170–180	170	0.29	0.1141	0.2148	$21.48 \approx 21$
180–190	180	0.95	0.3289	0.1174	$11.74 \approx 12$
190–200	190	1.61	0.4463	0.0418	$4.18 \approx 4$
200–210	200	2.26	0.4881	0.0101	$1.01 \approx 1$
210–220	210	2.92	0.4982		

EXERCISE 2.7

1. If X is normally distributed with a mean and standard deviation of 4, find (i) $P(5 \le X \le 10)$, (ii) $P(X \ge 15)$, (iii) $P(10 \le X \le 15)$, and (iv) $P(X \le 5)$.

[**Ans.:** (i) 0.3345 (ii) 0.003 (iii) 0.0638 (iv) 0.4013]

2. A normal distribution has a mean of 5 and a standard deviation of 3. What is the probability that the deviation from the mean of an item taken at random will be negative?

[**Ans.:** 0.0575]

3. If X is a normal variate with a mean of 30 and an SD of 6, find the value of $X = x_1$ such that $P(X \ge x_1) = 0.05$.

[**Ans.:** 39.84]

4. If X is a normal variate with a mean of 25 and SD of 5, find the value of $X = x_1$ such that $P(X \le x_1) = 0.01$.

[**Ans.:** 11.02]

5. The weights of 4000 students are found to be normally distributed with a mean of 50 kg and an SD of 5 kg. Find the probability that a student selected at random will have weight (i) less than 45 kg, and (ii) between 45 and 60 kg.

[**Ans.:** (i) 0.1587 (ii) 0.8185]

6. The daily sales of a firm are normally distributed with a mean of ₹ 8000 and a variance of ₹ 10000. (i) What is the probability that on a certain day the sales will be less than ₹ 8210? (ii) What is the percentage of days on which the sales will be between ₹ 8100 and ₹ 8200?

[**Ans.:** (i) 0.482 (ii) 14%]

7. The mean height of Indian soldiers is 68.22″ with a variance of 10.8″. Find the expected number of soldiers in a regiment of 1000 whose height will be more than 6 feet.

[**Ans.:** 125]

8. The life of army shoes is normally distributed with a mean of 8 months and a standard deviation of 2 months. If 5000 pairs are issued, how many pairs would be expected to need replacement after 12 months?

[**Ans.:** 2386]

9. In an intelligence test administered to 1000 students, the average was 42 and the standard deviation was 24. Find the number of students (i) exceeding 50, (ii) between 30 and 54, and (iii) the least score of top 1000 students.

[**Ans.:** (i) 129 (ii) 383 (iii) 72.72]

10. In a test of 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average of life of 2040 hours and a standard deviation of 60 hours. Estimate the number of bulbs likely to burn for (i) more than 2150 hours, and (ii) less than 1950 hours.

[**Ans.:** (i) 67 (ii) 184]

11. The marks of 1000 students of a university are found to be normally distributed with a mean of 70 and a standard of deviation 5. Estimate the number of students whose marks will be (i) between 60 and 75, (ii) more than 75, and (iii) less than 68.

[**Ans.:** (i) 910 (ii) 23 (iii) 37]

12. In a normal distribution, 31% items are under 45 and 8% are over 64. Find the mean and standard deviation. Find also, the percentage of items lying between 30 and 75.

[**Ans.:** 50, 10, 0.957]

13. Of a large group of men, 5% are under 60 inches in height and 40% are between 60 and 65 inches. Assuming a normal distribution, find the mean and standard deviation of distribution.

[**Ans.:** 65.42, 3.27]

14. The marks obtained by students in an examination follow a normal distribution. If 30% of the students got marks below 35 and 10% got marks above 60, find the mean and percentage of students who got marks between 40 and 50.

[**Ans.:** 42.23, 13.88, 28%]

15. Fit a normal distribution to the following data:

Class	60–65	65–70	70–75	75–80	80–85	85–90	90–95	95–100
Frequency	3	21	150	335	326	135	26	4

[**Ans.:** Expected frequency: 3, 31, 148, 322, 319, 144, 30, 3]

Points to Remember

Random Variables

A random variable X is a real-valued function of the elements of the sample space of a random experiment. In other words, a variable which takes the real values, depending on the outcome of a random experiment is called a *random variable*,

Discrete Random Variables: A random variable X is said to be discrete if it takes either finite or countably infinite values.

Continuous Random Variables: A random variable X is said to be continuous if it takes any values in a given interval.

Discrete Probability Distribution

Probability distribution of a random variable is the set of its possible values together with their respective probabilities.

Discrete Distribution Function

$$F(x) = P(X \le x) = \sum_{i=1}^{x} p(x_i)$$

Measures of Central Tendency for Discrete Probability Distribution

1. Mean

$$\mu = E(X) = \sum_{i=1}^{\infty} x_i \; p(x_i) = \sum x \; p(x)$$

2. Variance

$$\begin{aligned} \text{Var}(X) &= \sigma^2 = E(X-\mu)^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

3. Standard deviation

$$\text{SD} = \sigma = \sqrt{\sum_{i=1}^{\infty} x_i^2 \; p(x_i) - \mu^2}$$

$$= \sqrt{E\,(X^2) - \mu^2}$$

$$= \sqrt{E\,(X^2) - [E(X)]^2}$$

Continuous Distribution Function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x, \quad -\infty < x < \infty$$

Measures of Central Tendency for Continuous Probability Distribution

1. Mean

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x$$

2. Median

$$\int_{a}^{M} f(x)\,\mathrm{d}x = \int_{M}^{b} f(x)\,\mathrm{d}x = \frac{1}{2}$$

3. Variance

$$\text{Var}\,(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \; f(x)\,\mathrm{d}x$$

$$= \int_{-\infty}^{\infty} x^2 \; f(x)\,\mathrm{d}x - \mu^2$$

3. Standard Deviation

$$\text{SD} = \sqrt{\text{Var}(X)} = \sigma$$

Binomial Distribution

$$P(X = x) = p(x) = {}^nC_x \; p^x \; q^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

1. Mean of the Binomial Distribution

$$E(X) = np$$

2. Variance of the Binomial Distribution

$$\text{Var}(X) = npq$$

3. Standard Deviation of the Binomial Distribution

$$\text{SD} = \sqrt{\text{Variance}} = \sqrt{npq}$$

Recurrence Relation for the Binomial Distribution

$$P(X = x+1) = \frac{n-x}{x+1} \cdot \frac{p}{q} \cdot P(X = x)$$

Binomial Frequency Distribution

$$\sum_{x=0}^{n} f(x) = N \sum_{x=0}^{n} P(X = x) = N$$

Poisson Distribution

$$P(X = x) = p(x) = \frac{e^{-\lambda}\,\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

1. Mean of the Poisson Distribution

$$E(X) = \lambda$$

2. Variance of the Poisson Distribution

$$\text{Var}\,(X) = \lambda$$

3. Standard Deviation of the Poisson Distribution

$$\text{SD} = \sqrt{\text{Variance}} = \sqrt{\lambda}$$

Recurrence Relation for the Poisson Distribution

$$p(x+1) = \frac{\lambda}{x+1}\, p(x)$$

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < X < \infty, -\infty < \mu < \infty, \sigma > 0$$

1. Mean of the Normal Distribution

$$E(X) = \mu$$

2. Variance of the Normal Distribution

$$\text{Var}(X) = \sigma^2$$

3. Standard Deviation of the Normal Distribution

$$\text{SD} = \sigma$$

5. Median of the Normal Distribution

$$\mu = M$$

CHAPTER 3

Statistics

Chapter Outline

3.1 INTRODUCTION

Statistics is the science which deals with the collection, presentation, analysis, and interpretation of numerical data. Statistics should possess the following characteristics:

(i) Statistics are aggregates of facts.
(ii) Statistics are affected by a large number of causes.
(iii) Statistics are always numerically expressed.
(iv) Statistics should be enumerated or estimated.
(v) Statistics should be collected in a systematic manner.
(vi) Statistics should be collected for a pre-determined purpose.
(vii) Statistics should be placed in relation to each other.

The use of statistical methods help in presenting a complex mass of data in a simplified form so as to facilitate the process of comparison of characteristics in two or more situations. Statistics also provide important techniques for the study of relationship between two or more characteristics (or variables) in forecasting, testing of hypothesis, quality control, decision making, etc.

3.2 MEASURES OF CENTRAL TENDENCY

Summarization of data is a necessary function of any statistical analysis. The data is summarized in the form of tables and frequency distributions. In order to bring the characteristics of the data, these tables and frequency distributions need to be summarized further. A measure of central tendency or an average is very essential and an important summary measure in any statistical analysis.

An *average* is a single value which can be taken as a representative of the whole distribution. There are five types of measures of central tendency or averages which are commonly used.

(i) Arithmetic mean
(ii) Median
(iii) Mode
(iv) Geometric mean
(v) Harmonic mean

A good measure of average must have the following characteristics:

(i) It should be rigidly defined so that different persons obtain the same value for a given set of data.
(ii) It should be easy to understand and easy to calculate.
(iii) It should be based on all the observations of the data.
(iv) It should be easily subjected to further mathematical calculations.
(v) It should not be much affected by the fluctuations of sampling.
(vi) It should not be unduly affected by extreme observations.
(vii) It should be easy to interpret.

3.3 ARITHMETIC MEAN

The *arithmetic mean* of a set of observations is their sum divided by the number of observations. Let $x_1, x_2, ..., x_n$ be n observations. Then their average or arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}$$

For example, the marks obtained by 10 students in Class XII in a physics examination are 25, 30, 21, 55, 40, 45, 17, 48, 35, 42. The arithmetic mean of the marks is given by

$$\bar{x} = \frac{\sum x}{n} = \frac{25 + 30 + 21 + 55 + 40 + 45 + 17 + 48 + 35 + 42}{10} = \frac{358}{10} = 35.8$$

If n observations consist of n distinct values denoted by $x_1, x_2, ..., x_n$ of the observed variable x occurring with frequencies $f_1, f_2, ..., f_n$ respectively then the arithmetic mean is given by

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_ix_i}{N} = \frac{\sum fx}{N}$$

where $\quad N = \sum_{i=1}^{n} f_i = f_1 + f_2 + \cdots + f_n$

3.3.1 Arithmetic Mean of Grouped Data

In case of grouped or continuous frequency distribution the arithmetic mean is given by

$$\overline{x} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} = \frac{\sum fx}{N}, \text{ where } N = \sum_{i=1}^{n} f_i$$

and x is taken as the midvalue of the corresponding class.

Example 1

Find the arithmetic mean from the following frequency distribution:

x	5	6	7	8	9	10	11	12	13	14
f	25	45	90	165	112	96	81	26	18	12

Solution

x	f	fx
5	25	125
6	45	270
7	90	630
8	165	1320
9	112	1008
10	96	960
11	81	891
12	26	312
13	18	234
14	12	168
	$\sum f = 670$	$\sum fx = 5918$

$$N = \sum f = 670$$

$$\overline{x} = \frac{\sum fx}{N} = \frac{5918}{670} = 8.83$$

Example 2

Find the arithmetic mean of the marks from the following data:

Marks	0–10	10–20	20–30	30–40	40–50	50–60
Number of students	12	18	27	20	15	8

Solution

Marks	Number of students (f)	Midvalue (x)	fx
0–10	12	5	60
10–20	18	15	270
20–30	27	25	675
30–40	20	35	700
40–50	15	45	675
50–60	8	55	440
	$\sum f = 100$		$\sum fx = 2820$

$$N = \sum f = 100$$

$$\overline{x} = \frac{\sum fx}{N} = \frac{2820}{100} = 28.20$$

Example 3

A company is planning to improve plant safety. For this, accident data for the last 50 weeks was compiled. These data are grouped into the frequency distribution as shown below. Calculate the arithmetic mean of the number of accidents per week.

Number of accidents (x)	0–4	5–9	10–14	15–19	20–24
Number of weeks (f)	5	22	13	8	2

Solution

The given class intervals are inclusive. However, they need not be converted into exclusive class intervals for the calculation of mean.

Number of Accidents	Number of Weeks	Midvalue (x)	fx
0–4	5	2	10
5–9	22	7	154
10–14	13	12	156
15–19	8	17	136
20–24	2	22	44
	$\sum f = 50$		$\sum fx = 500$

$$N = \sum f = 50$$

$$\bar{x} = \frac{\sum fx}{N} = \frac{500}{50} = 10$$

3.3.2 Arithmetic Mean from Assumed Mean

If the values of x and (or) f are large, the calculation of mean becomes quite time-consuming and tedious. In such cases, the provisional man 'a' is taken as that value of x (midvalue of the class interval) which corresponds to the highest frequency or which comes near the middle value of the frequency distribution. This number is called the *assumed mean*.

Let $\quad d = x - a$

$$fd = f(x-a) = fx - af$$

$$\sum fd = \sum fx - a\sum f$$

$$= \sum fx - aN$$

Dividing both the sides by n,

$$\frac{\sum fd}{N} = \frac{\sum fx}{N} - a$$

$$= \bar{x} - a$$

$$\therefore \quad \bar{x} = a + \frac{\sum fd}{N}$$

Example 1

Ten coins were tossed together and the number of tails resulting from them were observed. The operation was performed 1050 *times and the frequencies thus obtained for different number of tail* (x) *are shown in the following table. Calculate the arithmetic mean.*

x	0	1	2	3	4	5	6	7	8	9	10
y	2	8	43	133	207	260	213	120	54	9	1

Solution

Let $a = 5$ be the assumed mean.

$d = x - a = x - 5$

x	f	$d = x - 5$	fd
0	2	–5	–10
1	8	–4	–32
2	43	–3	–129
3	133	–2	–266
4	207	–1	–207
5	260	0	0
6	213	1	213
7	120	2	240
8	54	3	162
9	9	4	36
10	1	5	5
	$\sum f = 1050$		$\sum fd = 12$

$$N = \sum f = 1050$$

$$\bar{x} = a + \frac{\sum fd}{N}$$

$$= 5 + \frac{12}{1050}$$

$$= 5.0114$$

Example 2

The daily earnings (in rupees) of employees working on a daily basis in a firm are

Daily earnings (₹)	100	120	140	160	180	200	220
Number of employees	3	6	10	15	24	42	75

Calculate the mean of daily earnings.

Solution

Let $a = 160$ be the assumed mean.

$d = x - a = x - 160$

Daily Earnings x	Number of Employees f	$d = x - 160$	fd
100	3	–60	–180
120	6	–40	–240
140	10	–20	–200
160	15	0	0
180	24	20	480
200	42	40	1680
220	75	60	4500
	$\sum f = 175$		$\sum fd = 6040$

$$N = \sum f = 175$$

$$\begin{aligned}\overline{x} &= a + \frac{\sum fd}{N} \\ &= 160 + \frac{6040}{175} \\ &= 194.51\end{aligned}$$

Example 3

Calculate the mean for the following frequency distribution

Class	0–8	8–16	16–24	24–32	32–40	40–48
Frequency	8	7	16	24	15	7

Solution

Let $a = 28$ be the assumed mean.

$d = x - a = x - 28$

Class	Frequency	Midvalue (x)	$d = x - 28$	fd
0–8	8	4	–24	–192
8–16	7	12	–16	–112
16–24	16	20	–8	–128
24–32	24	28	0	0
32–40	15	36	8	120
40–48	7	44	16	112
	$\sum f = 77$			$\sum fd = -200$

$$N = \sum f = 77$$

$$\bar{x} = a + \frac{\sum fd}{N}$$

$$= 28 + \frac{(-200)}{77}$$

$$= 25 \cdot 403$$

Example 4

Calculate the arithmetic mean of the following distribution:

Class Interval	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
Frequency	3	8	12	15	18	16	11	5

Solution

Let $a = 35$ be the assumed mean.

$d = x - a = x - 35$

Class Interval	Frequency f	Midvalue x	$d = x - 35$	fd
0–10	3	5	–30	–90
10–20	8	15	–20	–160
20–30	12	25	–10	–120
30–40	15	35	0	0
40–50	18	45	10	180
50–60	16	55	20	320
60–70	11	65	30	330
70–80	5	75	40	200
	$\sum f = 88$			$\sum fd = 660$

$$N = \sum f = 88$$

$$\bar{x} = a + \frac{\sum fd}{N}$$

$$= 35 + \frac{660}{88}$$

$$= 42.5$$

3.3.3 Arithmetic Mean by the Step-Deviation Method

When the class intervals in a grouped data are equal, calculation can be simplified by the step-deviation method. In such cases, deviation of variate x from the assumed mean

a (i.e., $d = x - a$) are divided by the common factor h which is equal to the width of the class interval.

Let $$d = \frac{x-a}{h}$$

$$\bar{x} = a + h\frac{\sum fd}{\sum f} = a + h\frac{\sum fd}{N}$$

where a is the assumed mean

$d = \dfrac{x-a}{h}$ is the deviation of any variate x from a

h is the width of the class interval

N is the number of observations

Example 1

Calculate the arithmetic mean of the following marks obtained by students in mathematics:

Marks (x)	5	10	15	20	25	30	35	40	45	50
Number of students (f)	20	43	75	67	72	45	39	9	8	6

Solution

Let $a = 30$ be the assumed mean and $h = 5$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-30}{5}$$

x	f	$d = \frac{x-30}{5}$	fd
5	20	–5	–100
10	43	–4	–172
15	75	–3	–225
20	67	–2	–134
25	72	–1	–72
30	45	0	0
35	39	1	39
40	9	2	18
45	8	3	24
50	6	4	24
	$\sum f = 384$		$\sum fd = -598$

$$N = \sum f = 384$$

$$\bar{x} = a + h\frac{\sum fd}{N}$$

$$= 30 + 5\left(\frac{-598}{384}\right)$$

$$= 22.214$$

Example 2

Calculate the average overtime work done per employee for the following distribution which gives the pattern of overtime work done by 100 employees of a company.

Overtime hours	10–15	15–20	20–25	25–30	30–35	35–40
Number of employees	11	20	35	20	8	6

Solution

Let $a = 22.5$ be the assumed mean and $h = 5$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-22.5}{5}$$

Overtime hours	Number of employees f	Midvalue x	$d = \frac{x-22.5}{5}$	fd
10–15	11	12.5	–2	–22
15–20	20	17.5	–1	–20
20–25	35	22.5	0	0
25–30	20	27.5	1	20
30–35	8	32.5	2	16
35–40	6	37.5	3	18
	$\Sigma f = 100$			$\Sigma fd = 12$

$$N = \sum f = 100$$

$$\bar{x} = a + h\frac{\sum fd}{N}$$

$$= 22.5 + 5\left(\frac{12}{100}\right)$$

$$= 23.1 \text{ hours}$$

Example 3

The following table gives the distribution of companies according to size of capital. Find the mean size of the capital of a company.

Capital (₹ in lacs)	<5	<10	<15	<20	<25	<30
No. of companies	20	27	29	38	48	53

Solution

This is a 'less than' type of frequency distribution. This will be first converted into class intervals. Let $a = 12.5$ be the assumed mean and $h = 5$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-12.5}{5}$$

Class intervals	Frequency f	Midvalue x	$d = \frac{x-12.5}{5}$	fd
0–5	20	2.5	–2	–40
5–10	7	7.5	–1	–7
10–15	2	12.5	0	0
15–20	9	17.5	1	9
20–25	10	22.5	2	20
25–30	5	27.5	3	15
	$\sum f = 53$			$\sum fd = -3$

$$N = \sum f = 53$$

$$\bar{x} = a + h\frac{\sum fd}{N}$$

$$= 12.5 + 5\left(\frac{-3}{53}\right)$$

$$= 12.22 \text{ lacs}$$

Example 4

Find the arithmetic mean from the following data:

Marks less than	10	20	30	40	50	60
No. of students	10	30	60	110	150	180

Solution

This is a 'less than' type of frequency distribution. This will be first converted into class intervals. Let a = 45 be the assumed mean and h = 10 be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-45}{10}$$

Marks	No. of students f	Midvalue x	$d = \frac{x-45}{10}$	fd
0–10	10	5	–4	–40
10–20	20	15	–3	–60
20–30	30	25	–2	–60
30–40	50	35	–1	–50
40–50	40	45	0	0
50–60	30	55	1	30
	$\sum f = 180$			$\sum fd = -180$

$$N = \sum f = 180$$

$$\bar{x} = a + h\frac{\sum fd}{N}$$

$$= 45 + 10\left(\frac{-180}{180}\right)$$

$$= 35$$

Example 5

Following is the distribution of marks obtained by 60 students in a mathematics test:

Marks	Number of students
More than 0	60
More than 10	56
More than 20	40
More than 30	20
More than 40	10
More than 50	3

Calculate the arithmetic mean.

Solution

This is a 'more than' type of frequency distribution. This will be first converted into class intervals. Let $a = 35$ be the assumed mean and $h = 10$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-35}{10}$$

Marks	No. of students f	Midvalue x	$d = \dfrac{x-35}{10}$	fd
0–10	4	5	–3	–12
10–20	16	15	–2	–32
20–30	20	25	–1	–20
30–40	10	35	0	0
40–50	7	45	1	7
50–60	3	55	2	6
	$\sum f = 60$			$\sum fd = -51$

$$N = \sum f = 60$$

$$\begin{aligned} \bar{x} &= a + h\frac{\sum fd}{N} \\ &= 35 + 10\left(\frac{-51}{60}\right) \\ &= 26.5 \end{aligned}$$

Example 6

Find the average marks of students from the following table:

Marks	No. of students	Marks	No. of students
Above 0	80	Above 60	23
Above 10	77	Above 70	16
Above 20	72	Above 80	10
Above 30	65	Above 90	8
Above 40	55	Above 100	0
Above 50	43		

Solution

This is a 'more than' type of frequency distribution. This will be first converted into class intervals. Let $a = 45$ be the assumed mean and $h = 10$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-45}{10}$$

Marks	No. of students f	Midvalue x	$d = \frac{x-45}{10}$	fd
0–10	3	5	–4	–12
10–20	5	15	–3	–15
20–30	7	25	–2	–14
30–40	10	35	–1	–10
40–50	12	45	0	0
50–60	20	55	1	20
60–70	7	65	2	14
70–80	6	75	3	18
80–90	2	85	4	8
90–100	8	95	5	40
	$\sum f = 80$			$\sum fd = 49$

$$N = \sum f = 80$$

$$\begin{aligned} \overline{x} &= a + h\frac{\sum fd}{N} \\ &= 45 + 10\left(\frac{49}{80}\right) \\ &= 51.125 \end{aligned}$$

Example 7

Find the arithmetic mean of the following data obtained during study on patients:

Age (in years)	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89
No. of cases	1	0	1	10	17	38	9	3

Solution

This is an inclusive series. The inclusive series can be converted into exclusive series by subtracting half the difference between the upper limit of a class and the lower limit of the next class from the lower limit of the class and adding the same to the upper limit of the class. Let a = 44.5 be the assumed mean and h = 10 be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-44.5}{10}$$

Age (in years)	No. of cases f	Midvalue x	$d = \frac{x - 44.5}{10}$	fd
9.5–19.5	1	14.5	–3	–3
19.5–29.5	0	24.5	–2	0
29.5–39.5	1	34.5	–1	–1
39.5–49.5	10	44.5	0	0
49.5–59.5	17	54.5	1	17
59.5–69.5	38	64.5	2	26
69.5–79.5	9	74.5	3	27
79.5–89.5	3	84.5	4	12
	$\sum f = 79$			$\sum fd = 128$

$$N = \sum f = 79$$

$$\begin{aligned}\bar{x} &= a + h\frac{\sum fd}{N} \\ &= 44.5 + 10\left(\frac{128}{79}\right) \\ &= 60.7\end{aligned}$$

3.3.4 Weighted Arithmetic Mean

In the calculation of arithmetic mean, equal importance is given to all the items. If all the items are not of equal importance, a simple arithmetic mean will not be a good representative of the given data. In such a case, proper weightage is to be given to various items. The weights are assigned to different items depending upon their importance, i.e., more important items are assigned more weight.

If $w_1, w_2, ..., w_n$ are the weights assigned to the values $x_1, x_2, ..., x_n$ respectively then the weighted arithmetic mean is given by

$$\text{Weighted arithmetic mean} = \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n}$$

$$\bar{x}_w = \frac{\sum wx}{\sum w}$$

When the assumed mean is used for calculation,

$$\bar{x}_w = a + \frac{\sum wd}{\sum w}$$

When the step-deviation method is used for calculation,

$$\bar{x}_w = a + h\frac{\sum wd}{\sum w}$$

Example 1

A candidate obtains the following percentages in an examination: English—46%, Mathematics—67%, Physics—72%, Chemistry—58%. It is agreed to give double weights to marks in English and Mathematics compared to other subjects. What is the weighted mean?

Solution

Subjects	Marks x	Weights w	wx
English	46	2	92
Mathematics	67	2	134
Physics	72	1	72
Chemistry	58	1	58
		$\sum w = 6$	$\sum wx = 356$

Weighted mean $\bar{x}_w = \dfrac{\sum wx}{\sum w} = \dfrac{356}{6} = 59.33$

Example 2

Comment on the performance of the students of two universities given below:

	GTU		Mumbai University	
Course	Pass %	No. of Students (in hundreds)	Pass %	No. of Students (in hundreds)
BE (Computer)	65	200	60	190
BE (Civil)	75	150	80	120
BE (IT)	55	180	60	130
BE (Mechanical)	60	130	65	150

Solution

Course	GTU			Mumbai University		
	Pass % x	No. of Students (in hundreds) w	wx	Pass % x	No. of Students (in hundreds) w	wx
BE (Computer)	65	200	13000	60	190	9000
BE (Civil)	75	150	11250	80	120	9600
BE (IT)	55	180	9900	60	130	7800
BE (Mechanical)	60	130	7800	65	150	9750
		$\sum w = 660$	$\sum wx = 41950$		$\sum w = 590$	$\sum wx = 36150$

$$\text{GTU} : \bar{x}_w = \frac{\sum wx}{\sum w} = \frac{41950}{660} = 63.56$$

$$\text{Mumbai University} : \bar{x}_w = \frac{\sum wx}{\sum w} = \frac{36150}{590} = 61.27$$

Hence, the performance of students of GTU is better than that of Mumbai University.

3.3.5 Properties of Arithmetic Mean

1. The algebraic sum of deviations from the mean is zero. If the mean of n observations $x_1, x_2, ..., x_n$ is $\bar{x}$ then $(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x}) = 0$, i.e., $\sum (x - \bar{x}) = 0$.
2. The sum of squares of the deviations is minimum when taken about the mean.
3. If $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ are the means of k series of sizes $n_1, n_2, ..., n_k$ respectively then the mean $\bar{x}$ of the composite series is given by

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

$$= \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i}$$

Example 1

Find the value of p for the following distribution whose mean is 11.37.

x	5	7	p	11	13	16	20
f	2	4	29	54	11	8	4

Solution

$$\bar{x} = \frac{\sum fx}{\sum f}$$

$$11.37 = \frac{(5\times 2)+(7\times 4)+29p+(11\times 54)+(13\times 11)+(16\times 8)+(20\times 4)}{2+4+29+54+11+8+4}$$

$$= \frac{10+28+29p+594+143+128+80}{112}$$

$$= \frac{983+29p}{112}$$

$$\therefore \quad p = 10.015 \approx 10$$

Example 2

The following is the distribution of weights (in lbs) of 60 *students of a class:*

Weight	No. of students	Weight	No. of students
93–97	2	113–117	14
98–102	5	118–122	?
103–107	12	123–127	3
108–112	?	128–132	1

If the mean weight of the students is 110.917, *find the missing frequencies.*

Solution

Let f_1 be the frequency of the class 108–112.

The frequency of the class 118–122 = 60 – (2 + 5 +12 + 14 + 3 + 1 + f_1)

= 23 – f_1

Let a = 110 be the assumed mean and h = 5 be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-110}{5}$$

Weights (lbs)	No. of students	Midvalue x	$d = \frac{x-110}{5}$	fd
93–97	2	95	–3	–6
98–102	5	100	–2	–10
103–107	12	105	–1	–12
108–112	f_1	110	0	0
113–117	14	115	1	14
118–122	$23 - f_1$	120	2	$46 - 2f_1$
123–127	3	125	3	9
128–132	1	130	4	4
	$\sum f = 60$			$\sum fd = 45 - 2f_1$

$$N = \sum f = 60$$

$$\bar{x} = a + h\frac{\sum fd}{N}$$

$$110.917 = 110 + 5\left(\frac{45 - 2f_1}{60}\right)$$

$$f_1 = 17$$

Hence, the frequency of the class 108–112 is 17 and the frequency of the class 118–122 is 23–17 = 6.

Example 3

The average salary of male employees in a company is ₹ 5200 *and that of females is* ₹ 4200. *The mean salary of all the employees is* ₹ 5000. *Find the percentage of male and female employees.*

Solution

Let $\bar{x}_1$ and $\bar{x}_2$ be the average salary of male and female employees respectively. Let n_1 and n_2 be the number of male and female employees in the company respectively. Let $\bar{x}$ be the average salary of all the employees in the company.

$$\bar{x}_1 = 5200, \bar{x}_2 = 4200, \bar{x} = 5000$$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$5000 = \frac{5200\, n_1 + 4200\, n_2}{n_1 + n_2}$$

$$5000(n_1+n_2)=5200n_1+4200n_2$$
$$200n_1=800n_2$$
$$\frac{n_1}{n_2}=\frac{4}{1}$$

Hence, the percentage of male employees = $\frac{4}{4+1}\times 100=80$

and the percentage of female employees = $\frac{1}{4+1}\times 100=20$

Example 4

There are 50 *students in a class of which* 40 *are boys and the rest girls. The average weight of the class is* 44 *kg and the average weight of the girls is* 40 *kg. Find the average weight of the boys.*

Solution

Let $\bar{x}_1$ and $\bar{x}_2$ be the average weight of boys and girls respectively. Let n_1 and n_2 be the number of boys and girls in the class respectively. Let $\bar{x}$ be the average weight of all the boys and girls in the class.

$$n_1=40,\quad n_2=10,\quad \bar{x}=44,\quad \bar{x}_2=40$$
$$\bar{x}=\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}$$
$$44=\frac{40\bar{x}_1+(10\times 40)}{40+10}$$
$$\therefore\quad \bar{x}_1=45$$

Example 5

The mean marks scored by 100 *students was found to be* 40. *Later on, it was discovered that a score of* 53 *was misread as* 83. *Find the correct mean.*

Solution

$$\bar{x}=40,\quad n=100$$
$$\bar{x}=\frac{\sum x}{n}$$
$$40=\frac{\sum x}{100}$$

$$\sum x = 4000$$

Incorrect $\sum x = 4000$

Correct $\sum x$ = Incorrect $\sum x$ – Incorrect Item + Correct Item

$= 4000 - 83 + 53 = 3970$

$$\therefore \qquad \text{Correct mean} = \frac{\text{Correct} \sum x}{n} = \frac{3970}{100} = 397$$

EXERCISE 3.1

1. Find the mean of the following marks obtained by students of a class:

Marks	15	20	25	30	35	40
No. of Students	9	7	12	14	15	6

[**Ans.:** 25.58]

2. The following table gives the distribution of total household expenditure (in rupees) of manual workers in a city:

Expenditure (in ₹)	100–150	150–200	200–250	250–300	300–350	350–400	400–450	450–500
Frequency	24	40	33	28	30	22	16	7

Find the average expenditure (in ₹) per household.

[**Ans.:** ₹ 266.25]

3. Calculate the mean for the following data:

Heights (in cm)	135–140	140–145	145–150	150–155	155–160	160–165	165–170	170–175
No. of boys	4	9	18	28	24	10	5	2

[**Ans.:** 153.45 cm]

4. The weights in kilograms of 60 workers in a factory are given below. Find the mean weight of a worker.

Weight (in kg)	60	61	62	63	64	65
No. of workers	5	8	14	16	10	7

[**Ans.:** 62.65 kg]

5. Calculate the mean from the following data:

Marks less than/up to	10	20	30	40	50	60
No. of students	10	30	60	110	150	180

[**Ans.:** 35]

6. Calculate the mean from the following data:

Marks more than	0	10	20	30	40	50	60
No. of students	180	170	150	120	70	30	0

[**Ans.:** 35]

7. Calculate the mean from the following data:

Marks	1–5	6–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45
No. of students	7	10	16	30	24	17	10	5	1

[**Ans.:** 20.33]

8. Find the missing frequency p for the following distribution whose mean is 50.

x	10	30	50	70	90
f	17	p	32	24	19

[**Ans.:** 28]

9. Find the missing value of p for the following distribution whose mean is 12.58.

x	5	8	10	12	p	20	25
f	2	5	8	22	7	4	2

[**Ans.:** 15]

10. The mean of the following frequency table is 50. But the frequencies f_1 and f_2 in the classes 20–40 and 60–80 are missing. Find the missing frequencies.

Class	0–20	20–40	40–60	60–80	80–100	Total
Frequency	17	f_1	32	f_2	19	120

[**Ans.:** $f_1 = 28$, $f_2 = 24$]

11. 100 students appeared in an examination. The result of the examination is as under:

Marks	4	5	6	7	8	9
No. of students	16	20	18	12	8	6

If the combined mean of the marks obtained by 100 students is 5.16, calculate the combined mean of the marks obtained by all the unsuccessful students.

[**Ans.:** 2.1]

12. Ram purchased equity shares of a company in 4 successive months. Find the average price per share.

Month	Price per share (in ₹)	No. of shares
Dec. 14	100	200
Jan. 15	150	250
Feb. 15	200	280
March 15	125	300

[**Ans.:** ₹ 146.60]

13. From the following results of two colleges *A* and *B*, find out which of the two is better.

Examination	College A		College B	
	Appeared	Passed	Appeard	Passed
M.Sc.	60	40	200	160
M.A.	100	60	240	200
B.Sc.	200	150	200	140
B.A.	120	75	160	100

[**Ans.:** College *B* is better]

14. The average rainfall for a week, excluding Sunday, was 10 cm. Due to heavy rainfall on Sunday, the average for the week rose to 15 cm. How much rainfall was recorded on Sunday?

[**Ans.:** 45 cm]

15. The arithmetic mean of 50 items of a series was calculated by a student as 20. However, it was later discovered that an item of 25 was misread as 35. Find the correct value of mean.

[**Ans.:** 19.8]

16. The sales of a balloon seller on seven days of a week are as given below:

Days	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Sales (in ₹)	100	150	125	140	160	200	250

If the profit is 20% of sales, find the average profit per day.

[**Ans.:** ₹ 32.14]

3.4 MEDIAN

Median is the central value of the variable when the values are arranged in ascending or descending order of magnitude. It divides the distribution into two equal parts. When the observations are arranged in the order of their size, median is the value of that item which has equal number of observations on either side.

In case of ungrouped data, if the number of observations is odd then the median is the middle value after the values have been arranged in ascending or descending order of magnitude. If the number of observations is even, there are two middle terms and the median is obtained by taking the arithmetic mean of the middle terms.

Examples

(i) The median of the values 20, 15, 25, 28, 18, 16, 30, i.e., 15, 16, 18, 20, 25, 28, 30 is 20 because $n = 7$, i.e., odd and the median is the middle value, i.e., 20.

(ii) The median of the values 8, 20, 50, 25, 15, 30, i.e., 8, 15, 20, 25, 30, 50 is the arithmetic mean of the middle terms, i.e., $\frac{20+25}{2} = 22.5$ because $n = 6$, i.e., even.

In case of discrete frequency distribution, the median is obtained by considering the cumulative frequencies. The steps for calculating the median are given below:

(i) Arrange the values of the variables in ascending or descending order of magnitudes.

(ii) Find $\frac{N}{2}$ where $N = \sum f$

(iii) Find the cumulative frequency just greater than $\frac{N}{2}$ and determine the corresponding value of the variable.

(iv) The corresponding value of x is the median.

Example 1

The following table represents the marks obtained by a batch of 12 *students in certain class tests in physics and chemistry.*

Marks (Physics)	53	54	32	30	60	46	28	25	48	72	33	65
Marks (Chemistry)	55	41	48	49	27	25	23	20	28	60	43	67

Indicate the subject in which the level of achievement is higher.

Solution

The level of achievement is higher in that subject for which the median marks are more. Arranging the marks in two subjects in ascending order,

Marks (Physics)	25	28	30	32	33	46	48	53	54	60	65	72
Marks (Chemistry)	20	23	25	27	28	41	43	48	49	55	60	67

Since the number of students is 12, the median is the arithmetic mean of the middle terms.

Median marks in Physics $= \dfrac{46+48}{2} = 47$

Median marks in Chemistry $= \dfrac{41+43}{2} = 42$

Since the median marks in physics are greater than the median marks in chemistry, the level of achievement is higher in physics.

Example 2

Obtain the median for the following frequency distribution.

x	0	1	2	3	4	5	6	7
f	7	14	18	36	51	54	52	18

Solution

x	f	Cumulative Frequency
0	7	7
1	14	21
2	18	39
3	36	75
4	51	126
5	54	180
6	52	232
7	18	250

$$N = 250$$

$$\frac{N}{2} = \frac{250}{2} = 125$$

The cumulative frequency just greater than $\frac{N}{2}$ = 125 is 126 and the value of x corresponding to 126 is 4. Hence, the median is 4.

Example 3

Find the median of the following distribution:

x	5	7	9	12	14	17	19	21
f	6	5	3	6	5	3	2	3

Solution

x	f	Cumulative Frequency
5	6	6
7	5	11
9	3	14
12	6	20
14	5	25
17	3	28
19	2	30
21	3	33

$$N = 33$$

$$\frac{N}{2} = \frac{33}{2} = 16.5$$

The cumulative frequency just greater than $\frac{N}{2}$ = 16.5 is 20 and the value of x corresponding to 20 is 12. Hence, the median is 12.

Median for Continuous Frequency Distribution

In case of continuous frequency distribution (less than frequency distribution), the class corresponding to the cumulative frequency just greater than $\frac{N}{2}$, is called the *median class*, and the value of the median is given by

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

where l is the lower limit of the median class

f is the frequency of the median class

h is the width of the median class

c is the cumulative frequency of the class preceding the median class

N is sum of frequencies, i.e., $N = \sum f$

In case of 'more than' or 'greater than' type of frequency distributions, the value of the median is given by

$$\text{Median} = u - \frac{h}{f}\left(\frac{N}{2} - c\right)$$

where u is the upper limit of the median class
f is the frequency of the median class
h is the width of the median class
c is the cumulative frequency of the class succeeding the median class

Example 1

The following table gives the weekly expenditures of 100 *workers. Find the median weekly expenditure.*

Weekly Expenditure (in ₹)	0–10	10–20	20–30	30–40	40–50
Number of workers	14	23	27	21	15

Solution

Weekly Expenditure (in ₹)	Number of Workers i.e., frequency (f)	Cumulative Frequency
0–10	14	14
10–20	23	37
20–30	27	64
30–40	21	85
40–50	15	100

$$N = 100$$

$$\frac{N}{2} = \frac{100}{2} = 50$$

The cumulative frequency just greater than $\frac{N}{2} = 50$ is 64 and the corresponding class 20–30 is the median class.

Here, $\frac{N}{2} = 50, \quad l = 20, \quad h = 10, \quad f = 27, \quad c = 37$

$$\begin{aligned}\text{Median} &= l + \frac{h}{f}\left(\frac{N}{2} - c\right)\\ &= 20 + \frac{10}{27}(50 - 37)\\ &= 24.815\end{aligned}$$

Example 2

From the following data, calculate the median:

Marks (Less than)	5	10	15	20	25	30	35	40	45
No. of Students	29	224	465	582	634	644	650	653	655

[Summer 2015]

Solution

This is a 'less than' type of frequency distribution. This will be first converted into class intervals.

Class Intervals	Frequency	Less than CF
0–5	29	29
5–10	195	224
10–15	241	465
15–20	117	582
20–25	52	634
25–30	10	644
30–35	6	650
35–40	3	653
40–45	2	655

$$N = 655$$

Since $\frac{N}{2} = \frac{655}{2} = 327.5,$ the median class is 10–15.

Here, $l = 10, \quad h = 5, \quad f = 241, \quad c = 224$

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$= 10 + \frac{5}{241}(327.5 - 224)$$

$$= 12.147$$

Example 3

Find the mean of the following data:

Age greater than (in years)	0	10	20	30	40	50	60	70
No. of Persons	230	218	200	165	123	73	28	8

Solution

This is a 'greater than' type of frequency distribution. This will be first converted into class intervals.

Class Intervals	Frequency	Greater than CF
0–10	12	230
10–20	18	218
20–30	35	200
30–40	42	165
40–50	50	123
50–60	45	73
60–70	20	28
70 and above	8	8

$N = 230$

Since $\frac{N}{2} = \frac{230}{2} = 115,$ the median class is 40–50.

Here, $u = 50,\quad h = 10,\quad f = 50,\quad c = 73$

$$\text{Median} = u - \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$= 50 - \frac{10}{50}(115 - 73)$$

$$= 41.6 \text{ years}$$

Example 4

The following table gives the marks obtained by 50 students in mathematics. Find the median.

Marks	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49
No. of Students	4	6	10	5	7	3	9	6

Solution

Since the class intervals are inclusive, it is necessary to convert them into exclusive series.

Marks	No. of Students	Cumulative Frequency
9.5–14.5	4	4
14.5–19.5	6	10
19.5–24.5	10	20
24.5–29.5	5	25
29.5–34.5	7	32
34.5–39.5	3	35
39.5–44.5	9	44
44.5–49.5	6	50

$N = 50$

Since $\dfrac{N}{2} = \dfrac{50}{2} = 25,$ the median class is 24.5–29.5.

Here, $l = 24.5, \quad h = 5, \quad f = 5, \quad c = 20$

$$\begin{aligned}\text{Median } &= l + \frac{h}{f}\left(\frac{N}{2} - c\right)\\ &= 24.5 + \frac{5}{5}(25 - 20)\\ &= 29.5\end{aligned}$$

Example 5

Find the median of the following distribution:

Midvalues	1500	2500	3500	4500	5500	6500	7500
Frequency	27	32	65	78	58	32	8

Solution

The difference between two midvalues is 1000. On subtracting and adding half of this, i.e., 500 to each of the midvalues, the lower and upper limits of the respective class intervals are obtained.

Class Intervals	Frequency	Cumulative Frequency
1000–2000	27	27
2000–3000	32	59
3000–4000	65	124
4000–5000	78	202
5000–6000	58	260
6000–7000	32	292
7000–8000	8	300

$N = 300$

Since $\frac{N}{2} = 150$, the median class is 4000–5000.

Here, $l = 4000$, $h = 1000$, $f = 78$, $c = 124$

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$= 4000 + \frac{1000}{78}(150 - 124)$$

$$= 4333.33$$

Example 6

The following table gives the distribution of daily wages of 900 workers. However, the frequencies of classes 40–50 and 60–70 are missing. If the median of the distribution is ₹ 59.25, find the missing frequencies.

Wages (in ₹)	30–40	40–50	50–60	60–70	70–80
No. of workers	120	?	200	?	185

Solution

Let f_1 and f_2 be the frequencies of the classes 40–50 and 60–70 respectively.

$$f_1 + f_2 = 900 - (120 + 200 + 185) = 395$$

Class Intervals	Frequency	Less than CF
30–40	120	120
40–50	f_1	$120 + f_1$
50–60	200	$320 + f_1$
60–70	f_2	$320 + f_1 + f_2$
70–80	185	900

$$N = 900$$

Since the median is 59.25, the median class is 50–60.

Here, $\frac{N}{2} = 450$, $l = 50$, $h = 10$, $f = 200$, $c = 120 + f_1$

$$\text{Median } = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$59.25 = 50 + \frac{10}{200}\left[450 - (120 + f_1)\right]$$

$$\therefore \quad f_1 = 145$$

$$f_2 = 395 - 145 = 250$$

EXERCISE 3.2

1. The heights (in cm) of 15 students of Class XII are 152, 147, 156, 149, 151, 159, 148, 160, 153, 154, 150, 143, 155, 157, 161. Find the median.

[**Ans.:** 153 cm]

2. The median of the following observations are arranged in the ascending order: 11, 12, 14, 18, x + 2, x + 4, 30, 32, 35, 41 is 24. Find x.

[**Ans.:** 21]

3. Find the median of the following frequency distribution:

x	10	11	12	13	14	15	16
f	8	15	25	20	12	10	5

[**Ans.:** 12]

4. Find the median of the following frequency distribution:

Wages (in ₹)	20–30	30–40	40–50	50–60	60–70
No. of workers	3	5	20	10	5

[**Ans.:** 46.75]

5. Calculate the median of the following data:

x	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11
f	3	7	12	16	22	20	13	7

[**Ans.:** 7.55]

6. The weekly wages of 1000 workers of a factory are shown in the following table:

Weekly wages (less than)	425	475	525	575	625	675	725	775	825	875
No. of Workers	2	10	43	123	293	506	719	864	955	1000

[**Ans.:** 673.59]

7. Calculate the mean of the following distribution of marks obtained by 50 students in advanced engineering mathematics.

Marks more than	0	10	20	30	40	50
No. of Students	50	46	40	20	10	3

[**Ans.:** 27.5]

8. Calculate the median from the following data:

Midvalues	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

[**Ans.:** 153.79]

9. The following incomplete table gives the number of students in different age groups of a town. If the median of the distribution is 11 years, find out the missing frequencies.

Age group	0–5	5–10	10–15	15–20	20–25	25–30	Total
No. of Students	15	125	?	66	?	4	300

[**Ans.:** 50, 40]

10. An incomplete frequency distribution is given as follows:

Variable	10–20	20–30	30–40	40–50	50–60	60–70	70–80	Total
Frequency	12	30	?	65	?	25	18	229

Given that the median value is 46, calculate the missing frequencies.

[**Ans.:** 34, 45]

3.5 MODE

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set are heavily distributed. In other words, mode is the value of the variable which is most frequent or predominant in the series. In case of a discrete frequency distribution, mode is the value of x corresponding to the maximum frequency.

Examples

(i) In the series 6, 5, 3, 4, 3, 7, 8, 5, 9, 5, 4, the value 5 occurs most frequently. Hence, the mode is 5.

(ii) Consider the following frequency distribution:

x	1	2	3	4	5	6	7	8
f	4	9	16	25	22	15	7	3

The value of x corresponding to the maximum frequency, viz., 25, is 4. Hence, the mode is 4.

For an asymmetrical frequency distribution, the difference between the mean and the mode is approximately three times the difference between the mean and the median.

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

This is known as the *empirical formula for calculation of the mode*.

3.5.1 Mode by Method of Grouping

The grouping method is used when the frequency distribution is not regular. In such cases, the difference between the maximum frequency and the frequency preceding or succeeding it is very small, and the items are heavily concentrated on either side. In such cases, the value of the mode is determined by preparing a grouping table and an analysis table.

Grouping Table

A grouping table has the following six columns:

Column I: This column has original frequencies.

Column II: In this column, the frequencies of Column I are combined 'two by two'.

Column III: In this column, the first frequency of Column I is left out and the remaining frequencies of Column I are combined 'two by two'.

Column IV: In this column, the frequencies of Column I are combined 'three by three'.

Column V: In this column, the first frequency of Column I is left out and the remaining frequencies of Column I are combined 'three by three'.

Column VI: In this column, the first two frequencies of Column I are left out and the remaining frequencies of Column I are combined 'three by three'.

The maximum frequency in each column is identified and circled.

Analysis Table

After preparing the grouping table, an analysis table is prepared. The column numbers are put on the left-hand side and various probable values of the mode, i.e., x are put on the right-hand side. The values against which frequencies are marked maximum in the grouping table are entered in the analysis table corresponding to the values they represent.

Example 1

Calculate the mode from the following frequency distribution:

x	50	51	52	53	54	55	56	57	58	59	60
f	2	4	5	6	8	5	4	7	11	5	3

Solution

Since the frequency distribution is not regular, the method of grouping is used for calculation of the mode.

Grouping Table

x	(I) f	(II) Column of two	(III) Column of two leaving out the first	(IV) Column of three	(V) Column of three leaving out the first	(VI) Column of three leaving out the first two
50	2	6		11		
51	4		9		15	
52	5	11				19
53	6		14	19		
54	8	13			17	
55	5		9			16
56	4	11		22		
57	7		18		23	
58	11	16				19
59	5		8			
60	3					

Analysis Table

Columns ↓ / $x \rightarrow$	50	51	52	53	54	55	56	57	58	59	60
I									1		
II									1	1	
III								1	1		
IV							1	1	1		
V								1	1	1	
VI			1	1	1				1	1	1
$\sum$	0	0	1	1	1	0	1	3	6	3	1

Since the value 58 has occurred the maximum number of times, the mode of the frequency distribution is 58.

Example 2

Find the mode of the following frequency distribution:

x	10	11	12	13	14	15	16	17	18	19
f	8	15	20	100	98	95	90	75	50	30

Solution

Since the frequency distribution is not regular, the method of grouping is used for calculation of the mode.

Grouping Table

x	(I) f	(II) Column of two	(III) Column of two leaving out the first	(IV) Column of three	(V) Column of three leaving out the first	(VI) Column of three leaving out the first two
10	8	23		43		
11	15		35		135	
12	20	120				218
13	100		198	293		
14	98	193			283	
15	95		185			260
16	90	165		215		
17	75		125		155	
18	50	80				
19	30					

Analysis Table

Columns ↓ / $x \rightarrow$	10	11	12	13	14	15	16	17	18	19
I				1						
II					1	1				
III				1	1					
IV				1	1	1				
V					1	1	1			
VI						1	1	1		
Σ	0	0	0	3	4	4	2	1	0	0

Since the values 14 and 15 have occurred the maximum number of times, the mode is ill-defined. In such a case, mode = 3 median – 2 mean.

Calculation of Mean and Median

x	f	fx	CF
10	8	80	8
11	15	165	23
12	20	240	43
13	100	1300	143
14	98	1372	241
15	95	1425	336
16	90	1440	426
17	75	1275	501
18	50	900	551
19	30	570	581
	$\sum f = 581$	$\sum fx = 8767$	

$$N = \sum f = 581$$

$$\bar{x} = \frac{\sum fx}{N} = \frac{8767}{581} = 15.09$$

$$\frac{N}{2} = \frac{581}{2} = 290.5$$

The cumulative frequency just greater than $\frac{N}{2} = 290.5$ is 95 and the value of x corresponding to 95 is 15. Hence, the median is 15.

$$\text{Mode} = 3 \times 15 - 2 \times 15.09 = 14.82$$

3.5.2 Mode for a Continuous Frequency Distribution

In case of a continuous frequency distribution, the class in which the mode lies is called the *modal class* and the value of the mode is given by

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

where l is the lower limit of the modal class

h is the width of the modal class

f_m is the frequency of the modal class

f_1 is the frequency of the class preceding the modal class

f_2 is the frequency of the class succeeding the modal class

This method of finding mode is called the *method of interpolation*. This formula is applicable only to a unimodal frequency distribution.

Example 1

Find the mode for the following data:

Profit per shop	0–100	100–200	200–300	300–400	400–500	500–600
No. of Shops	12	18	27	20	17	6

Solution

Since the maximum frequency is 27 which lies in the class 200–300, the modal class is 200–300.

Here, $l = 200$, $h = 100$, $f_m = 27$, $f_1 = 18$, $f_2 = 20$

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

$$= 200 + 100\left[\frac{27-18}{2(27)-18-20}\right]$$

$$= 256.25$$

Example 2

The frequency distribution of marks obtained by 60 students of a class in a college is given by

Marks	30–34	35–39	40–44	45–49	50–54	55–59	60–64
Frequency	3	5	12	18	14	6	2

Find the mode of the distribution.

Solution

The class intervals are first converted into a continuous exclusive series as shown in the following table:

Marks	Frequency
29.5–34.5	3
34.5–39.5	5
39.5–44.5	12
44.5–49.5	18
49.5–54.5	14
54.5–59.5	6
59.5–64.5	2

Since the maximum frequency is 18 which lies in the interval 44.5–49.5, the modal class is 44.5–49.5.

Here, $l = 44.5$, $h = 5$, $f_m = 18$, $f_1 = 12$, $f_2 = 14$

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

$$= 44.5 + 5\left[\frac{18-12}{2(18)-12-14}\right]$$

$$= 47.5$$

Example 3

Calculate the mode of the following distribution:

Midvalues	5	15	25	35	45	55	65	75
Frequency	7	15	18	30	31	4	3	1

Solution

Since the frequency distribution is not regular, the method of grouping is used for calculation of the mode.

The difference between the two midvalues is 10. On subtracting and adding half of this, i.e., 5, to each of the midvalues, the lower and upper limits of the respective class intervals are obtained.

Grouping Table

Class Intervals	I f	II Column of two	III Column of two leaving out the first	IV Column of three	V Column of three leaving out the first	VI Column of three leaving out the first two
0–10	7	22 (0–20)		40 (0–30)		
10–20	15		33 (10–30)		63 (10–40)	
20–30	18	48 (20–40)				79 (20–50)
30–40	30		61 (30–50)	65 (30–60)		
40–50	31	35 (40–60)			38 (40–70)	
50–60	4		7 (50–70)			8 (50–80)
60–70	3	4 (60–80)				
70–80	1					

Analysis Table

Columns ↓	$x \rightarrow$ 10–20	20–30	30–40	40–50	50–60
I				1	
II		1	1		
III			1	1	
IV			1	1	1
V	1	1	1		
VI		1	1	1	
	1	3	5	4	1

From the analysis table, the modal class is 30–40.

Here, $l = 30$, $h = 10$, $f_m = 30$, $f_1 = 18$, $f_2 = 31$

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

$$= 30 + 10\left[\frac{30 - 18}{2(30) - 18 - 31}\right]$$

$$= 40.91$$

Example 4

Find the mode for the following distribution:

Class intervals	0–10	10–20	20–30	30–40	40–50
Frequency	45	20	14	7	3

Solution

Since the highest frequency occurs in the first class interval, the interpolation formula is not applicable. Thus, empirical formula is used for calculation of mode.

Class intervals	Frequency	CF	Midvalue	$d = \frac{x - 25}{10}$	fd
0–10	45	45	5	–2	–90
10–20	20	65	15	–1	–20
20–30	14	79	25	0	0
30–40	7	86	35	1	7
40–50	3	89	45	2	6
	$\sum f = 89$				$\sum fd = -97$

$$N = \sum f = 89$$

Since $\frac{N}{2} = \frac{89}{2} = 44.5$, the median class is 0–10.

Here, $l = 0$, $h = 10$, $f = 45$, $c = 0$

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$= 0 + \frac{10}{45}(44.5 - 0)$$

$$= 9.89$$

$$\text{Mean} = a + h\frac{\sum fd}{N}$$

$$= 25 + 10\left(\frac{-97}{89}\right)$$

$$= 14.1$$

$$\text{Hence, mode} = 3\,\text{Median} - 2\,\text{Mean}$$

$$= 3(9.89) - 2(14.1)$$

$$= 1.47$$

Example 5

The following table gives the incomplete income distribution of 300 *workers of a company, where the frequencies of the classes* 3000–4000 *and* 5000–6000 *are missing. If the mode of the distribution is* ₹ 4428.57, *find the missing frequencies.*

Monthly Income (in ₹)	No. of Workers
1000–2000	30
2000–3000	35
3000–4000	?
4000–5000	75
5000–6000	?
6000–7000	30
7000–8000	15

Solution

Let f_1 and f_2 be the frequencies of the classes 3000–4000 and 5000–6000 respectively.

$$f_1 + f_2 = 300 - (30 + 35 + 75 + 30 + 15) = 115$$

Since the mode is 4428.57, the modal class is 4000–5000.

Here, $l = 4000$, $h = 1000$, $f_m = 75$

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

$$= l + h\left[\frac{f_m - f_1}{2f_m - (f_1 + f_2)}\right]$$

$$4428.57 = 4000 + 1000\left[\frac{75 - f_1}{2(75) - 115}\right]$$

$$\therefore \quad f_1 = 60$$

$$f_2 = 115 - 60 = 55$$

EXERCISE 3.3

1. Calculate the mode for the following distribution:

x	6	12	18	24	30	36
f	12	24	36	38	37	6

[Ans.: 24**]**

2. Calculate the mode for the following distribution:

x	10	20	30	40	50	60	70
f	17	22	31	39	27	15	13

[Ans.: 40**]**

3. Calculate the mode for the following distribution:

Class interval	0–4	4–8	8–12	12–16
Frequency	4	8	5	6

[Ans.: 6.28**]**

4. Calculate the mode of the following distribution:

x	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45
f	20	24	32	28	20	16	37	10	18

[Ans.: 13.33**]**

5. Calculate the mode for the following data:

Class	10–20	20–30	30–40	40–50	50–60	60–70	70–80
f	24	42	56	66	108	130	154

[**Ans.:** 71.348]

6. Find the mode of the following distribution:

Class	55–64	65–74	75–84	85–94	95–104	105–114	115–124	125–134	135–144
f	1	2	9	22	33	22	8	2	1

[**Ans.:** 99.5]

7. Calculate the modal marks from the following distribution of marks of 100 students of a class:

Marks (more than)	90	80	70	60	50	40	30	20	10
No. of Students	0	4	15	33	53	76	92	98	100

[**Ans.:** 47]

8. If the mode and mean of a moderately asymmetrical series are 80 and 68, what will be the most probable median?

[**Ans.:** 72]

9. Calculate the mode from the following data:

Midpoint	1	2	3	4	5	6	7	8
Frequency	5	50	45	30	20	10	15	5

[**Ans.:** 2.875]

10. Calculate the mode from the following series:

Class intervals	10–19	20–29	30–39	40–49	50–59	60–69
Frequency	4	6	8	5	4	2

[**Ans.:** 33.5]

11. Calculate the mode from the following distribution:

Marks (less than)	7	14	21	28	35	42	49
No. of Students	20	25	33	41	45	50	52

[**Ans.:** 11.26]

12. Calculate the mode from the following distribution:

Class intervals	6–10	11–15	16–20	21–25	26–30
Frequency	20	30	50	40	10

[**Ans.:** 18.83]

3.6 GEOMETRIC MEAN

The *geometric mean* of a set of n observations is the n^{th} root of their product. If there are n observations, $x_1, x_2, .., x_n$ such that $x_i > 0$ for each i, their geometric mean GM is given by

$$\text{GM} = (x_1 \cdot x_2 \cdot \cdots \cdot x_n)^{\frac{1}{n}}$$

The n^{th} root is calculated with the help of logarithms. Taking logarithms of both the sides,

$$\begin{aligned}\log \text{GM} &= \log(x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}} \\ &= \frac{1}{n}\log(x_1 \cdot x_2 \cdots x_n) \\ &= \frac{1}{n}(\log x_1 + \log x_2 + \cdots + \log x_n) \\ &= \frac{\sum \log x}{n}\end{aligned}$$

$$\text{GM} = \text{antilog}\left(\frac{\sum \log x}{n}\right)$$

In case of a frequency distribution consisting of n observations $x_1, x_2, .., x_n$ with respective frequencies $f_1, f_2, .., f_n$, the geometric mean is given by

$$\text{GM} = (x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n})^{\frac{1}{N}}, \text{ where } N = \sum f$$

Taking logarithms of both the sides,

$$\begin{aligned}\log \text{GM} &= \frac{1}{N}(f_1 \log x_1 + f_2 \log x_2 + \cdots + f_n \log x_n) \\ &= \frac{\sum f \log x}{N}\end{aligned}$$

$$\text{GM} = \text{antilog}\left(\frac{\sum f \log x}{N}\right)$$

Thus, the geometric mean is the antilog of the weighted mean of the different values of log x_i whose weights are their frequencies f_i.

In case of a continuous or grouped frequency distribution, x is taken to be the value corresponding to the midpoints of the class intervals.

Example 1

Calculate the geometric mean of the following data:

10, 110, 120, 50, 52, 80

Solution

$$\begin{aligned}
\text{GM} &= \text{antilog}\left(\frac{\sum \log x}{n}\right) \\
&= \text{antilog}\left(\frac{\log 10 + \log 110 + \log 120 + \log 50 + \log 52 + \log 80}{6}\right) \\
&= \text{antilog}\left(\frac{2.3026 + 4.7005 + 4.7875 + 3.9120 + 3.9512 + 4.3820}{6}\right) \\
&= \text{antilog}\ (4.006) \\
&= 54.9267
\end{aligned}$$

Example 2

Find the geometric mean of the following data:

x	5	10	15	20	25	30
f	13	18	50	40	10	6

Solution

x	f	$\log x$	$f \log x$
5	13	1.6094	20.9227
10	18	2.3026	41.4465
15	50	2.7081	135.4025
20	40	2.9957	119.8293
25	10	3.2189	32.1888
30	6	3.4012	20.4072
	$\sum f = 137$		$\sum f \log x = 370.197$

$$N = \sum f = 137$$

$$\text{GM} = \text{antilog}\left(\frac{\sum f \log x}{N}\right)$$
$$= \text{antilog}\left(\frac{370.197}{137}\right)$$
$$= 14.912$$

Example 3

Find the geometric mean of the following data:

Marks	0–10	10–20	20–30	30–40
No. of Students	5	8	3	4

Solution

Marks	No. of Students f	Midvalue x	$\log x$	$f \log x$
0–10	5	5	1.6094	8.047
10–20	8	15	2.7081	21.6648
20–30	3	25	3.2189	9.6567
30–40	4	35	3.5553	14.2212
	$\sum f = 20$			$\sum f \log x = 53.5897$

$$N = \sum f = 20$$
$$\text{GM} = \text{antilog}\left(\frac{\sum f \log x}{N}\right)$$
$$= \text{antilog}\left(\frac{53.5897}{20}\right)$$
$$= 14.5776$$

3.7 HARMONIC MEAN

The *harmonic mean* of a number of observations, none of which is zero, is the reciprocal of the arithmetic mean of the reciprocals of the given values.

The harmonic mean of n observations $x_1, x_2, ..., x_n$ is given by

$$\text{HM} = \frac{1}{\frac{1}{n}\sum\left(\frac{1}{x}\right)}$$

$$= \frac{1}{\frac{1}{n}\left(\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}\right)}$$

$$= \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$$

For example, the harmonic mean of 2, 4 and 5 is

$$\text{HM} = \frac{3}{\frac{1}{2}+\frac{1}{4}+\frac{1}{5}} = 3.16$$

In case of a frequency distribution consisting of n observations $x_1, x_2, ..., x_n$ with respective frequencies $f_1, f_2, ..., f_n$, the harmonic mean is given by

$$\text{HM} = \frac{f_1+f_2+\cdots+f_n}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\cdots+\frac{f_n}{x_n}}$$

$$= \frac{\sum f}{\sum\left(\frac{f}{x}\right)}$$

If $x_1, x_2, ..., x_n$ are n observations with weights $w_1, w_2, ..., w_n$ respectively, their weighted harmonic mean is given by

$$\text{HM} = \frac{\sum w}{\sum\left(\frac{w}{x}\right)}$$

Example 1

Calculate the harmonic mean of the following data:

x	20	21	22	23	24	25
f	4	2	7	1	3	1

Solution

x	f	$\frac{f}{x}$
20	4	0.2
21	2	0.095
22	7	0.318
23	1	0.043
24	3	0.125
25	1	0.04
	$\sum f = 18$	$\sum\left(\frac{f}{x}\right) = 0.821$

$$\text{HM} = \frac{\sum f}{\sum\left(\frac{f}{x}\right)} = \frac{18}{0.821} = 21.924$$

Example 2

Find the harmonic mean of the following distribution:

Class interval	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
Frequency	5	8	11	21	35	30	22	18

Solution

Class Interval	Frequency f	Midvalue x	$\frac{f}{x}$
0–10	5	5	1
10–20	8	15	0.533
20–30	11	25	0.44
30–40	21	35	0.6
40–50	35	45	0.778
50–60	30	55	0.545
60–70	22	65	0.338
70–80	18	75	0.24
	$\sum f = 150$		$\sum\left(\frac{f}{x}\right) = 4.474$

$$\text{HM} = \frac{\sum f}{\sum\left(\frac{f}{x}\right)} = \frac{150}{4.474} = 33.527$$

Relation between Arithmetic Mean, Geometric Mean, and Harmonic Mean

The arithmetic mean (AM), geometric mean (GM), and harmonic mean (HM) for a given set of observations of a series are related as

$$\text{AM} \geq \text{GM} \geq \text{HM}$$

For two observations x_1 and x_2 of a series,

$$\text{AM} = \frac{x_1 + x_2}{2}$$

$$\text{GM} = \sqrt{x_1 \ x_2}$$

$$\text{HM} = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} = \frac{2x_1 \ x_2}{x_1 + x_2}$$

$$\text{AM} \cdot \text{HM} = \left(\frac{x_1 + x_2}{2}\right)\left(\frac{2x_1 \ x_2}{x_1 + x_2}\right) = x_1 \ x_2 = (\text{GM})^2$$

$$\therefore \qquad \text{GM} = \sqrt{\text{AM} \cdot \text{HM}}$$

Example 1

If the AM of two observations is 15 *and their GM is* 9, *find their HM and the two observations.*

Solution

$$\text{GM} = \sqrt{\text{AM} \cdot \text{HM}}$$

$$9 = \sqrt{15 \times \text{HM}}$$

$$\therefore \quad \text{HM} = 5.4$$

Let the two observations be x_1 and x_2.

$$\text{AM} = \frac{x_1 + x_2}{2} = 15$$

$$x_1 + x_2 = 30 \qquad \text{...(1)}$$

$$\text{GM} = \sqrt{x_1 \ x_2} = 9$$

$$x_1 \ x_2 = 81 \qquad \text{...(2)}$$

Solving Eqs (1) and (2),

$$x_1 = 27, x_2 = 3$$

EXERCISE 3.4

1. Calculate the geometric and harmonic means of the following series of monthly expenditure of a batch of students:

₹	125	130	75	10	45	0.5	0.4	500	1505

[Ans.: ₹ 22.98, ₹ 2.06]

2. Calculate the geometric mean of the following distribution:

Class intervals	5–15	15–25	25–35	35–45	45–55
Frequency	10	22	25	20	8

[Ans.: 26.65]

3. Calculate the harmonic mean of the following data:

x	10	11	12	13	14
f	5	8	10	9	6

[Ans.: 11.94]

4. An investor buys ₹ 1200 worth of shares in a company each month. During the first 5 months, he bought the shares at a price of ₹ 10, ₹ 12, ₹ 15, ₹ 20, ₹ 24 per share. After 5 months, what is the average price paid for the shares?

[Ans.: ₹ 14.63]

5. Calculate the geometric mean of the following distribution:

Marks (less than)	10	20	30	40	50
No. of Students	12	27	72	93	100

[Ans.: 21.35]

6. Calculate the GM and HM for the following data:

Class intervals	5–15	15–25	25–35	35–45	45–55
Frequency	6	9	15	8	4

[Ans.: 26.07, 22.92]

3.8 STANDARD DEVIATION

Standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by the Greek letter σ. Let X be a random variable which takes on values, viz., $x_1, x_2, ..., x_n$. The standard deviation of these n observations is given by

$$\sigma = \sqrt{\frac{\sum (x-\overline{x})^2}{n}}$$

where $\overline{x} = \frac{\sum x}{n}$ is the arithmetic mean of these observations.

This equation can be modified further.

$$\sigma = \sqrt{\frac{\sum (x^2 - 2x\,\overline{x} + \overline{x})^2}{n}}$$

$$= \sqrt{\frac{\sum x^2 - 2\overline{x}\sum x + \overline{x}^2 \sum 1}{n}}$$

$$= \sqrt{\frac{\sum x^2}{n} - 2\frac{\sum x}{n}\frac{\sum x}{n} + \left(\frac{\sum x}{n}\right)^2 \cdot \frac{n}{n}} \qquad \left[\because \; \sum 1 = n\right]$$

$$= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \sqrt{\text{Mean of squares} - \text{Square of mean}}$$

In case of a frequency distribution consisting of n observations $x_1, x_2, ..., x_n$ with respective frequencies $f_1, f_2, ..., f_n$, the standard deviation is given by

$$\sigma = \sqrt{\frac{\sum f(x-\overline{x})^2}{N}}$$

This equation can also be modified.

$$\sigma = \sqrt{\frac{\sum f\,(x^2 - 2x\overline{x} + \overline{x}^2)}{N}}$$

$$= \sqrt{\frac{\sum fx^2}{N} - \frac{2\overline{x}\sum fx}{N} + \overline{x}^2\frac{\sum f}{N}}$$

$$= \sqrt{\frac{\sum fx^2}{N} - 2\frac{\sum fx}{N}\frac{\sum fx}{N} + \left(\frac{\sum fx}{N}\right)^2} \qquad \left[\because \; \sum f = N \text{ and } \overline{x} = \frac{\sum fx}{N}\right]$$

$$= \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2}$$

3.8.1 Variance

The *variance* is the square of the standard deviation and is denoted by σ^2. The method for calculating variance is same as that given for the standard deviation.

Example 1

Calculate the standard deviation of the weights of ten persons.

Weight (in kg)	45	49	55	50	41	44	60	58	53	55

Solution

$$n = 10$$

$$\sum x = 45 + 49 + 55 + 50 + 41 + 44 + 60 + 58 + 53 + 55 = 510$$

$$\sum x^2 = 45^2 + 49^2 + 55^2 + 50^2 + 41^2 + 44^2 + 60^2 + 58^2 + 53^2 + 55^2 = 26366$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \sqrt{\frac{26366}{10} - \left(\frac{510}{10}\right)^2}$$

$$= 5.967$$

Aliter:

$$\overline{x} = \frac{\sum x}{n} = \frac{510}{10} = 51$$

x	$x - \overline{x}$	$(x - \overline{x})^2$
45	–6	36
49	–2	4
55	4	16
50	–1	1
41	–10	100
44	–7	49
60	9	81
58	7	49
53	2	4
55	4	16
		$\sum(x - \overline{x})^2 = 356$

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}$$
$$= \sqrt{\frac{356}{10}}$$
$$= 5.967$$

Example 2

Calculate the standard deviation of the following data:

x	10	11	12	13	14	15	16	17	18
f	2	7	10	12	15	11	10	6	3

Solution

x	f	fx	x^2	fx^2
10	2	20	100	200
11	7	77	121	847
12	10	120	144	1440
13	12	156	169	2028
14	15	210	196	2940
15	11	165	225	2475
16	10	160	256	2560
17	6	102	289	1734
18	3	54	324	972
	$\sum f = 76$	$\sum fx = 1064$		$\sum fx^2 = 15196$

$$N = \sum f = 76$$
$$\sigma = \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2}$$
$$= \sqrt{\frac{15196}{76} - \left(\frac{1064}{76}\right)^2}$$
$$= 1.987$$

Aliter:

$$N = \sum f = 76$$
$$\bar{x} = \frac{\sum fx}{N} = \frac{1064}{76} = 14$$

x	f	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
10	2	–4	16	32
11	7	–3	9	63
12	10	–2	4	40
13	12	–1	1	12
14	15	0	0	0
15	11	1	1	11
16	10	2	4	40
17	6	3	9	54
18	3	4	16	48
				$\sum f(x-\bar{x})^2 = 300$

$$\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$$

$$= \sqrt{\frac{300}{76}}$$

$$= 1.987$$

3.8.2 Standard Deviation from the Assumed Mean

If the values of x and f are large, the calculation of fx, fx^2 becomes tedious. In such a case, the assumed mean a is taken to simplify the calculation.

Let a be the assumed mean.

$$d = x - a$$

$$x = a + d$$

$$\sum fx = \sum f(a+d) = Na + \sum fd$$

Dividing both the sides by N,

$$\frac{\sum fx}{N} = a + \frac{\sum fd}{N}$$

$$\bar{x} = a + \bar{d}$$

$$x - \bar{x} = d - \bar{d}$$

$$\sigma_x = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$$

$$= \sqrt{\frac{\sum f(d-\bar{d})^2}{N}}$$

$$= \sigma_d$$

Hence, the standard deviation is independent of change of origin.

$$\therefore \qquad \sigma_x = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

Example 1

Find the standard deviation from the following data:

Size of the item	10	11	12	13	14	15	16
Frequency	2	7	11	15	10	4	1

Solution

Let $a = 13$ be the assumed mean.

$$d = x - a = x - 13$$

Size of item (x)	Frequency (f)	$d = x - a$	d^2	fd	fd^2
10	2	–3	9	–6	18
11	7	–2	4	–14	28
12	11	–1	1	–11	11
13	15	0	0	0	0
14	10	1	1	10	10
15	4	2	4	8	16
16	1	3	9	3	9
	$\sum f = 50$			$\sum fd = -10$	$\sum fd^2 = 92$

$$N = \sum f = 50$$

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= \sqrt{\frac{92}{50} - \left(\frac{-10}{50}\right)^2}$$

$$= 1.342$$

3.8.3 Standard Deviation by Step-Deviation Method

Let a be the assumed mean and h be the width of the class interval.

$$d = \frac{x - a}{N}$$

$$x = a + hd$$

$$\sum fx = \sum f\,(a + hd) = Na + h\sum fd$$

Dividing both the sides by N,

$$\frac{\sum fx}{N} = a + h\frac{\sum fd}{N}$$

$$\bar{x} = a + h\bar{d}$$

$$x - \bar{x} = h(d - \bar{d})$$

$$\sigma_x = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$$

$$= \sqrt{\frac{\sum f\,h^2\,(d-\bar{d})^2}{N}}$$

$$= h\sqrt{\frac{\sum f\,(d-\bar{d})^2}{N}}$$

$$= h\,\sigma_d$$

Hence, the standard deviation is independent of change of origin but not of scale.

$$\therefore \qquad \sigma_x = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

Example 1

Find the standard deviation for the following distribution:

Marks	10–20	20–30	30–40	40–50	50–60	60–70	70–80
Number of Students	5	12	15	20	10	4	2

Solution

Let $a = 45$ be the assumed mean and $h = 10$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-45}{10}$$

Marks	Number of students f	Midvalue x	$d = \frac{x-45}{10}$	d^2	fd	fd^2
10–20	5	15	–3	9	–15	45
20–30	12	25	–2	4	–24	48
30–40	15	35	–1	1	–15	15
40–50	20	45	0	0	0	0
50–60	10	55	1	1	10	10
60–70	4	65	2	4	8	16
70–80	2	75	3	9	6	18
	$\sum f = 68$				$\sum fd = -30$	$\sum fd^2 = 152$

$$N = \sum f = 68$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= 10\sqrt{\frac{152}{68} - \left(\frac{-30}{68}\right)^2}$$

$$= 14.285$$

Example 2

Find the standard deviation for the following data:

Class interval	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	6	14	10	8	1	3	8

Solution

Let $a = 35$ be the assumed mean and $h = 10$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-35}{10}$$

Class interval	f	Midvalue x	$d = \frac{x-35}{10}$	d^2	fd	fd^2
0–10	6	5	–3	9	–18	54
10–20	14	15	–2	4	–28	56
20–30	10	25	–1	1	–10	10
30–40	8	35	0	0	0	0
40–50	1	45	1	1	1	1
50–60	3	55	2	4	6	12
60–70	8	65	3	9	24	72
	$\sum f = 50$				$\sum fd = -25$	$\sum fd^2 = 205$

$$N = \sum f = 50$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= 10\sqrt{\frac{205}{50} - \left(\frac{-25}{50}\right)^2}$$

$$= 19.62$$

Example 3

Find the standard deviation for the following data:

Age (in years)	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89
Number of cases	1	0	1	10	17	38	9	3

Solution

Let $a = 44.5$ be the assumed mean and $h = 10$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-44.5}{10}$$

Age (in years)	No. of cases f	Midvalue x	$d = \frac{x-44.5}{10}$	fd	fd^2
10–19	1	14.5	–3	–3	9
20–29	0	24.5	–2	0	0
30–39	1	34.5	–1	–1	1
40–49	10	44.5	0	0	0
50–59	17	54.5	1	17	17
60–69	38	64.5	2	76	152
70–79	9	74.5	3	27	81
80–89	3	84.5	4	12	48
	$\sum f = 79$			$\sum fd = 128$	$\sum fd^2 = 308$

$$N = \sum f = 79$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= 10\sqrt{\frac{308}{79} - \left(\frac{128}{79}\right)^2}$$

$$= 11.285$$

Example 4

Find the mean and standard deviation of the following distribution:

Age (in years)	No. of Persons
less than 20	0
less than 25	170
less than 30	280
less than 35	360
less than 40	405
less than 45	445
less than 50	480

Solution

This is a 'less than' type of frequency distribution. This is first converted into an exclusive series. Let $a = 32.5$ be the assumed mean and $h = 5$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-32.5}{5}$$

Class intervals	No. of Persons f	Midvalue x	$d = \frac{x-32.5}{5}$	fd	fd^2
20–25	170	22.5	–2	–340	680
25–30	110	27.5	–1	–110	110
30–35	80	32.5	0	0	0
35–40	45	37.5	1	45	45
40–45	40	42.5	2	80	160
45–50	35	47.5	3	105	315
	$\sum f = 480$			$\sum fd = -220$	$\sum fd^2 = 1310$

$$N = \sum f = 480$$

$$\bar{x} = a + h\frac{\sum fd}{N}$$

$$= 32.5 + 5\left(\frac{-220}{480}\right)$$

$$= 30.21 \text{ years}$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= 5\sqrt{\frac{1310}{480} - \left(\frac{-220}{480}\right)^2}$$

$$= 7.94 \text{ years}$$

Example 5

A student obtained the mean and standard deviation of 100 *observations as* 40 *and* 5.1 *respectively. It was later discovered that he had wrongly copied down an observation as* 50 *instead of* 40*. Calculate the correct mean and standard deviation.*

Solution

$$n = 100, \quad \bar{x} = 40, \quad \sigma = 5.1$$

$$\bar{x} = \frac{\sum x}{n}$$

$$40 = \frac{\sum x}{100}$$

$$\therefore \quad \sum x = 4000$$

Correct $\sum x$ = Uncorrect $\sum x$ – Wrong observation + Correct observation

$$= 4000 - 50 + 40$$

$$= 3990$$

Correct $\bar{x} = \dfrac{\text{Correct} \sum x}{n}$

$$= \frac{3990}{100}$$

$$= 39.9$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$5.1 = \sqrt{\frac{\sum x^2}{100} - \left(\frac{4000}{100}\right)^2}$$

$$\therefore \quad \sum x^2 = 162601$$

Correct $\sum x^2$ = Uncorrect $\sum x^2$ – Wrong observation + Correct observation

$$= 162601 - (50)^2 + (40)^2$$

$$= 161701$$

$$\text{Correct} \quad \sigma = \sqrt{\frac{\text{Correct}\sum x^2}{n} - \left(\frac{\text{Correct}\sum x}{n}\right)^2}$$

$$= \sqrt{\frac{161701}{100} - \left(\frac{3990}{100}\right)^2}$$

$$= 5$$

3.8.4 Coefficient of Variation

The *standard deviation* is an absolute measure of dispersion. The coefficient of variation is a relative measure of dispersion and is denoted by CV.

$$\text{CV} = \frac{\sigma}{\bar{x}} \times 100$$

where σ is the standard deviation and $\bar{x}$ is the mean of the given series. The coefficient of variation has great practical significance and is the best measure of comparing the variability of two series. The series or groups for which the coefficient of variation is greater is said to be more variable or less consistent. On the other hand, the series for which the variation is lesser is said to be less variable or more consistent.

Example 1

The arithmetic mean of the runs scored by three batsmen Amit, Sumeet, and Nayan in the series are 50, 48, *and* 12 *respectively. The standard deviations of their runs are* 15, 12, *and* 2 *respectively. Who is the more consistent of the three?*

Solution

Let $\bar{x}_1, \bar{x}_2, \bar{x}_3$ be the arithmetic means and σ_1, σ_2, σ_3 be the standard deviations of the runs scored by Amit, Sumeet, and Nayan.

$$\bar{x}_1 = 50, \bar{x}_2 = 48, \bar{x}_3 = 12, \sigma_1 = 15, \sigma_2 = 12, \sigma_3 = 2$$

$$\text{CV}_1 = \frac{\sigma_1}{\bar{x}_1} \times 100$$

$$= \frac{15}{50} \times 100$$

$$= 30\%$$

$$\begin{aligned}
CV_2 &= \frac{\sigma_2}{\bar{x}_2} \times 100 \\
&= \frac{12}{48} \times 100 \\
&= 25\% \\
CV_3 &= \frac{\sigma_3}{\bar{x}_3} \times 100 \\
&= \frac{2}{12} \times 100 \\
&= 16.67\%
\end{aligned}$$

Since the coefficient of variation of Nayan is least, he is the most consistent.

Example 2

The runs scored by two batsmen A and B in 9 consecutive matches are given below:

A	85	20	62	28	74	5	69	4	13
B	72	4	15	30	59	15	49	27	26

Which of the batsmen is more consistent?

Solution

$$n = 9$$

For the batsman A,

$$\sum x_A = 85 + 20 + 62 + 28 + 74 + 5 + 69 + 4 + 13 = 360$$

$$\sum x_A^2 = 85^2 + 20^2 + 62^2 + 28^2 + 74^2 + 5^2 + 69^2 + 4^2 + 13^2 = 22700$$

$$\begin{aligned}
\sigma_A &= \sqrt{\frac{\sum x_A^2}{n} - \left(\frac{\sum x_A}{n}\right)^2} \\
&= \sqrt{\frac{22700}{9} - \left(\frac{360}{9}\right)^2} \\
&= 30.37
\end{aligned}$$

$$\bar{x}_A = \frac{\sum x_A}{n} = \frac{360}{9} = 40$$

$$\begin{aligned}
\text{CV}_A &= \frac{\sigma_A}{\bar{x}_A} \times 100 \\
&= \frac{30.37}{40} \times 100 \\
&= 75.925\%
\end{aligned}$$

For the batsman B,

$$\sum x_B = 72 + 4 + 15 + 30 + 59 + 15 + 49 + 27 + 26 = 297$$

$$\sum x_B^2 = 72^2 + 4^2 + 15^2 + 30^2 + 59^2 + 15^2 + 49^2 + 27^2 + 26^2 = 13837$$

$$\begin{aligned}
\sigma_B &= \sqrt{\frac{\sum x_B^2}{n} - \left(\frac{\sum x_B}{n}\right)^2} \\
&= \sqrt{\frac{13837}{9} - \left(\frac{297}{9}\right)^2} \\
&= 21.18
\end{aligned}$$

$$\bar{x}_B = \frac{\sum x_B}{n} = \frac{297}{9} = 33$$

$$\begin{aligned}
\text{CV}_B &= \frac{\sigma_B}{\bar{x}_B} \times 100 \\
&= \frac{21.18}{33} \times 100 \\
&= 64.18\%
\end{aligned}$$

Since $\text{CV}_B < \text{CV}_A$, the batsman B is more consistent.

Example 3

The following is the record goals scored by Team A in a football season:

No. of goals scored by Team A in the match	0	1	2	3	4
No. of matches played in the month	1	9	7	5	3

For Team B, the average number of goals scored per match was 2.5 with a SD of 1.25 goals. Find which team may be considered more consistent.

Solution

$$N = 1 + 9 + 7 + 5 + 3 = 25$$

For Team A,

$$\sum fx_A = (1\times 0)+(9\times 1)+(7\times 2)+(5\times 3)+(3\times 4)=50$$

$$\sum fx_A^2 = (1\times 0^2)+(9\times 1^2)+(7\times 2^2)+(5\times 3^2)+(3\times 4^2)=130$$

$$\sigma_A = \sqrt{\frac{\sum fx_A^2}{N}-\left(\frac{\sum fx_A}{N}\right)^2}$$

$$= \sqrt{\frac{130}{25}-\left(\frac{50}{25}\right)^2}$$

$$= 1.095$$

$$\bar{x}_A = \frac{\sum fx_A}{n} = \frac{50}{25} = 2$$

$$\text{CV}_A = \frac{\sigma_A}{\bar{x}_A}\times 100$$

$$= \frac{1.095}{2}\times 100$$

$$= 54.75\%$$

$$\sigma_B = 1.25, \quad \bar{x}_B = 2.5$$

$$\text{CV}_B = \frac{\sigma_B}{\bar{x}_B}\times 100$$

$$= \frac{1.25}{2.5}\times 100$$

$$= 50\%$$

Since $\text{CV}_B < \text{CV}_A$, Team B is more consistent.

Example 4

The number of matches played and goals scored by two teams A and B in World Cup Football 2002 *were as follows:*

Matches played by Team A	27	9	8	5	4
Matches played by Team B	17	9	6	5	3
No. of goals scored in a match	0	1	2	3	4

Find which team may be considered more consistent.

Solution

For Team A,

$$N_A = 27+9+8+5+4 = 53$$

$$\sum fx_A = (27\times 0)+(9\times 1)+(8\times 2)+(5\times 3)+(4\times 4) = 56$$

$$\sum fx_A^2 = (27\times 0^2)+(9\times 1^2)+(8\times 2^2)+(5\times 3^2)+(4\times 4^2) = 150$$

$$\sigma_A = \sqrt{\frac{\sum fx_A^2}{N_A} - \left(\frac{\sum fx_A}{N_A}\right)^2}$$

$$= \sqrt{\frac{150}{53} - \left(\frac{56}{53}\right)^2}$$

$$= 1.31$$

$$\bar{x}_A = \frac{\sum fx_A}{N_A} = \frac{56}{53} = 1.06$$

$$\text{CV}_A = \frac{\sigma_A}{\bar{x}_A}\times 100$$

$$= \frac{1.31}{1.06}\times 100$$

$$= 123.58\%$$

For Team B,

$$N_B = 17+9+6+5+3 = 40$$

$$\sum fx_B = (17\times 0)+(9\times 1)+(6\times 2)+(5\times 3)+(3\times 4) = 48$$

$$\sum fx_B^2 = (17\times 0^2)+(9\times 1^2)+(6\times 2^2)+(5\times 3^2)+(3\times 4^2) = 126$$

$$\sigma_B = \sqrt{\frac{\sum fx_B^2}{N_B} - \left(\frac{\sum fx_B}{N_B}\right)^2}$$

$$= \sqrt{\frac{126}{40} - \left(\frac{48}{40}\right)^2}$$

$$= 1.31$$

$$\bar{x}_B = \frac{\sum fx_B}{N_B} = \frac{48}{40} = 1.2$$

$$\text{CV}_B = \frac{\sigma_B}{\bar{x}_B}\times 100$$

$$= \frac{1.31}{1.2}\times 100$$

$$= 109.17\%$$

Since $CV_B < CV_A$, Team B is more consistent in performance.

Example 5

Two automatic filling machines A and B are used to fill a mixture of cement concrete in a beam. A random sample of beams on each machine showed the following information:

Machine A	32	28	47	63	71	39	10	60	96	14
Machine B	19	31	48	53	67	90	10	62	40	80

Find the standard deviation of each machine and also comment on the performances of the two machines.

[Summer 2015]

Solution

$$n = 10$$

$$\bar{x} = \frac{\sum x}{n} = \frac{460}{10} = 46$$

$$\bar{y} = \frac{\sum y}{n} = \frac{500}{10} = 50$$

Machine A			Machine B		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
32	–14	196	19	–31	961
28	–18	324	31	–19	361
47	1	1	48	–2	4
63	17	289	53	3	9
71	25	625	67	17	289
39	–7	49	90	40	1600
10	–36	1296	10	–40	1600
60	14	196	62	12	144
96	50	2500	40	–10	100
14	–32	1024	80	30	900
$\sum x = 460$		$\Sigma(x-\bar{x})^2 = 6500$	$\sum y = 500$		$\Sigma(y-\bar{y})^2 = 5968$

$$\begin{aligned}\sigma_A &= \sqrt{\frac{\sum(x-\bar{x})^2}{n}}\\ &= \sqrt{\frac{6500}{10}}\\ &= 25.495\end{aligned}$$

$$\begin{aligned}\sigma_B &= \sqrt{\frac{\sum(y-\bar{y})^2}{n}}\\ &= \sqrt{\frac{5968}{10}}\\ &= 24.429\end{aligned}$$

$$\begin{aligned}\text{CV}_A &= \frac{\sigma_A}{\bar{x}}\times 100\\ &= \frac{25.495}{46}\times 100\\ &= 55.423\%\end{aligned}$$

$$\begin{aligned}\text{CV}_B &= \frac{\sigma_B}{\bar{y}}\times 100\\ &= \frac{24.429}{50}\times 100\\ &= 48.858\%\end{aligned}$$

Since $\text{CV}_B < \text{CV}_A$, there is less variability in the performance of the machine B.

3.8.5 Combined Standard Deviation

If $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ be the arithmetic means, $\sigma_1, \sigma_2, ..., \sigma_k$ be the standard deviations, and $n_1, n_2, ..., n_k$ be the number of observations of k groups then the combined standard deviation is given by

$$\sigma = \sqrt{\frac{n_1(\sigma_1^2+d_1^2)+n_2(\sigma_2^2+d_2^2)+\cdots+n_k(\sigma_k^2+d_k^2)}{n_1+n_2+\cdots+n_k}}$$

where $\quad d_1 = \bar{x}_1-\bar{x},\ d_2 = \bar{x}_2-\bar{x}, ..., d_k = \bar{x}_k - x$

Example 1

A sample of 90 *values has a means of* 55 *and a standard deviation of* 3. *A second sample of* 110 *values has a mean of* 60 *and a standard deviation of* 2. *Find the mean and standard deviation of the combined sample of* 200 *values.*

Solution

$$n_1 = 90, \quad \bar{x}_1 = 55, \quad \sigma_1 = 3$$
$$n_2 = 110, \quad \bar{x}_2 = 60, \quad \sigma_2 = 2$$

Combined mean $\bar{x} = \dfrac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$

$$= \frac{(90 \times 55) + (110 \times 60)}{90 + 110}$$
$$= 57.75$$

$$d_1 = \bar{x}_1 - \bar{x} = 55 - 57.75 = -2.75$$
$$d_2 = x_2 - \bar{x} = 60 - 57.75 = 2.25$$

Combined standard deviation $\sigma = \sqrt{\dfrac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$

$$= \sqrt{\frac{90\left[3^2 + (-2.75)^2\right] + 110(2^2 + 2.25)^2}{90 + 110}}$$
$$= 3.53$$

EXERCISE 3.5

1. Find the standard deviation of 10 persons whose income in rupees is given below:

312, 292, 227, 235, 269, 255, 333, 348, 321, 299

[**Ans.:** 39.24]

2. Calculate the standard deviation from the following data:

Heights in cm	150	155	160	165	170	175	180
No. of students	15	24	32	33	24	16	6

[**Ans.:** 8.038 cm]

3. Find the standard deviation of the following data:

Size of items	10	11	12	13	14	15	16
Frequency	2	7	11	15	10	4	1

[**Ans.:** 1.342]

4. Calculate the standard deviation for the following frequency distribution:

Class interval	0–4	4–8	8–12	12–16
Frequency	4	8	2	1

[**Ans.:** 3.27]

5. Calculate the standard deviation of the following series:

Marks	0–10	10–20	20–30	30–40	40–50
Frequency	10	8	15	8	4

[**Ans.:** 12.37]

6. Calculate the SD for the following distributions of 300 telephone calls according to their durations in seconds:

Duration (in seconds)	0–30	30–60	60–90	90–120	120–150	150–180	180–210
No. of calls	9	17	43	82	81	44	24

[**Ans.:** 42.51]

7. Calculate the standard deviation from the following data:

Age less than (in years)	10	20	30	40	50	60	70	80
No. of Persons	15	30	53	75	100	110	115	125

[**Ans.:** 19.75]

8. Find the standard deviation from the following data:

Midvalue	30	35	40	45	50	55	60	65	70	75	80
Frequency u	1	2	4	7	9	13	17	12	7	6	3

[**Ans.:** 11.04]

9. Two cricketers scored the following runs in ten innings. Find who is a better run-getter and who is a more consistent player.

A	42	17	83	59	72	76	64	45	40	32
B	28	70	31	0	59	108	82	14	3	95

[**Ans.:** *A* is a better run-getter and is more consistent]

10. Two workers on the same job show the following results over a long period of time:

	Worker *A*	Worker *B*
Mean time (in minutes)	30	25
Standard deviation (in minutes)	6	4

[**Ans.:** *B* is more consistent]

11. The mean and standard deviation of 100 items are found to be 40 and 10. At the time of calculations, two items are wrongly taken as 30 and 72 instead of 3 and 27. Find the correct mean and correct standard deviation.

[**Ans.:** 39.28, 10.18]

12. The mean and standard deviation of distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the mean and standard deviation of all the 250 items taken together.

[**Ans.:** 44, 7.46]

3.9 SKEWNESS

Skewness is a measure that refers to the extent of symmetry or asymmetry in a distribution. A distribution is said to be *symmetrical* when its mean, median, and mode are equal, and the frequencies are symmetrically distributed about the mean. A symmetrical distribution when plotted on a graph will give a perfectly bell-shaped curve which is known as a *normal curve* (Fig. 3.1).

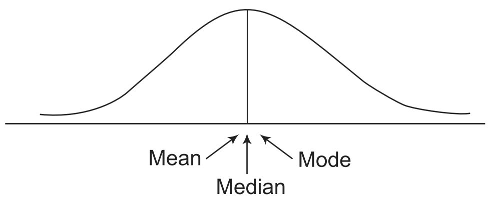
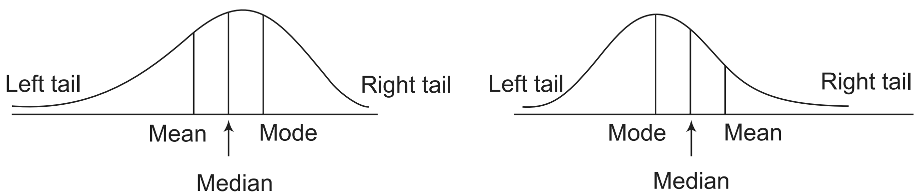


Fig. 3.1

A distribution is said to be asymmetrical or skewed when the mean, median, and mode are not equal, i.e., the mean median, and mode do not coincide. If the curve has a longer tail towards the left, it is said to be a negatively skewed distribution (Fig. 3.2*a*). If the curve has a longer tail towards the right, it is said to be positively skewed (Fig. 3.2*b*).



(a) Negatively skewed distribution (b) Positively skewed distribution

Fig. 3.2

Skewness gives an idea of the nature and degree of concentration of observations about the mean.

3.9.1 Measures of Skewness

A measure of skewness gives the extent and direction of skewness of a distribution. These measures can be absolute or relative. The absolute measures are also known as measures of skewness.

$$\text{Absolute skewness} = \text{Mean} - \text{Mode}$$

If the value of the mean is greater than the mode, the skewness will be positive and if the value of the mean is less than the mode, the skewness will be negative.

The relative measures of skewness is called the *coefficient of skewness*.

3.9.2 Karl Pearson's Coefficient of Skewness

Karl Pearson's coefficient of skewness denoted by S_k, is given by

$$S_k = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}$$

$$= \frac{\text{Mean} - \text{Mode}}{\sigma}$$

When the mode is ill-defined and the distribution is moderately skewed, the averages have the following relationship:

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}$$

$$S_k = \frac{\text{Mean} - (3\ \text{Median} - 2\ \text{Mean})}{\text{Standard Deviation}}$$

$$= \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}}$$

$$= \frac{3(\text{Mean} - \text{Median})}{\sigma}$$

The coefficient of skewness usually lies between -1 and 1.

For a positively skewed distribution, $S_k > 0$.

For a negatively skewed distribution, $S_k < 0$.
For a symmetrical distribution, $S_k = 0$.

Example 1

Calculate Karl Pearson's coefficient of skewness for the following data:

x	0	1	2	3	4	5	6	7
y	12	17	29	19	8	4	1	0

Solution

Let $a = 4$ be the assumed mean.
$d = x - a = x - 4$

x	f	d	d^2	fd	fd^2
0	12	–4	16	–48	192
1	17	–3	9	–51	153
2	29	–2	4	–58	116
3	19	–1	1	–19	19
4	8	0	0	0	0
5	4	1	1	4	4
6	1	2	4	2	4
7	0	3	9	0	0
	$\sum f = 90$			$\sum fd = -170$	$\sum fd^2 = 488$

$$N = \sum f = 90$$

$$\bar{x} = a + \frac{\sum fd}{N}$$

$$= 4 + \left(\frac{-170}{90}\right)$$

$$= 2.11$$

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= \sqrt{\frac{488}{90} - \left(\frac{-170}{90}\right)^2}$$

$$= 1.36$$

Since the maximum frequency is 29, the mode is 2.

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma}$$
$$= \frac{2.11 - 2}{1.36}$$
$$= 0.08$$

Example 2

Calculate Karl Pearson's coefficient of skewness from the following data:

Wages (₹)	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50
No. of Workers	8	16	30	45	62	32	15	6

Solution

Let $a = 32.5$ be the assumed mean and $h = 5$ be the width of the class interval.

$$d = \frac{x - a}{h} = \frac{x - 32.5}{5}$$

Wages (₹)	No. of workers f	Midvalue x	$d = \frac{x - 32.5}{5}$	d^2	fd	fd^2
10–15	8	12.5	–4	16	–32	128
15–20	16	17.5	–3	9	–48	144
20–25	30	22.5	–2	4	–60	120
25–30	45	27.5	–1	1	–45	45
30–35	62	32.5	0	0	0	0
35–40	32	37.5	1	1	32	32
40–45	15	42.5	2	4	30	60
45–50	6	47.5	3	9	18	54
	$\sum f = 214$				$\sum fd = -105$	$\sum fd^2 = 583$

$$N = \sum f = 214$$

$$\bar{x} = a + h\frac{\sum fd}{N}$$
$$= 32.5 + 5\left(\frac{-105}{214}\right)$$
$$= 30.05$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$
$$= 5\sqrt{\frac{583}{214} - \left(\frac{-105}{214}\right)^2}$$
$$= 7.88$$

Since the maximum frequency is 32, the mode lies in the interval 30–35.

Here, $l = 30$, $h = 5$, $f_m = 62$, $f_1 = 45$, $f_2 = 32$

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$
$$= 30 + 5\left[\frac{62 - 45}{2(62) - 45 - 32}\right]$$
$$= 31.81$$

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma}$$
$$= \frac{30.05 - 31.81}{7.88}$$
$$= -0.223$$

Example 3

The scores at an aptitude test by 100 *candidates are given below. Calculate Karl Pearson's coefficient of skewness.*

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70
No. of candidates	10	15	24	25	10	10	6

Solution

Let $a = 35$ be the assumed mean and $h = 10$ be the width of the class interval.

$$d = \frac{x - a}{h} = \frac{x - 35}{10}$$

Marks	No. of candidates f	Midvalue x	$d=\frac{x-35}{10}$	d^2	fd	fd^2
0–10	10	5	–3	9	–30	90
10–20	15	15	–2	4	–30	60
20–30	24	25	–1	1	–24	24
30–40	25	35	0	0	0	0
40–50	10	45	1	1	10	10
50–60	10	55	2	4	20	40
60–70	6	65	3	9	18	54
	$\sum f = 100$				$\sum fd = -36$	$\sum fd^2 = 278$

$$N = \sum f = 100$$

$$\bar{x} = a + h\left(\frac{\sum fd}{N}\right)$$

$$= 35 + 10\left(\frac{-36}{100}\right)$$

$$= 31.4$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= 10\sqrt{\frac{278}{100} - \left(\frac{-36}{100}\right)^2}$$

$$= 16.28$$

Since the maximum frequency is 25, the mode lies in the interval 30–40.

Here, $l = 30$, $h = 10$, $f_m = 25$, $f_1 = 24$, $f_2 = 10$

$$\text{Mode } = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

$$= 30 + 10\left[\frac{25 - 24}{2(25) - 24 - 10}\right]$$

$$= 30.625$$

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

$$= \frac{31.4 - 30.625}{16.28}$$

$$= 0.0476$$

Example 4

Calculate Karl Pearson's coefficient of skewness from the following data:

Weekly wages	40–50	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130–140
No. of workers	5	6	8	10	25	30	36	50	60	70

Solution

The maximum frequency 70 occurs at the end of the frequency distribution. Hence, the mode is ill-defined and Karl Pearson's coefficient of skewness is obtained using the median. Let a = 85 be the assumed mean and h = 10 be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-85}{10}$$

Weekly wages	No. of workers f	Midvalue x	$d = \frac{x-85}{10}$	d^2	fd	fd^2	CF
40–50	5	45	–4	16	–20	80	5
50–60	6	55	–3	9	–18	54	11
60–70	8	65	–2	4	–16	32	19
70–80	10	75	–1	1	–10	10	29
80–90	25	85	0	0	0	0	54
90–100	30	95	1	1	30	30	84
100–110	36	105	2	4	72	144	120
110–120	50	115	3	9	150	450	170
120–130	60	125	4	16	240	960	230
130–140	70	135	5	25	350	1750	300
	$\sum f = 300$				$\sum fd = 778$	$\sum fd^2 = 3510$	

$$N = \sum f = 300$$

$$\bar{x} = a + h\left(\frac{\sum fd}{N}\right)$$

$$= 85 + 10\left(\frac{778}{300}\right)$$

$$= 110.93$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= 10\sqrt{\frac{3510}{300} - \left(\frac{778}{300}\right)^2}$$

$$= 22.304$$

$$\frac{N}{2} = \frac{300}{2} = 150$$

The cumulative frequency just greater than 150 is 170 and the corresponding class 110–120 is the median class.

Here, $l = 110$, $h = 10$, $f = 50$, $c = 120$

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$= 110 + \frac{10}{50}(150 - 120)$$

$$= 116$$

$$S_k = \frac{3(\text{Mean} - \text{median})}{\sigma}$$

$$= \frac{3(110.93 - 116)}{22.304}$$

$$= -0.682$$

Example 5

From the marks scored by 100 *students in Section A and* 100 *students in Section B of a class, the following measures were obtained:*

Section *A*	$\overline{x}_A = 55$	$\sigma_A = 15.4$	Mode = 58.72
Section *B*	$\overline{x}_A = 53$	$\sigma_B = 15.4$	Mode = 48.83

Determine which distribution of marks is more skewed.

Solution

$$S_{k_A} = \frac{\text{Mean} - \text{Mode}}{\sigma_A} = \frac{55 - 58.72}{15.4} = -0.24$$

$$S_{k_B} = \frac{\text{Mean} - \text{Mode}}{\sigma_B} = \frac{53 - 48.83}{15.4} = 0.27$$

$$|0.27| > |-0.24|$$

Hence, the distribution of marks of Section *B* is more skewed.

Example 6

For a group of 10 *items,* $\sum x = 452, \sum x^2 = 24270$, *and mode* = 43.7. *Find Karl Pearson's coefficient of skewness.*

Solution

$$n = 10, \quad \sum x = 452, \quad \sum x^2 = 24270, \quad \text{mode} = 43.7$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{452}{10} = 45.2$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{24270}{10} - \left(\frac{452}{10}\right)^2}$$

$$= 19.59$$

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

$$= \frac{45.2 - 43.7}{19.59}$$

$$= 0.077$$

Example 7

In a distribution, the mean = 65, *median* = 70, *coefficient of skewness* = –0.6. *Find the mode and coefficient of variation.*

Solution

$$\bar{x} = 65, \quad \text{Median} = 70, \quad S_k = -0.6$$

$$\text{Mode} = 3\text{ Median} - 2\text{ Mean} = 3(70) - 2(65) = 80$$

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

$$-0.6 = \frac{65 - 80}{\sigma}$$

$$\therefore \quad \sigma = 25$$

$$\text{CV} = \frac{\sigma}{\bar{x}} \times 100 = \frac{25}{65} \times 100 = 38.64\%$$

Example 8

The following information was obtained from the records of a factory relating to wages:

Arithmetic mean = ₹ 56.8, Median = ₹ 59.5, *Standard deviation* = ₹ 12.4

Give the information about the distribution of wages.

Solution

$$\bar{x} = 56.8, \quad \text{Median} = 59.5, \quad \sigma = 12.4$$

$$S_k = \frac{3(\text{Mean} - \text{Median})}{\sigma} = \frac{3(56.8 - 59.5)}{12.4} = -0.65$$

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean} = 3(59.5) - 2(56.8) = 64.9$$

Hence, the maximum wages is ₹ 64.9.

There is a negative skewness in wages.

Example 9

For a moderately skewed distribution of retail price for men's shoes, it is found that the mean price is ₹ 20 *and the median price is* ₹ 17. *If the coefficient of variation is* 20%, *find the Pearson's coefficient of skewness.*

Solution

$$\bar{x} = 20, \quad \text{Median} = 17, \quad \text{CV} = 20\%$$

$$\text{CV} = \frac{\sigma}{\bar{x}} \times 100$$

$$20 = \frac{\sigma}{20} \times 100$$

$$\therefore \quad \sigma = 4$$

$$S_k = \frac{3(\text{Mean} - \text{Median})}{\sigma} = \frac{3(20 - 17)}{4} = 2.25$$

EXERCISE 3.6

1. Calculate Karl Pearson's coefficient of skewness for the following data:

   25, 15, 23, 40, 27, 25, 23, 25, 30

   [**Ans.:** – 0.03]

2. Calculate Karl Pearson's coefficient of skewness for the following data:

Size	1	2	3	4	5	6	7
Frequency	10	18	30	25	12	3	2

[**Ans.:** 0.2075]

3. Find the coefficient of skewness for the following data:

Weekly wages (in ₹)	15	20	25	30	35	40	45
No. of earners	3	25	19	16	4	5	6

[**Ans.:** 0.88]

4. Find Karl Pearson's coefficient of skewness for the following data:

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70
No. of students	10	12	18	25	16	14	8

[**Ans.:** 0.013]

5. Calculate the coefficient of skewness from the following data:

Marks less than	20	30	40	50	60	70	80
No. of students	10	25	40	65	80	95	100

[**Ans.:** – 0.089]

6. For the data given below, calculate Karl Pearson's coefficient of skewness:

x	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140–150
f	12	18	35	42	50	45	20	8

[**Ans.:** 0.243]

7. Karl Pearson's measure of skewness of a distribution is 0.5. Its median and mode are respectively 42 and 36. Find the coefficient of variation.

[**Ans.:** 40]

8. From the marks scored by 120 students in Section *A* and 120 students in Section *B* of a class, the following measures are obtained:

Section *A*	$\bar{x} = 46.83$	SD = 14.8	mode = 51.67
Section *B*	$\bar{x} = 47.83$	SD = 14.8	mode = 47.07

Determine which distribution of marks is more skewed.

[**Ans.:** Section *A*]

9. For a moderately skewed data, the arithmetic mean is 200, the coefficient of variation is 8, and Karl Pearson's coefficient of skewness is 0.3. Find the mode and median.

[**Ans.:** 195.2, 198.4]

10. Karl Pearson's coefficient of skewness of a distribution is 0.32. Its standard deviation is 6.5 and the mean is 29.6. Find the mode and median for the distribution.

[**Ans.:** 27.52, 28.9]

11. The median, mode and coefficient of skewness for a certain distribution are respectively 17.4, 15.3, and 0.35. Find the coefficient of variation.

[**Ans.:** 48.78%]

12. In a distribution, mean = 65, median = 70, coefficient of skewness = –6. Find the mode and coefficient of variation.

[**Ans.:** 80, 39.78%]

Points to Remember

Arithmetic Mean

The *arithmetic mean* of a set of observations is their sum divided by the number of observations. If $x_1, x_2, \ldots, x_n$ be n observations then their average or arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}$$

If n observations consist of n distinct values denoted by $x_1, x_2, \ldots, x_n$ of the observed variable x occurring with frequencies $f_1, f_2, \ldots, f_n$ respectively then the arithmetic mean is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_ix_i}{N} = \frac{\sum fx}{N}$$

1. Arithmetic Mean of Grouped Data

In case of grouped or continuous frequency distribution the arithmetic mean is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} = \frac{\sum fx}{N}, \text{ where } N = \sum_{i=1}^{n} f_i$$

and x is taken as the midvalue of the corresponding class.

2. Arithmetic Mean from Assumed Mean

$$\bar{x} = a + \frac{\sum fd}{N}$$

3. Arithmetic Mean by the Step-Deviation Method

$$\bar{x} = a + h\frac{\sum fd}{N}$$

4. Weighted Arithmetic Mean

Weighted arithmetic mean $= \dfrac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n}$

$$\bar{x}_w = \frac{\sum wx}{\sum w}$$

When the assumed mean is used for calculation,

$$\bar{x}_w = a + \frac{\sum wd}{\sum w}$$

When the step-deviation method is used for calculation,

$$\bar{x}_w = a + h\frac{\sum wd}{\sum w}$$

Combined Arithmetic Mean

If $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ are the means of k series of sizes $n_1, n_2, ..., n_k$ respectively then the mean $\bar{x}$ of the composite series is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\,\bar{x}_2 + \cdots + n_k\,\bar{x}_k}{n_1 + n_2 + \cdots + n_k}$$

$$= \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i}$$

Median

Median is the central value of the variable when the values are arranged in ascending or descending order of magnitude.

In case of ungrouped data, if the number of observations is odd then the median is the middle value after the values have been arranged in ascending or descending order of magnitude. If the number of observations is even, there are two middle terms and the median is obtained by taking the arithmetic mean of the middle terms.

In case of discrete frequency distribution, the median is obtained by considering the cumulative frequencies. The steps for calculating the median are given below:

(i) Arrange the values of the variables in ascending or descending order of magnitudes.

(ii) Find $\dfrac{N}{2}$ where $N = \sum f$

(iii) Find the cumulative frequency just greater than $\dfrac{N}{2}$ and determine the corresponding value of the variable.

(iv) The corresponding value of x is the median.

Median for Continuous Frequency Distribution

In case of continuous frequency distribution (less than frequency distribution), the class corresponding to the cumulative frequency just greater than $\dfrac{N}{2}$, is called the *median class*, and the value of the median is given by

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

In case of 'more than' or 'greater than' type of frequency distributions, the value of the median is given by

$$\text{Median} = u - \frac{h}{f}\left(\frac{N}{2} - c\right)$$

where u is the upper limit of the median class

f is the frequency of the median class

h is the width of the median class

c is the cumulative frequency of the class succeeding the median class

Mode

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set are heavily distributed.

Mode for a Continuous Frequency Distribution

In case of a continuous frequency distribution, the class in which the mode lies is called the *modal class* and the value of the mode is given by

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

where l is the lower limit of the modal class

h is the width of the modal class

f_m is the frequency of the modal class

f_1 is the frequency of the class preceding the modal class

f_2 is the frequency of the class succeeding the modal class

Geometric Mean

The *geometric mean* of a set of n observations is the n^{th} root of their product.

$$\text{GM} = (x_1 \cdot x_2 \cdot \cdots \cdot x_n)^{\frac{1}{n}}$$

$$\text{GM} = \text{antilog}\left(\frac{\sum \log x}{n}\right)$$

In case of a frequency distribution consisting of n observations x_1, x_2, .., x_n with respective frequencies f_1, f_2, .., f_n, the geometric mean is given by

$$\text{GM} = (x_1^{f_1} \cdot x_2^{f_2} \cdots\cdot x_n^{f_n})^{\frac{1}{N}}, \text{ where } N = \sum f$$

$$\text{GM} = \text{antilog}\left(\frac{\sum f \log x}{N}\right)$$

Harmonic Mean

The *harmonic mean* of a number of observations, none of which is zero, is the reciprocal of the arithmetic mean of the reciprocals of the given values.

$$\text{HM} = \frac{1}{\frac{1}{n}\sum\left(\frac{1}{x}\right)}$$

$$= \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

In case of a frequency distribution consisting of n observations $x_1, x_2, ..., x_n$ with respective frequencies $f_1, f_2, ..., f_n$, the harmonic mean is given by

$$\text{HM} = \frac{f_1 + f_2 + \cdots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}}$$

$$= \frac{\sum f}{\sum \left(\frac{f}{x}\right)}$$

If $x_1, x_2, ..., x_n$ are n observations with weights $w_1, w_2, ..., w_n$ respectively, their weighted harmonic mean is given by

$$\text{HM} = \frac{\sum w}{\sum \left(\frac{w}{x}\right)}$$

Relation between Arithmetic Mean, Geometric Mean and Harmonic Mean

$$\text{AM} \geq \text{GM} \geq \text{HM}$$

For two observations x_1 and x_2 of a series,

$$\text{GM} = \sqrt{\text{AM} \cdot \text{HM}}$$

Standard Deviation

Standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean.

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

In case of a frequency distribution consisting of n observations $x_1, x_2, ..., x_n$ with respective frequencies $f_1, f_2, ..., f_n$, the standard deviation is given by

$$\sigma = \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2}$$

1. Standard Deviation from the Assumed Mean

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

2. Standard Deviation by Step-Deviation Method

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

3. Variance

The *variance* is the square of the standard deviation and is denoted by σ^2. The method for calculating variance is same as that given for the standard deviation.

4. Coefficient of Variation

The *standard deviation* is an absolute measure of dispersion. The coefficient of variation is a relative measure of dispersion and is denoted by CV.

$$\text{CV} = \frac{\sigma}{\bar{x}} \times 100$$

Skewness

Skewness is a measure that refers to the extent of symmetry or asymmetry in a distribution.

Karl Pearson's Coefficient of Skewness

Karl Pearson's coefficient of skewness denoted by S_k, is given by

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

When the mode is ill-defined and the distribution is moderately skewed, the averages have the following relationship:

$$S_k = \frac{\text{Mean} - (3\text{ Median} - 2\text{ Mean})}{\text{Standard Deviation}}$$

$$= \frac{3(\text{Mean} - \text{Median})}{\sigma}$$

CHAPTER 4

Correlation and Regression

Chapter Outline

4.1 INTRODUCTION

Correlation and regression are the most commonly used techniques for investigating the relationship between two quantitative variables. *Correlation* refers to the relationship of two or more variables. It measures the closeness of the relationship between the variables. *Regression* establishes a functional relationship between the variables. In correlation, both the variables x and y are random variables, whereas in regression, x is a random variable and y is a fixed variable. The coefficient of correlation is a relative measure whereas the regression coefficient is an absolute figure.

4.2 CORRELATION

Correlation is the relationship that exists between two or more variables. Two variables are said to be correlated if a change in one variable affects a change in the other variable. Such a data connecting two variables is called *bivariate data*. Thus, correlation is a statistical analysis which measures and analyses the degree or extent to which two variables fluctuate with reference to each other. Some examples of such a relationship are as follows:

1. Relationship between heights and weights.
2. Relationship between price and demand of commodity.
3. Relationship between rainfall and yield of crops.
4. Relationship between age of husband and age of wife.

4.3 TYPES OF CORRELATIONS

Correlation is classified into four types:

1. Positive and negative correlations
2. Simple and multiple correlations
3. Partial and total correlations
4. Linear and nonlinear correlations

4.3.1 Positive and Negative Correlations

Depending on the variation in the variables, correlation may be positive or negative.

1. Positive Correlation If both the variables vary in the same direction, the correlation is said to be positive. In other words, if the value of one variable increases, the value of the other variable also increases, or, if value of one variable decreases, the value of the other variable decreases, e.g., the correlation between heights and weights of group of persons is a positive correlation.

Height (cm)	150	152	155	160	162	165
Weight (kg)	60	62	64	65	67	69

2. Negative Correlation If both the variables vary in the opposite direction, correlation is said to be negative. In other words, if the value of one variable increases, the value of the other variable decreases, or, if the value of one variable decreases, the value of the other variable increases, e.g., the correlation between the price and demand of a commodity is a negative correlation.

Price (₹ per unit)	10	8	6	5	4	1
Demand (units)	100	200	300	400	500	600

4.3.2 Simple and Multiple Correlations

Depending upon the study of the number of variables, correlation may be simple or multiple.

1. Simple Correlation When only two variables are studied, the relationship is described as simple correlation, e.g., the quantity of money and price level, demand and price, etc.

2. Multiple Correlation When more than two variables are studied, the relationship is described as multiple correlation, e.g., relationship of price, demand, and supply of a commodity.

4.3.3 Partial and Total Correlations

Multiple correlation may be either partial or total.

1. Partial Correlation When more than two variables are studied excluding some other variables, the relationship is termed as partial correlation.

2. Total Correlation When more than two variables are studied without excluding any variables, the relationship is termed total correlation.

4.3.4 Linear and Nonlinear Correlations

Depending upon the ratio of change between two variables, the correlation may be linear or nonlinear.

1. Linear Correlation If the ratio of change between two variables is constant, the correlation is said to be linear. If such variables are plotted on a graph paper, a straight line is obtained, e.g.,

Milk (*l*)	5	10	15	20	25	30
Curg (kg)	2	4	6	8	10	12

2. Nonlinear Correlation If the ratio of change between two variables is not constant, the correlation is said to nonlinear. The graph of a nonlinear or curvilinear relationship will be a curve, e.g.,

Advertising expenses (₹ in lacs)	3	6	9	12	15
Sales (₹ in lacs)	10	12	15	15	16

4.4 METHODS OF STUDYING CORRELATION

There are two different methods of studying correlation, (1) Graphic methods (2) Mathematical methods.

Graphic methods are (a) scatter diagram, and (b) simple graph.

Mathematical methods are (a) Karl Pearson's coefficient of correlation, and (b) Spearman's rank coefficient of correlation.

4.5 SCATTER DIAGRAM

The scatter diagram is a diagrammatic representation of bivariate data to find the correlation between two variables. There are various correlationships between two variables represented by the following scatter diagrams.

1. Perfect Positive Correlation If all the plotted points lie on a straight line rising from the lower left-hand corner to the upper right-hand corner, the correlation is said to be perfectly positive (Fig. 4.1).

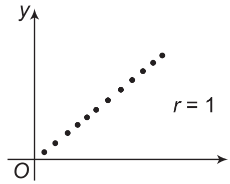


Fig. 4.1

2. Perfect Negative Correlation If all the plotted points lie on a straight line falling from the upper-left hand corner to the lower right-hand corner, the correlation is said to be perfectly negative (Fig. 4.2).

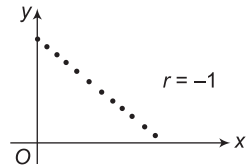


Fig. 4.2

3. High Degree of Positive Correlation If all the plotted points lie in the narrow strip, rising from the lower left-hand corner to the upper right-hand corner, it indicates a high degree of positive correlation (Fig. 4.3).

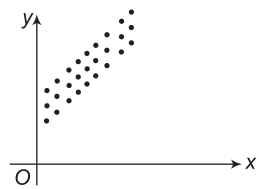


Fig. 4.3

4. High Degree of Negative Correlation If all the plotted points lie in a narrow strip, falling from the upper left-hand corner to the lower right-hand corner, it indicates the existence of a high degree of negative correlation (Fig. 4.4).

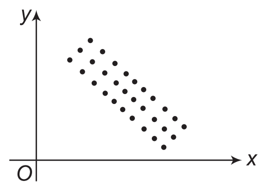


Fig. 4.4

5. No Correlation If all the plotted points lie on a straight line parallel to the *x*-axis or *y*-axis or in a haphazard manner, it indicates the absence of any relationship between the variables (Fig. 4.5).

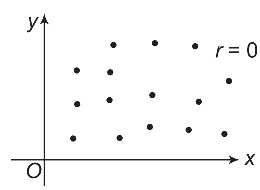


Fig. 4.5

Merits of a Scatter Diagram

1. It is simple and nonmathematical method to find out the correlation between the variables.

2. It gives an indication of the degree of linear correlation between the variables.
3. It is easy to understand.
4. It is not influenced by the size of extreme items.

4.6 SIMPLE GRAPH

A *simple graph* is a diagrammatic representation of bivariate data to find the correlation between two variables. The values of the two variables are plotted on a graph paper. Two curves are obtained, one for the variable x and the other for the variable y. If both the curves move in the same direction, the correlation is said to be positive. If both the curves move in the opposite direction, the correlation is said to be negative. This method is used in the case of a time series. It does not reveal the extent to which the variables are related.

4.7 KARL PEARSON'S COEFFICIENT OF CORRELATION

The coefficient of correlation is the measure of correlation between two random variables X and Y, and is denoted by r.

$$r = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}$$

where cov (X, Y) is covariance of variables X and Y,

σ_X is the standard deviation of variable X,

and σ_Y is the standard deviation of variable Y.

This expression is known as Karl Pearson's coefficient of correlation or Karl Pearson's product-moment coefficient of correlation.

$$\operatorname{cov}(X,Y) = \frac{1}{n}\sum (x-\overline{x})\,(y-\overline{y})$$

$$\sigma_X = \sqrt{\frac{\sum (x-\overline{x})^2}{n}}$$

$$\sigma_Y = \sqrt{\frac{\sum (y-\overline{y})^2}{n}}$$

$$\therefore \qquad r = \frac{\sum (x-\overline{x})\,(y-\overline{y})}{\sqrt{\sum (x-\overline{x})^2}\sqrt{\sum (y-\overline{y})^2}}$$

The above expression can be further modified.

Expanding the terms,

$$r = \frac{\sum (xy - x\overline{y} - \overline{x}y + \overline{x}\,\overline{y})}{\sqrt{\sum \left(x^2 - 2x\overline{x} + \overline{x}^2\right)}\sqrt{\sum \left(y^2 - 2y\overline{y} + \overline{y}^2\right)}}$$

$$= \frac{\sum xy - \overline{y}\sum x - \overline{x}\sum y + \overline{x}\,\overline{y}\sum 1}{\sqrt{\sum x^2 - 2\overline{x}\sum x + \overline{x}^2 \sum 1}\sqrt{\sum y^2 - 2\overline{y}\sum y + \overline{y}^2 \sum 1}}$$

$$= \frac{\sum xy - \frac{\sum y}{n}\sum x - \frac{\sum x}{n}\sum y + \frac{\sum x}{n}\frac{\sum y}{n}\cdot n}{\sqrt{\sum x^2 - 2\frac{\sum x}{n}\sum x + \left(\frac{\sum x}{n}\right)^2 n}\sqrt{\sum y^2 - 2\frac{\sum y}{n}\sum y + \left(\frac{\sum y}{n}\right)^2 n}}$$

$$= \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}}$$

4.8 PROPERTIES OF COEFFICIENT OF CORRELATION

1. The coefficient of correlation lies between −1 and 1, i.e., −1 ≤ *r* ≤ 1.

Proof Let $\overline{x}$ and $\overline{y}$ be the mean of x and y series and σ_x and σ_y be their respective standard deviations.

Let $\sum \left(\frac{x-\overline{x}}{\sigma_x} \pm \frac{y-\overline{y}}{\sigma_y}\right)^2 \ge 0$ $\left[\because \begin{array}{l}\text{sum of squares of real quantities}\\ \text{cannot be negative}\end{array}\right]$

$$\frac{\sum (x-\overline{x})^2}{\sigma_x^2} + \frac{\sum (y-\overline{y})^2}{\sigma_y^2} \pm \frac{2\sum (x-\overline{x})(y-\overline{y})}{\sigma_x \sigma_y} \ge 0$$

$$n + n \pm 2nr \ge 0$$

$$2n \pm 2nr \ge 0$$

$$2n(1 \pm r) \ge 0$$

$$1 \pm r \ge 0$$

i.e., $1 + r \ge 0$ or $1 - r \ge 0$

$r \ge -1$ or $r \le 1$

Hence, the coefficient of correlation lies between –1 and 1, i.e., $-1 \le r \le 1$.

2. Correlation coefficient is independent of change of origin and change of scale.

Proof Let $d_x = \dfrac{x-a}{h}, \quad d_y = \dfrac{y-b}{k}$

$$x = a + hd_x, \quad y = b + kd_y$$

where a, b, h (>0) and k(>0) are constants.

$$x = a + hd_x \Rightarrow \overline{x} = a + h\overline{d}_x \Rightarrow x - \overline{x} = h(d_x - \overline{d}_x)$$
$$y = b + kd_y \Rightarrow \overline{y} = b + h\,\overline{d}_y \Rightarrow y - \overline{y} = k(d_y - \overline{d}_y)$$

$$\begin{aligned} r_{xy} &= \frac{\sum (x-\overline{x})(y-\overline{y})}{\sqrt{\sum (x-\overline{x})^2}\sqrt{\sum (y-\overline{y})^2}} \\ &= \frac{\sum h(d_x-\overline{d}_x)\,k(d_y-\overline{d}_y)}{\sqrt{\sum h^2(d_x-\overline{d}_x)^2}\sqrt{\sum k^2(d_y-\overline{d}_y)^2}} \\ &= \frac{\sum (d_x-\overline{d}_x)(d_y-\overline{d}_y)}{\sqrt{\sum (d_x-\overline{d}_x)^2}\sqrt{(d_y-\overline{d}_y)^2}} \\ &= r_{d_x d_y} \end{aligned}$$

Hence, the correlation coefficient is independent of change of origin and change of scale.

Note Since correlation coefficient is independent of change of origin and change of scale,

$$r = \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \dfrac{\left(\sum d_x\right)^2}{n}}\sqrt{\sum d_y^2 - \dfrac{\left(\sum d_y\right)^2}{n}}}$$

3. Two independent variables are uncorrelated.

Proof If random variables X and Y are independent,

$$\sum (x-\overline{x})(y-\overline{y}) = 0 \text{ or } \operatorname{cov}(X, Y) = 0$$

$$\therefore \qquad r = 0$$

Thus, if X and Y are independent variables, they are uncorrelated.

Note The converse of the above property is not true, i.e., two uncorrelated variables may not be independent.

Example 1

Calculate the correlation coefficient between x and y using the following data:

x	2	4	5	6	8	11
y	18	12	10	8	7	5

Solution

$n = 6$

x	y	x^2	y^2	xy
2	18	4	324	36
4	12	16	144	48
5	10	25	100	50
6	8	36	64	48
8	7	64	49	56
11	5	121	25	55
$\sum x = 36$	$\sum y = 60$	$\sum x^2 = 266$	$\sum y^2 = 706$	$\sum xy = 293$

$$r = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}}$$

$$= \frac{293 - \dfrac{(36)(60)}{6}}{\sqrt{266 - \dfrac{(36)^2}{6}}\sqrt{706 - \dfrac{(60)^2}{6}}}$$

$$= -0.9203$$

Example 2

Calculate the coefficient of correlation from the following data:

x	12	9	8	10	11	13	7
y	14	8	6	9	11	12	3

Solution

$n = 7$

x	y	x^2	y^2	xy
12	14	144	196	168
9	8	81	64	72
8	6	64	36	48
10	9	100	81	90
11	11	121	121	121
13	12	169	144	156
7	3	49	9	21
$\sum x = 70$	$\sum y = 63$	$\sum x^2 = 728$	$\sum y^2 = 651$	$\sum xy = 676$

$$r = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}}$$

$$= \frac{676 - \dfrac{(70)\,(63)}{7}}{\sqrt{728 - \dfrac{(70)^2}{7}}\sqrt{651 - \dfrac{(63)^2}{7}}}$$

$$= 0.949$$

Example 3

Calculate the coefficient of correlation for the following data:

x	9	8	7	6	5	4	3	2	1
y	15	16	14	13	11	12	10	8	9

Solution

$n = 9$

x	y	x^2	y^2	xy
9	15	81	225	135
8	16	64	256	128
7	14	49	196	98
6	13	36	169	78
5	11	25	121	55
4	12	16	144	48
3	10	9	100	30
2	8	4	64	16
1	9	1	81	9
$\sum x = 45$	$\sum y = 108$	$\sum x^2 = 285$	$\sum y^2 = 1356$	$\sum xy = 597$

$$r = \frac{\sum x\,y - \dfrac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}}$$

$$= \frac{597 - \dfrac{(45)(108)}{9}}{\sqrt{285 - \dfrac{(45)^2}{9}}\sqrt{1356 - \dfrac{(108)^2}{9}}}$$

$$= 0.95$$

Example 4

Calculate the correlation coefficient between the following data:

x	5	9	13	17	21
y	12	20	25	33	35

Solution

$$n = 5$$

$$\overline{x} = \frac{\sum x}{n} = \frac{65}{5} = 13$$

$$\overline{y} = \frac{\sum y}{n} = \frac{125}{5} = 25$$

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
5	12	–8	–13	64	169	104
9	20	–4	–5	16	25	20
13	25	0	0	0	0	0
17	33	4	8	16	64	32
21	35	8	10	64	100	80
$\sum x = 65$	$\sum y = 125$	$\sum(x-\overline{x}) = 0$	$\sum(y-\overline{y}) = 0$	$\sum(x-\overline{x})^2 = 160$	$\sum(y-\overline{y})^2 = 358$	$\sum(x-\overline{x})(y-\overline{y}) = 236$

$$\begin{aligned} r &= \frac{\sum (x-\overline{x})(y-\overline{y})}{\sqrt{\sum (x-\overline{x})^2}\sqrt{\sum (y-\overline{y})^2}} \\ &= \frac{236}{\sqrt{160}\sqrt{358}} \\ &= 0.986 \end{aligned}$$

Example 5

Calculate the correlation coefficient between for the following values of demand and the corresponding price of a commodity:

Demand in Quintals	65	66	67	67	68	69	70	72
Price in rupees per kg	67	68	65	68	72	72	69	71

Solution

Let the demand in quintal be denoted by x and the price in rupees per kg be denoted by y.

$$n = 8$$

$$\bar{x} = \frac{\sum x}{n} = \frac{544}{8} = 68$$

$$\bar{y} = \frac{\sum y}{n} = \frac{552}{8} = 69$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
65	67	–3	–2	9	4	6
66	68	–2	–1	4	1	2
67	65	–1	–4	1	16	4
67	68	–1	–1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
$\sum x = 544$	$\sum y = 552$	$\sum(x-\bar{x}) = 0$	$\sum(y-\bar{y}) = 0$	$\sum(x-\bar{x})^2 = 36$	$\sum(y-\bar{y})^2 = 44$	$\sum(x-\bar{x})(y-\bar{y}) = 24$

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

$$= \frac{24}{\sqrt{36}\sqrt{44}}$$

$$= 0.603$$

Example 6

Calculate the coefficient of correlation for the following pairs of x and y:

x	17	19	21	26	20	28	26	27
y	23	27	25	26	27	25	30	33

Solution

Let $a = 23$ and $b = 27$ be the assumed means of x and y series respectively.

$$d_x = x - a = x - 23$$
$$d_y = y - b = y - 27$$
$$n = 8$$

x	y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
17	23	–6	–4	36	16	24
19	27	–4	0	16	0	0
21	25	–2	–2	4	4	4
26	26	3	–1	9	1	–3
20	27	–3	0	9	0	0
28	25	5	–2	25	4	–10
26	30	3	3	9	9	9
27	33	4	6	16	36	24
		$\Sigma d_x = 0$	$\Sigma d_y = 0$	$\Sigma d_x^2 = 124$	$\Sigma d_y^2 = 70$	$\Sigma d_x d_y = 48$

$$r = \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \dfrac{\left(\sum dx\right)^2}{n}}\sqrt{\sum d_y^2 - \dfrac{\left(\sum d_y\right)^2}{n}}}$$

$$= \frac{48 - 0}{\sqrt{124 - 0}\,\sqrt{70 - 0}}$$

$$= 0.515$$

Example 7

Calculate the correlation coefficient from the following data:

x	23	27	28	29	30	31	33	35	36	39
y	18	22	23	24	25	26	28	29	30	32

Solution

Let $a = 30$ and $b = 25$ be the assumed means of x and y series respectively.

$$d_x = x - a = x - 30$$
$$d_y = y - b = x - 25$$
$$n = 10$$

x	y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
23	18	–7	–7	49	49	49
27	22	–3	–3	9	9	9
28	23	–2	–2	4	4	4
29	24	–1	–1	1	1	1
30	25	0	0	0	0	0
31	26	1	1	1	1	1
33	28	3	3	9	9	9
35	29	5	4	25	16	20
36	30	6	5	36	25	30
39	32	9	7	81	49	63
		$\Sigma d_x = 11$	$\Sigma d_y = 7$	$\Sigma d_x^2 = 215$	$\Sigma d_y^2 = 163$	$\Sigma d_x d_y = 186$

$$r = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \frac{\left(\sum d_x\right)^2}{n}}\sqrt{\sum d_y^2 - \frac{\left(\sum d_y\right)^2}{n}}}$$

$$= \frac{186 - \frac{(11)(7)}{10}}{\sqrt{215 - \frac{(11)^2}{10}}\sqrt{163 - \frac{(7)^2}{10}}}$$

$$= 0.996$$

Example 8

Calculate the coefficient of correlation between the ages of cars and annual maintenance costs.

Age of cars (year)	2	4	6	7	8	10	12
Annual maintenance cost (₹)	1600	1500	1800	1900	1700	2100	2000

Solution

Let the ages of cars in years be denoted by x and annual maintenance costs in rupees be denoted by y.

Let $a = 7$ and $b = 1800$ be the assumed means of x and y series respectively.

Let $h = 1$, $k = 100$

$$d_x = \frac{x-a}{h} = \frac{x-7}{1} = x-7$$

$$d_y = \frac{y-b}{k} = \frac{y-1800}{100}$$

$$n = 7$$

x	y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
2	1600	–5	–2	25	4	10
4	1500	–3	3	9	9	9
6	1800	–1	0	1	0	0
7	1900	0	1	0	1	0
8	1700	1	–1	1	1	–1
10	2100	3	3	9	9	9
12	2000	5	2	25	4	10
		$\sum d_x = 0$	$\sum d_y = 0$	$\sum d_x^2 = 70$	$\sum d_y^2 = 28$	$\sum d_x d_y = 37$

$$r = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \frac{\left(\sum d_x\right)^2}{n}}\sqrt{\sum d_y^2 - \frac{\left(\sum d_y\right)^2}{n}}}$$

$$= \frac{37-0}{\sqrt{70-0}\sqrt{28-0}}$$

$$= 0.836$$

Example 9

Calculate Karl Pearson's coefficient of correlation for the data given below:

x	10	14	18	22	26	30
y	18	12	24	6	30	36

Solution

Let $a = 22$ and $b = 24$ be the assumed means of x and y series respectively.

Let $h = 4$, $k = 6$

$$d_x = \frac{x-a}{h} = \frac{x-22}{4}$$

$$d_y = \frac{y-b}{k} = \frac{y-24}{6}$$

$$n = 6$$

x	y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
10	18	–3	–1	9	1	3
14	12	–2	–2	4	4	4
18	24	–1	0	1	0	0
22	6	0	–3	0	9	0
26	30	1	1	1	1	1
30	36	2	2	4	4	4
		$\sum d_x = -3$	$\sum d_y = -3$	$\sum d_x^2 = 19$	$\sum d_y^2 = 19$	$\sum d_x d_y = 12$

$$r = \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \dfrac{\left(\sum d_x\right)^2}{n}}\sqrt{\sum d_y^2 - \dfrac{\left(\sum d_y\right)^2}{n}}}$$

$$= \frac{12 - \dfrac{(-3)(-3)}{6}}{\sqrt{19 - \dfrac{(-3)^2}{6}}\sqrt{19 - \dfrac{(-3)^2}{6}}}$$

$$= 0.6$$

Example 10

The coefficient of correlation between two variables X and Y is 0.48. The covariance is 36. The variance of X is 16. Find the standard deviation of Y.

Solution

$$r = 0.48, \qquad \text{cov}(X, Y) = 36, \qquad \sigma_X^2 = 16$$

$$\therefore \qquad \sigma_X = 4$$

$$r = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$0.48 = \frac{36}{4\sigma_Y}$$

$$\therefore \quad \sigma_Y = 18.75$$

Example 11

Given $n = 10$, $\sigma_X = 5.4$, $\sigma_Y = 6.2$, *and sum of the product of deviations from the mean of x and y is* 66. *Find the correlation coefficient.*

Solution

$$n = 10,\ \sigma_X = 5.4,\ \sigma_Y = 6.2$$

$$\sum (x - \overline{x})(y - \overline{y}) = 66$$

$$\sigma_X = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$$

$$5.4 = \sqrt{\frac{\sum (x - \overline{x})^2}{10}}$$

$$\therefore \quad \sum (x - \overline{x})^2 = 291.6$$

$$\sigma_Y = \sqrt{\frac{\sum (y - \overline{y})^2}{n}}$$

$$6.2 = \sqrt{\frac{\sum (y - \overline{y})^2}{10}}$$

$$\therefore \quad \sum (y - \overline{y})^2 = 384.4$$

$$r = \frac{\sqrt{\sum (x - \overline{x})(y - \overline{y})}}{\sqrt{\sum (x - \overline{x})^2}\ \sqrt{\sum (y - \overline{y})^2}}$$

$$= \frac{66}{\sqrt{291.6}\ \sqrt{384.4}}$$

$$= 0.197$$

Example 12

From the following information, calculate the value of n.

$$\sum x = 4, \sum y = 4, \sum x^2 = 44, \sum y^2 = 44, \sum xy = -40, r = -1$$

Solution

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}}$$

$$-1 = \frac{-40 - \frac{(4)(4)}{n}}{\sqrt{44 - \frac{(4)^2}{n}}\sqrt{44 - \frac{(4)^2}{n}}}$$

$$\therefore \qquad n = 8$$

Example 13

From the following data, find the number of items n.

$$r = 0.5, \sum (x - \overline{x})(y - \overline{y}) = 120, \sigma_Y = 8, \sum (x - \overline{x})^2 = 90$$

Solution

$$\sigma_Y = \sqrt{\frac{\sum (y - \overline{y})^2}{n}}$$

$$8 = \sqrt{\frac{\sum (y - \overline{y})^2}{n}}$$

$$\sum (y - \overline{y})^2 = 64\,n$$

$$r = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sqrt{\sum (x - \overline{x})^2}\sqrt{\sum (y - \overline{y})^2}}$$

$$0.5 = \frac{120}{\sqrt{90}\sqrt{64\,n}}$$

$$\therefore \qquad n = 10$$

Example 14

Calculate the correlation coefficient between x and y from the following data:

$$n = 10, \sum x = 140, \sum y = 150, \sum (x - 10)^2 = 180$$

$$\sum (y - 15)^2 = 215, \sum (x - 10)\,(y - 15) = 60$$

Solution

$$\sum d_x^2 = \sum (x-10)^2 = 180$$

$$\sum d_y^2 = \sum (y-15)^2 = 215$$

$$\sum d_x\, d_y = \sum (x-10)\,(y-15) = 60$$

$$a = 10$$

$$b = 15$$

$$n = 10$$

$$\bar{x} = \frac{\sum x}{n} = \frac{140}{10} = 14$$

$$\bar{y} = \frac{\sum y}{n} = \frac{150}{10} = 15$$

$$\bar{x} = a + \frac{\sum d_x}{n}$$

$$14 = 10 + \frac{\sum d_x}{10}$$

$$\therefore \qquad \sum d_x = 40$$

$$\bar{y} = b + \frac{\sum d_y}{n}$$

$$15 = 15 + \frac{\sum d_y}{10}$$

$$\therefore \qquad \sum d_y = 0$$

$$r = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \frac{\left(\sum d_x\right)^2}{n}}\sqrt{\sum d_y^2 - \frac{\left(\sum d_y\right)^2}{n}}}$$

$$= \frac{60 - \frac{(40)(0)}{10}}{\sqrt{180 - \frac{(40)^2}{10}}\sqrt{215 - \frac{0}{10}}}$$

$$= 0.915$$

Example 15

A computer operator while calculating the coefficient between two variates x and y for 25 pairs of observations obtained the following constants:

$$n = 25, \sum x = 125, \sum x^2 = 650, \sum y = 100,$$
$$\sum y^2 = 460, \sum xy = 508$$

It was later discovered at the time of checking that he had copied down two pairs as (6, 14) *and* (8, 6) *while the correct pairs were* (8,12) *and* (6, 8). *Obtain the correct value of the correlation coefficient.*

Solution

$$n = 25$$

$$\begin{aligned} \text{Corrected} \sum x &= \text{Incorrect} \sum x - (\text{Sum of incorrect } x) + (\text{Sum of correct } x) \\ &= 125 - (6+8) + (8+6) \\ &= 125 \end{aligned}$$

Similarly,

$$\text{Corrected} \sum y = 100 - (14+6) + (12+8) = 100$$
$$\text{Corrected} \sum x^2 = 650 - (6^2 + 8^2) + (8^2 + 6^2) = 650$$
$$\text{Corrected} \sum y^2 = 460 - (14^2 + 6^2) + (12^2 + 8^2) = 436$$
$$\text{Corrected} \sum xy = 508 - (84+48) + (96+48) = 520$$

Correct value of correlation coefficient

$$r = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}}$$

$$= \frac{520 - \dfrac{(125)(100)}{25}}{\sqrt{650 - \dfrac{(125)^2}{25}}\sqrt{436 - \dfrac{(100)^2}{25}}}$$

$$= 0.67$$

EXERCISE 4.1

1. Draw a scatter diagram to represent the following data:

x	2	4	5	6	8	11
y	18	12	10	8	7	5

Calculate the coefficient of correlation between *x* and *y*.

[Ans.: –0.92]

2. Find the coefficient of correlation between x and y for the following data:

x	10	12	18	24	23	27
y	13	18	12	25	30	10

[Ans.: 0.223**]**

3. From the following information relating to the stock exchange quotations for two shares A and B, ascertain by using Pearson's coefficient of correlation how shares A and B are correlated in their prices?

Price share (A) ₹	160	164	172	182	166	170	178
Price share (B) ₹	292	280	260	234	266	254	230

[Ans.: –0.96**]**

4. Find the correlation coefficient between the income and expenditure of a wage earner.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul
Income	46	54	56	56	58	60	62
Expenditure	36	40	44	54	42	58	54

[Ans.: 0.769**]**

5. From the following data, examine whether the input of oil and output of electricity can be said to be correlated.

Input of oil	6.9	8.2	7.8	4.8	9.6	8.0	7.7
Output of Electricity	1.9	3.5	6.5	1.3	5.5	3.5	2.2

[Ans.: 0.696**]**

6. For the following data, show that cov $(x, x^2) = 0$.

x	–3	–2	–1	0	1	2	3
x^2	9	4	1	0	1	4	9

7. Find the coefficient of correlation between x and y for the following data:

x	62	64	65	69	70	71	72	74
y	126	125	139	145	165	152	180	208

[Ans.: 0.9032**]**

8. The following data gave the growth of employment in lacs in the organized sector in India between 1988 and 1995:

Year	1988	1989	1990	1991	1992	1993	1994	1995
Public sector	98	101	104	107	113	120	125	128
Private sector	65	65	67	68	68	69	68	68

Find the correlation coefficient between the employment in public and private sectors.

[Ans.: 0.77]

9. Calculate Karl Pearson's coefficient of correlation from the following data, using 20 as the working mean for price and 70 as working mean for demand.

Price	14	16	17	18	19	20	21	22	23
Demand	84	78	70	75	66	67	62	58	60

[Ans.: −0.954]

10. A sample of 25 pairs of values x and y lead to the following results:

$$\sum x = 127, \sum y = 100, \sum x^2 = 760, \sum y^2 = 449, \sum xy = 500$$

Later on, it was found that two pairs of values were taken as (8, 14) and (8, 6) instead of the correct values (8, 12) and (6, 8). Find the corrected coefficient between x and y.

[Ans.: −0.31]

4.9 RANK CORRELATION

Let a group of n individuals be arranged in order of merit with respect to some characteristics. The same group would give a different order (rank) for different characteristics. Considering the orders corresponding to two characteristics A and B, the correlation between these n pairs of ranks is called the *rank correlation* in the characteristics A and B for that group of individuals.

4.9.1 Spearman's Rank Correlation Coefficient

Let x, y be the ranks of the i^{th} individuals in two characteristics A and B respectively where $i = 1, 2, ..., n$. Assuming that no two individuals have the same rank either for x or y, each of the variables x and y take the values 1, 2, ..., n.

$$\bar{x} = \bar{y} = \frac{1+2+3+\cdots+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

$$\begin{aligned}
\sum (x-\bar{x})^2 &= \sum (x^2 - 2x\,\bar{x} + \bar{x}^2) \\
&= \sum x^2 - 2\bar{x} \sum x + \bar{x}^2 \sum 1 \\
&= \sum x^2 - 2n\bar{x}^2 + n\bar{x}^2 \qquad \left[\because \quad \sum x = n\bar{x} \text{ and } \sum 1 = n\right] \\
&= \sum x^2 - n\,\bar{x}^2 \\
&= (1^2 + 2^2 + \cdots + n^2) - n\left(\frac{n+1}{2}\right)^2
\end{aligned}$$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4}$$

$$= \frac{1}{12}(n^3 - n)$$

Similarly, $\sum (y - \bar{y})^2 = \frac{1}{12}(n^3 - n)$

If d denotes the difference between the ranks of the i^{th} individuals in the two variables,

$$d = x - y = (x - \bar{x}) - (y - \bar{y}) \qquad \left[\because \quad \bar{x} = \bar{y}\right]$$

Squaring and summing over i from 1 to n,

$$\begin{aligned}\sum d^2 &= \sum \left[(x - \bar{x}) - (y - \bar{y})\right]^2 \\ &= \sum (x - \bar{x})^2 + \sum (y - \bar{y})^2 - 2\sum (x - \bar{x})(y - \bar{y})\end{aligned}$$

$$\begin{aligned}\sum (x - \bar{x})(y - \bar{y}) &= \frac{1}{2}\left[\sum (x - \bar{x})^2 + \sum (y - \bar{y})^2 - \sum d^2\right] \\ &= \frac{1}{12}(n^3 - n) - \frac{1}{2}\sum d^2\end{aligned}$$

Hence, the coefficient of correlation between these variables is

$$\begin{aligned}r &= \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \\ &= \frac{\frac{1}{12}(n^3 - n) - \frac{1}{2}\sum d^2}{\frac{1}{12}(n^3 - n)} \\ &= 1 - \frac{6\sum d^2}{n^3 - n} \\ &= 1 - \frac{6\sum d^2}{n(n^2 - 1)}\end{aligned}$$

This is called Spearman's rank correlation coefficient and is denoted by ρ.

Note $\sum d = \sum (x - y) = \sum x - \sum y = n(\bar{x} - \bar{y}) = 0$

Example 1

Ten participants in a contest are ranked by two judges as follows:

x	1	3	7	5	4	6	2	10	9	8
y	3	1	4	5	6	9	7	8	10	2

Calculate the rank correlation coefficient.

Solution

$n = 10$

Rank by first Judge x	Rank by second Judge y	$d = x - y$	d^2
1	3	–2	4
3	1	2	4
7	4	3	9
5	5	0	0
4	6	–2	4
6	9	–3	9
2	7	–5	25
10	8	2	4
9	10	–1	1
8	2	6	36
		$\sum d = 0$	$\sum d^2 = 96$

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

$$= 1 - \frac{6(96)}{10\left[(10)^2 - 1\right]}$$

$$= 0.418$$

Example 2

Ten competitors in a musical test were ranked by the three judges A, B, and C in the following order:

Rank by A	1	6	5	10	3	2	4	9	7	8
Rank by B	3	5	8	4	7	10	2	1	6	9
Rank by C	6	4	9	8	1	2	3	10	5	7

Using the rank correlation method, find which pair of judges has the nearest approach to common liking in music. **[Summer 2015]**

Solution

$n = 10$

Rank by A x	Rank by B y	Rank by C z	$d_1 = x - y$	$d_2 = y - z$	$d_3 = z - x$	d_1^2	d_2^2	d_3^2
1	3	6	–2	–3	5	4	9	25
6	5	4	1	1	–2	1	1	4
5	8	9	–3	–1	4	9	1	16
10	4	8	6	–4	–2	36	16	4
3	7	1	–4	6	–2	16	36	4
2	10	2	–8	8	0	64	64	0
4	2	3	2	–1	–1	4	1	1
9	1	10	8	–9	1	64	81	1
7	6	5	1	1	–2	1	1	4
8	9	7	–1	2	–1	1	4	1
			$\sum d_1 = 0$	$\sum d_2 = 0$	$\sum d_3 = 0$	$\sum d_1^2 = 200$	$\sum d_2^2 = 214$	$\sum d_3^2 = 60$

$$r(x, y) = 1 - \frac{6\sum d_1^2}{n(n^2-1)}$$

$$= 1 - \frac{6(200)}{10\left[(10)^2 - 1\right]}$$

$$= -0.21$$

$$r(y, z) = 1 - \frac{6\sum d_2^2}{n(n^2-1)}$$

$$= 1 - \frac{6(214)}{10\left[(10)^2 - 1\right]}$$

$$= -0.296$$

$$r(z, x) = 1 - \frac{6\sum d_3^2}{n(n^2-1)}$$

$$= 1 - \frac{6(60)}{10\left[(10)^2 - 1\right]}$$

$$= 0.64$$

Since $r(z, x)$ is maximum, the pair of judges A and C has the nearest common approach.

Example 3

Ten students got the following percentage of marks in mathematics and physics:

Mathematics (x)	8	36	98	25	75	82	92	62	65	35
Physics (y)	84	51	91	60	68	62	86	58	35	49

Find the rank correlation coefficient.

Solution

$n = 10$

x	y	Rank in mathematics x	Rank in Physics y	$d = x - y$	d^2
8	84	10	3	7	49
36	51	7	8	–1	1
98	91	1	1	0	0
25	60	9	6	3	9
75	68	4	4	0	0
82	62	3	5	–2	4
92	86	2	2	0	0
62	58	6	7	–1	1
65	35	5	10	–5	25
35	49	8	9	–1	1
				$\sum d = 0$	$\sum d^2 = 90$

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

$$= 1 - \frac{6(90)}{10\left[(10)^2 - 1\right]}$$

$$= 0.455$$

Example 4

The coefficient of rank correlation of the marks obtained by 10 students in physics and chemistry was found to be 0.5. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 3 instead of 7. Find the rank coefficient of the rank correlation.

Solution

$n = 10$

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

$$0.5 = 1 - \frac{6\sum d^2}{10(100-1)}$$

$$\therefore \qquad \sum d^2 = 82.5$$

$$\text{Correct} \sum d^2 = \text{Incorrect} \sum d^2 - (\text{Incorrect rank difference})^2 + (\text{Correct rank difference})^2$$

$$= 82.5 - (3)^2 + (7)^2$$
$$= 122.5$$

Correct coefficient of rank correlation $r = 1 - \frac{6(122.5)}{10(100-1)}$

$$= 0.26$$

4.9.2 Tied Ranks

If there is a tie between two or more individuals ranks, the rank is divided among equal individuals, e.g., if two items have fourth rank, the 4^{th} and 5^{th} rank is divided between them equally and is given as $\frac{4+5}{2} = 4.5^{\text{th}}$ rank to each of them. If three items have the same 4^{th} rank, each of them is given $\frac{4+5+6}{3} = 5^{\text{th}}$ rank. As a result of this, the following adjustment or correction is made in the rank correlation formula. If m is the number of item having equal ranks then the factor $\frac{1}{12}(m^3 - m)$ is added to $\sum d^2$. If there are more than one cases of this type, this factor is added corresponding to each case.

$$r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots\right]}{n(n^2-1)}$$

Example 1

Obtain the rank correlation coefficient from the following data:

x	10	12	18	18	15	40
y	12	18	25	25	50	25

Solution

Here, $n = 6$

x	y	Rank x	Rank y	$d = x - y$	d^2
10	12	1	1	0	0
12	18	2	2	0	0
18	25	4.5	4	0.5	0.25
18	25	4.5	4	0.5	0.25
15	50	3	6	–3	9
40	25	6	4	2	4
					$\sum d^2 = 13.5$

There are two items in the x series having equal values at the rank 4. Each is given the rank 4.5. Similarly, there are three items in the y series at the rank 3. Each of them is given the rank 4.

$$m_1 = 2, m_2 = 3$$

$$r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2)\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[13.50 + \frac{1}{12}(8-2) + \frac{1}{12}(27-3)\right]}{6\left[(6)^2 - 1\right]}$$

$$= 0.5429$$

EXERCISE 4.2

1. Compute Spearman's rank correlation coefficient from the following data:

x	18	20	34	52	12
y	39	23	35	18	46

[Ans.: –0.9]

2. Two judges gave the following ranks to a series of eight one-act plays in a drama competition. Examine the relationship between their judgements.

Judge *A*	8	7	6	3	2	1	5	4
Judge *B*	7	5	4	1	3	2	6	8

[Ans.: 0.62]

3. From the following data, calculate Spearman's rank correlation between x and y.

x	36	56	20	42	33	44	50	15	60
y	50	35	70	58	75	60	45	80	38

[Ans.: 0.92**]**

4. Ten competitors in a voice test are ranked by three judges in the following order:

Rank by First Judge	6	10	2	9	8	1	5	3	4	7
Rank by Second Judge	5	4	10	1	9	3	8	7	2	6
Rank by Third Judge	4	8	2	10	7	6	9	1	3	6

Use the method of rank correlation to gauge which pairs of judges has the nearest approach to common liking in voice.

[Ans.: The first and third judge**]**

5. The following table gives the scores obtained by 11 students in English and Tamil translation. Find the rank correlation coefficient.

Scores in English	40	46	54	60	70	80	82	85	85	90	95
Scores in Tamil	45	45	50	43	40	75	55	72	65	42	70

[Ans.: 0.36**]**

6. Calculate Spearman's coefficient of rank correlation for the following data:

x	53	98	95	81	75	71	59	55
y	47	25	32	37	30	40	39	45

[Ans.: –0.905**]**

7. Following are the scores of ten students in a class and their IQ:

Score	35	40	25	55	85	90	65	55	45	50
IQ	100	100	110	140	150	130	100	120	140	110

Calculate the rank correlation coefficient between the score IQ.

[Ans.: 0.47**]**

4.10 REGRESSION

Regression is defined as a method of estimating the value of one variable when that of the other is known and the variables are correlated. *Regression analysis* is used to predict or estimate one variable in terms of the other variable. It is a highly valuable tool for prediction purpose in economics and business. It is useful in statistical estimation of demand curves, supply curves, production function, cost function, consumption function, etc.

4.11 TYPES OF REGRESSION

Regression is classified into two types:

1. Simple and multiple regressions
2. Linear and nonlinear regressions

4.11.1 Simple and Multiple Regressions

Depending upon the study of the number of variables, regression may be simple or multiple.

1. Simple Regression The regression analysis for studying only two variables at a time is known as simple regression.

2. Multiple Regression The regression analysis for studying more than two variables at a time is known as multiple regression.

4.11.2 Linear and Nonlinear Regressions

Depending upon the regression curve, regression may be linear or nonlinear.

1. Linear Regression If the regression curve is a straight line, the regression is said to be linear.

2. Nonlinear Regression If the regression curve is not a straight line i.e., not a first-degree equation in the variables x and y, the regression is said to be nonlinear or curvilinear. In this case, the regression equation will have a functional relation between the variables x and y involving terms in x and y of the degree higher than one, i.e., involving terms of the type x^2, y^2, x^3, y^3, xy, etc.

4.12 METHODS OF STUDYING REGRESSION

There are two methods of studying correlation:

(i) Method of scatter diagram
(ii) Method of least squares

4.12.1 Method of Scatter Diagram

It is the simplest method of obtaining the lines of regression. The data are plotted on a graph paper by taking the independent variable on the x-axis and the dependent variable on the y-axis. Each of these points are generally scattered in a narrow strip. If the correlation is perfect, i.e., if r is equal to one, positive, or negative, the points will lie on a line which is the line of regression.

4.12.2 Method of Least Squares

This is a mathematical method which gives an objective treatment to find a line of regression. It is used for obtaining the equation of a curve which fits best to a given set of observations. It is based on the assumption that the sum of squares of differences between the estimated values and the actual observed values of the observations is minimum.

4.13 LINES OF REGRESSION

If the variables, which are highly correlated, are plotted on a graph then the points lie in a narrow strip. If all the points in the scatter diagram cluster around a straight line, the line is called the *line of regression*. The line of regression is the line of best fit and is obtained by the principle of least squares.

Line of Regression of *y* on *x*

It is the line which gives the best estimate for the values of y for any given values of x. The regression equation of y on x is given by

$$y - \overline{y} = r\frac{\sigma_y}{\sigma_x}(x - \overline{x})$$

It is also written as

$$y = a + bx$$

Line of Regression of *x* on *y*

It is the line which gives the best estimate for the values of x for any given values of y. The regression equation for x on y is given by

$$x - \overline{x} = r\frac{\sigma_x}{\sigma_y}(y - \overline{y})$$

It is also written as

$$x = a + by$$

where $\overline{x}$ and $\overline{y}$ are means of x series and y series respectively, σ_x and σ_y are standard deviations of x series and y series respectively, r is the correlation coefficient between x and y.

4.14 REGRESSION COEFFICIENTS

The slope b of the line of regression of y on x is also called the *coefficient of regression* of y on x. It represents the increment in the value of y corresponding to a unit change in the value of x.

$$b_{yx} = \text{Regression coefficient of } y \text{ on } x$$

$$= r\frac{\sigma_y}{\sigma_x}$$

Similarly, the slope b of the line of regression of x on y is called the coefficient of regression of x on y. It represents the increment in the value of x corresponding to a unit change in the value of y.

$$b_{xy} = \text{Regression coefficient of } x \text{ on } y$$
$$= r\,\frac{\sigma_x}{\sigma_y}$$

Expressions for Regression Coefficients

(i) We know that

$$r = \frac{\sum (x-\bar{x})\,(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\;\sqrt{\sum (y-\bar{y})^2}}$$

$$\sigma_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum (y-\bar{y})^2}{n}}$$

$$b_{yx} = r\,\frac{\sigma_y}{\sigma_x}$$
$$= \frac{\sum (x-\bar{x})\,(y-\bar{y})}{\sum (x-\bar{x})^2}$$

and $$b_{xy} = r\,\frac{\sigma_x}{\sigma_y}$$
$$= \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (y-\bar{y})^2}$$

(ii) We know that

$$r = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}\;\sqrt{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}}$$

$$\sigma_x = \sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}$$

$$\sigma_y = \sqrt{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}$$

$$b_{yx} = r\,\frac{\sigma_y}{\sigma_x}$$

$$= \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}$$

and $$b_{xy} = r\,\frac{\sigma_x}{\sigma_y}$$

$$= \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}$$

(iii) We know that

$$r = \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \dfrac{\left(\sum d_x\right)^2}{n}}\,\sqrt{\sum d_y^2 - \dfrac{\left(\sum d_y\right)^2}{n}}}$$

$$\sigma_x = \sqrt{\sum d_x^2 - \frac{\left(\sum d_x\right)^2}{n}}$$

$$\sigma_y = \sqrt{\sum d_y^2 - \frac{\left(\sum d_y\right)^2}{n}}$$

$$b_{yx} = r\,\frac{\sigma_y}{\sigma_x}$$

$$= \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \dfrac{\left(\sum d_x\right)^2}{n}}}$$

and $$b_{xy} = r\,\frac{\sigma_x}{\sigma_y}$$

$$= \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_y^2 - \dfrac{\left(\sum d_y\right)^2}{n}}}$$

4.15 PROPERTIES OF REGRESSION COEFFICIENTS

1. The coefficient of correlation is the geometric mean of the coefficients of regression, i.e., $r = \sqrt{b_{yx} b_{xy}}$.

Proof We know that

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$b_{yx}\, b_{xy} = r\frac{\sigma_y}{\sigma_x}\cdot r\frac{\sigma_x}{\sigma_y}$$

$$= r^2$$

$$r = \sqrt{b_{yx}\, b_{xy}}$$

2. If one of the regression coefficients is greater than one, the other must be less than one.

Proof Let $b_{yx} > 1$

We know that

$$r^2 \le 1 \text{ and } r^2 = b_{yx}\, b_{xy}$$

$$b_{yx}\, b_{xy} \le 1$$

$$b_{yx} \le \frac{1}{b_{xy}}$$

Hence, if $b_{yx} < 1$, $b_{xy} > 1$

3. The arithmetic mean of regression coefficients is greater than or equal to the coefficient of correlation.

Proof We have to prove that

$$\frac{1}{2}(b_{yx} + b_{xy}) \ge r$$

i.e., $$\frac{1}{2}\left(r\frac{\sigma_y}{\sigma_x} + r\frac{\sigma_x}{\sigma_y}\right) \ge r$$

i.e., $$\frac{\sigma_y}{\sigma_x} + \frac{\sigma_x}{\sigma_y} \ge 2$$

i.e., $$\sigma_y^2 + \sigma_x^2 - 2\sigma_x\sigma_y \ge 0$$

i.e., $(\sigma_y - \sigma_x)^2 \geq 0$

which is always true, since the square of a real quantity is $1 \geq 0$.

4. Regression Coefficients are independent of the change of origin but not of scale.

Proof Let
$$d_x = \frac{x-a}{h},\ d_y = \frac{y-b}{k}$$
$$x = a + hd_x,\ y = b + kd_y$$

where a, b, h (> 0) and k(> 0) are constants.

$$r_{d_x d_y} = r_{xy},\ \sigma_{d_x}^2 = \frac{1}{h^2}\sigma_x^2,\ \sigma_{d_y}^2 = \frac{1}{k^2}\sigma_y^2$$

$$\begin{aligned} b_{d_x d_y} &= r_{d_x d_y}\frac{\sigma_{d_x}}{\sigma_{d_y}} \\ &= r_{xy}\frac{\sigma_x}{h}\frac{k}{\sigma_y} \\ &= \frac{k}{h} r_{xy}\frac{\sigma_x}{\sigma_y} \\ &= \frac{k}{h} b_{xy} \end{aligned}$$

Similarly, $$b_{d_y d_x} = \frac{h}{k} b_{yx}$$

5. Both regression coefficients will have the same sign i.e., either both are positive or both are negative.

6. The sign of correlation is same as that of the regression coefficients, i.e., $r > 0$ if $b_{xy} > 0$ and $b_{yx} > 0$; and $r < 0$ if $b_{xy} < 0$ and $b_{yx} < 0$.

4.16 PROPERTIES OF LINES OF REGRESSION (LINEAR REGRESSION)

1. The two regression lines x on y and y on x always intersect at their means $(\bar{x}, \bar{y})$.
2. Since $r^2 = b_{yx}\, b_{xy}$, i.e., $r = \sqrt{b_{yx}\, b_{xy}}$, therefore, r, b_{yx}, b_{xy} all have the same sign.
3. If $r = 0$, the regression coefficients are zero.
4. The regression lines become identical if $r = \pm 1$. It follows from the regression equations that $x = \bar{x}$ and $y = \bar{y}$. If $r = 0$, these lines are perpendicular to each other.

Example 1

The regression lines of a sample are $x + 6y = 6$ *and* $3x + 2y = 10$*. Find*
(i) sample means $\bar{x}$ and $\bar{y}$, *and*
(ii) the coefficient of correlation between x and y.
(iii) Also estimate y when $x = 12$*.*

Solution

(i) The regression lines pass through the point $(\bar{x}, \bar{y})$.

$$\bar{x} + 6\bar{y} = 6 \qquad \text{...(1)}$$

$$3\bar{x} + 2\bar{y} = 10 \qquad \text{...(2)}$$

Solving Eqs (1) and (2),

$$\bar{x} = 3, \quad \bar{y} = \frac{1}{2}$$

(ii) Let the line $x + 6y = 6$ be the line of regression of y on x.

$$6y = -x + 6$$

$$y = -\frac{1}{6}x + 1$$

$$\therefore \qquad b_{yx} = -\frac{1}{6}$$

Let the line $3x + 2y = 10$ be the line of regression of x on y.

$$3x = -2y + 10$$

$$x = -\frac{2}{3}y + \frac{10}{3}$$

$$\therefore \qquad b_{xy} = -\frac{2}{3}$$

$$r = \sqrt{b_{yx}\, b_{xy}} = \sqrt{\left(-\frac{1}{6}\right)\left(-\frac{2}{3}\right)} = \frac{1}{3}$$

Since b_{yx} and b_{xy} are negative, r is negative.

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

Estimated value of y when $x = 12$ is

$$y = -\frac{1}{6}(12) + 1 = -1$$

Example 2

If the two lines of regression are $4x - 5y + 30 = 0$ *and* $20x - 9y - 107 = 0$, *which of these are lines of regression of x on y and y on x? Find* r_{xy} *and* σ_y *when* $\sigma_x = 3$.

Solution

For the line $4x - 5y + 30 = 0,$

$$-5y = -4x - 30$$

$$y = 0.8x + 6$$

$$\therefore \quad b_{yx} = 0.8$$

For the line $20x - 9y - 107 = 0$

$$20x = 9y + 107$$

$$x = 0.45y + 5.35$$

$$\therefore \quad b_{xy} = 0.45$$

Both b_{yx} and b_{xy} are positive.

Hence, line $4x - 5y + 30 = 0$ is the line of regression of y one x and line $20x - 9y - 107 = 0$ is the line of regression of x on y.

$$r = \sqrt{b_{yx}\, b_{xy}} = \sqrt{(0.8)(0.45)} = 0.6$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$0.8 = 0.6\left(\frac{\sigma_y}{3}\right)$$

$$\therefore \quad \sigma_y = 4$$

Example 3

The following data regarding the heights (y) *and weights* (x) *of* 100 *college students are given:*

$$\sum x = 15000, \quad \sum x^2 = 2272500, \quad \sum y = 6800$$

$$\sum y^2 = 463025, \quad \sum xy = 1022250$$

Find the coefficient of correlation between height and weight and also the equation of regression of height and weight.

Solution

$$n = 100$$

$$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}$$

$$= \frac{1022250 - \frac{(15000)(6800)}{100}}{2272500 - \frac{(15000)^2}{100}}$$

$$= 0.1$$

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}$$

$$= \frac{1022250 - \frac{(15000)(6800)}{100}}{463025 - \frac{(6800)^2}{100}}$$

$$= 3.6$$

$$r = \sqrt{b_{yx}\, b_{xy}} = \sqrt{(0.1)(3.6)} = 0.6$$

$$\overline{x} = \frac{\sum x}{n} = \frac{15000}{100} = 150$$

$$\overline{y} = \frac{\sum y}{n} = \frac{6800}{100} = 68$$

The equation of the line of regression of y on x is

$$y - \overline{y} = b_{yx}\,(x - \overline{x})$$

$$y - 68 = 0.1(x - 150)$$

$$y = 0.1x + 53$$

The equation of the line of regression of x on y is

$$x - \overline{x} = b_{xy}(y - \overline{y})$$

$$x - 150 = 3.6(y - 68)$$

$$x = 3.6y - 94.8$$

Example 4

For a bivariate data, the mean value of x is 20 and the mean value of y is 45. The regression coefficient of y on x is 4 and that of x on y is $\frac{1}{9}$. *Find*

(i) the coefficient of correlation, and

(ii) the standard deviation of x if the standard deviation of y is 12.

(iii) Also write down the equations of regression lines.

Solution

$$\bar{x} = 20, \quad \bar{y} = 45, \quad b_{yx} = 4, \quad b_{xy} = \frac{1}{9}$$

(i) $$r = \sqrt{b_{yx}\, b_{xy}} = \sqrt{(4)\left(\frac{1}{9}\right)} = \frac{2}{3} = 0.667$$

(ii) $$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$4 = \frac{2}{3}\left(\frac{12}{\sigma_x}\right)$$

$$\therefore \quad \sigma_x = 2$$

(iii) The equation of the regression line of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 45 = 4(x - 20)$$

$$y = 4x - 35$$

The equation of the regression line of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 20 = \frac{1}{9}(y - 45)$$

$$x = \frac{1}{9}y + 15$$

Example 5

From the following results, obtain the two regression equations and estimate the yield when the rainfall is 29 cm and the rainfall, when the yield is 600 kg:

	Yield in kg	Rainfall in cm
Mean	508.4	26.7
SD	36.8	4.6

The coefficient of correlation between yield and rainfall is 0.52.

Solution

Let rainfall in cm be denoted by x and yield in kg be denoted by y.

$$\bar{x} = 26.7, \quad \bar{y} = 508.4, \quad \sigma_x = 4.6, \quad \sigma_y = 36.8, \quad r = 0.52$$

$$\begin{aligned} b_{yx} &= r\frac{\sigma_y}{\sigma_x} \\ &= 0.52\left(\frac{36.8}{4.6}\right) \\ &= 4.16 \end{aligned}$$

$$\begin{aligned} b_{xy} &= r\frac{\sigma_x}{\sigma_y} \\ &= 0.52\left(\frac{4.6}{36.8}\right) \\ &= 0.065 \end{aligned}$$

The equation of the line of regression of y on x is

$$\begin{aligned} y - \bar{y} &= b_{yx}(x - \bar{x}) \\ y - 508.4 &= 4.16(x - 26.7) \\ y &= 4.16x + 397.328 \end{aligned}$$

The equation of the line of regression of x on y is

$$\begin{aligned} x - \bar{x} &= b_{xy}(y - \bar{y}) \\ x - 26.7 &= 0.065(y - 508.4) \\ x &= 0.065y - 6.346 \end{aligned}$$

Estimated yield when the rainfall is 29 cm is

$$y = 4.16\,(29) + 397.328 = 517.968 \text{ kg}$$

Estimated rainfall when the yield is 600 kg is

$$x = 0.065\,(600) - 6.346 = 32.654 \text{ cm}$$

Example 6

Find the regression coefficients b_{yx} and b_{xy} and hence, find the correlation coefficient between x and y for the following data:

x	4	2	3	4	2
y	2	3	2	4	4

Solution

$n = 5$

x	y	x^2	y^2	xy
4	2	16	4	8
2	3	4	9	6
3	2	9	4	6
4	4	16	16	16
2	4	4	16	8
$\Sigma x = 15$	$\Sigma y = 15$	$\Sigma x^2 = 49$	$\Sigma y^2 = 49$	$\Sigma xy = 44$

$$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}$$

$$= \frac{44 - \frac{(15)(15)}{5}}{49 - \frac{(15)^2}{5}}$$

$$= -0.25$$

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}$$

$$= \frac{44 - \frac{(15)(15)}{5}}{49 - \frac{(15)^2}{5}}$$

$$= -0.25$$

$$r = \sqrt{b_{yx}\, b_{xy}} = \sqrt{(-0.25)(-0.25)} = 0.25$$

Since b_{yx} and b_{xy} are negative, r is negative.

$$r = -0.25$$

Example 7

The following data give the experience of machine operators and their performance rating as given by the number of good parts turned out per 100 pieces.

Operator	1	2	3	4	5	6
Performance rating (x)	23	43	53	63	73	83
Experience (y)	5	6	7	8	9	10

Calculate the regression line of performance rating on experience and also estimate the probable performance if an operator has 11 *years of experience.* **[Summer 2015]**

Solution

$$n = 6$$

x	y	y^2	xy
23	5	25	115
43	6	36	258
53	7	49	371
63	8	64	504
73	9	81	657
83	10	100	830
$\sum x = 338$	$\sum y = 45$	$\sum y^2 = 355$	$\sum xy = 2735$

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}$$

$$= \frac{2735 - \frac{(338)(45)}{6}}{355 - \frac{(45)^2}{6}}$$

$$= 11.429$$

$$\bar{x} = \frac{\sum x}{n} = \frac{338}{6} = 56.33$$

$$\bar{y} = \frac{\sum y}{n} = \frac{45}{6} = 7.5$$

The equation of regression line of x on y is

$$x - \bar{x} = b_{xy}\ (y - \bar{y})$$

$$x - 56.33 = 11.429(y - 7.5)$$

$$x = 11.429\ y - 29.3875$$

Estimated performance if $y = 11$ is

$$x = 11.429(11) - 29.3875 = 96.3315$$

Example 8

The number of bacterial cells (y) per unit volume in a culture at different hours (x) is given below:

x	0	1	2	3	4	5	6	7	8	9
y	43	46	82	98	123	167	199	213	245	272

Fit lines of regression of y on x and x on y. Also, estimate the number of bacterial cells after 15 *hours.*

Solution

$n = 10$

x	y	x^2	xy	y^2
0	43	0	0	1849
1	46	1	46	2116
2	82	4	164	6724
3	98	9	294	9604
4	123	16	492	15129
5	167	25	835	27889
6	199	36	1194	39601
7	213	49	1491	45369
8	245	64	1960	60025
9	272	81	2448	73984
$\sum x = 45$	$\sum y = 1488$	$\sum x^2 = 285$	$\sum xy = 8924$	$\sum y^2 = 282290$

$$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}$$

$$= \frac{8924 - \frac{(45)(1488)}{10}}{285 - \frac{(45)^2}{10}}$$

$$= 27.0061$$

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}$$

$$= \frac{8924 - \frac{(45)(1488)}{10}}{282290 - \frac{(1488)^2}{10}}$$

$$= 0.0366$$

$$\bar{x} = \frac{\sum x}{n} = \frac{45}{10} = 4.5$$

$$\bar{y} = \frac{\sum y}{n} = \frac{1488}{10} = 148.8$$

The equation of the line of regression of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 148.8 = 27.0061\,(x - 4.5)$$

$$y = 27.0061x + 27.2726$$

The equation of the line of regression of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 4.5 = 0.0366(y - 148.8)$$

$$x = 0.366\,y - 0.9461$$

At $x = 15$ hours,

$y = 27.0061\ (15) + 27.2726 = 432.3641$

Example 9

Find the regression coefficient of y on x for the following data:

x	1	2	3	4	5
y	160	180	140	180	200

Solution

$$n = 5$$

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{\sum y}{n} = \frac{860}{5} = 172$$

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
1	160	–2	–12	4	24
2	180	–1	8	1	–8
3	140	0	–32	0	0
4	180	1	8	1	8
5	200	2	28	4	56
$\sum x = 15$	$\sum y = 860$	$\sum(x - \bar{x}) = 0$	$\sum(y - \bar{y}) = 0$	$\sum(x - \bar{x})^2 = 10$	$\sum(x - \bar{x})(y - \bar{y}) = 80$

$$\begin{aligned} b_{yx} &= \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} \\ &= \frac{80}{10} \\ &= 8 \end{aligned}$$

Example 10

Calculate the two regression coefficients from the data and find correlation coefficient.

x	7	4	8	6	5
y	6	5	9	8	2

Solution

$$n = 5$$

$$\bar{x} = \frac{\sum x}{n} = \frac{30}{5} = 6$$

$$\bar{y} = \frac{\sum y}{n} = \frac{30}{5} = 6$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
7	6	1	0	1	0	0
4	5	–2	–1	4	1	2
8	9	2	3	4	9	6
6	8	0	2	0	4	0
5	2	–1	–4	1	16	4
$\sum x = 30$	$\sum y = 30$	$\sum(x-\bar{x}) = 0$	$\sum(y-\bar{y}) = 0$	$\sum(x-\bar{x})^2 = 10$	$\sum(y-\bar{y})^2 = 30$	$\sum(x-\bar{x})(y-\bar{y}) = 12$

$$b_{yx} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$= \frac{12}{10}$$

$$= 1.2$$

$$b_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}$$

$$= \frac{12}{30}$$

$$= 0.4$$

$$r = \sqrt{b_{yx}\, b_{xy}} = \sqrt{(1.2)(0.4)} = 0.693$$

Example 11

Obtain the two regression lines from the following data and hence, find the correlation coefficient.

x	6	2	10	4	8
y	9	11	5	8	7

[Summer 2015]

Solution

$$n = 5$$

$$\bar{x} = \frac{\sum x}{n} = \frac{30}{5} = 6$$

$$\bar{y} = \frac{\sum y}{n} = \frac{40}{5} = 8$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
6	9	0	1	0	1	0
2	11	–4	3	16	9	–12
10	5	4	–3	16	9	–12
4	8	–2	0	4	0	0
8	7	2	–1	4	1	–2
$\sum x = 30$	$\sum y = 40$	$\sum(x-\bar{x}) = 0$	$\sum(y-\bar{y}) = 0$	$\sum(x-\bar{x})^2 = 40$	$\sum(y-\bar{y})^2 = 20$	$\sum(x-\bar{x})(y-\bar{y})^2 = -26$

$$b_{yx} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{-26}{40} = -0.65$$

$$b_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2} = \frac{-26}{20} = -1.3$$

The equation of regression line of y on x is

$$y-\bar{y} = b_{yx}(x-\bar{x})$$
$$y-8 = -0.65(x-6)$$
$$y = -0.65\,x + 11.9$$

The equation of regression line of x on y is

$$x-\bar{x} = b_{xy}(y-\bar{y})$$
$$x-6 = -1.3(y-8)$$
$$x = -1.3\,y + 16.4$$
$$r = \sqrt{b_{yx}\,b_{xy}} = \sqrt{(-0.65)(-1.3)} = 0.9192$$

Since b_{yx} and b_{xy} are negative, r is negative.

$$r = -0.9192.$$

Example 12

Calculate the regression coefficients and find the two lines of regression from the following data:

x	57	58	59	59	60	61	62	64
y	67	68	65	68	72	72	69	71

Find the value of y when x = 66.

Solution

$$n = 8$$

$$\overline{x} = \frac{\sum x}{n} = \frac{480}{8} = 60$$

$$\overline{y} = \frac{\sum y}{n} = \frac{552}{8} = 69$$

x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
57	67	–3	–2	9	4	6
58	68	–2	–1	4	1	2
59	65	–1	–4	1	16	4
59	68	–1	–1	1	1	1
60	72	0	3	0	9	0
61	72	1	3	1	9	3
62	69	2	0	4	0	0
64	71	4	2	16	4	8
$\sum x = 480$	$\sum y = 552$	$\sum(x-\overline{x}) = 0$	$\sum(y-\overline{y}) = 0$	$\sum(x-\overline{x})^2 = 36$	$\sum(y-\overline{y})^2 = 44$	$\sum(x-\overline{x})(y-\overline{y}) = 24$

$$\begin{aligned} b_{yx} &= \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2} \\ &= \frac{24}{36} \\ &= 0.667 \end{aligned}$$

$$b_{xy} = \frac{\sum (x-\overline{x})(y-\overline{y})}{\sum (y-\overline{y})^2}$$
$$= \frac{24}{44}$$
$$= 0.545$$

The equation of regression line of y on x is

$$y-\overline{y} = b_{yx}\,(x-\overline{x})$$
$$y-69 = 0.667(x-60)$$
$$y = 0.667x + 28.98$$

The equation of regression line of x on y is

$$x-\overline{x} = b_{xy}\,(y-\overline{y})$$
$$x-60 = 0.545(y-69)$$
$$x = 0.545y + 22.395$$

Value of y when $x = 66$ is

$$y = 0.667\ (66) + 28.98 = 73.002$$

Example 13

The following data represents rainfall (x) and yield of paddy per hectare (y) in a particular area. Find the linear regression of x on y.

x	113	102	95	120	140	130	125
y	1.8	1.5	1.3	1.9	1.1	2.0	1.7

Solution

Let $a = 120$ and $b = 1.8$ be the assumed means of x and y series respectively.

$$d_x = x - a = x - 120$$
$$d_y = y - b = y - 1.8$$
$$n = 7$$

x	y	d_x	d_y	d_y^2	$d_x d_y$
113	1.8	–7	0	0	0
102	1.5	–18	–0.3	0.09	5.4
95	1.3	–25	–0.5	0.25	12.5
120	1.9	0	0.1	0.01	0
140	1.1	20	–0.7	0.49	–14
130	2.0	10	0.2	0.04	2.0
125	1.7	5	–0.1	0.01	–0.5
$\sum x = 825$	$\sum y = 11.3$	$\sum d_x = -15$	$\sum d_y = -1.3$	$\sum d_y^2 = 0.89$	$\sum d_x d_y = 5.4$

$$b_{xy} = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{n}}{\sum d_y^2 - \frac{\left(\sum d_y\right)^2}{n}}$$

$$= \frac{5.4 - \frac{(-15)(-1.3)}{7}}{0.89 - \frac{(-1.3)^2}{7}}$$

$$= 4.03$$

$$\overline{x} = \frac{\sum x}{n} = \frac{825}{7} = 117.86$$

$$\overline{y} = \frac{\sum y}{n} = \frac{11.3}{7} = 1.614$$

The equation of the regression line of x on y is

$$x - \overline{x} = b_{xy}(y - \overline{y})$$

$$x - 117.86 = 4.03\,(y - 1.614)$$

$$x = 4.03\,y + 111.36$$

Example 14

Find the two lines of regression from the following data:

Age of husband (x)	25	22	28	26	35	20	22	40	20	18
Age of wife (y)	18	15	20	17	22	14	16	21	15	14

Hence, estimate (i) the age of the husband when the age of the wife is 19, *and (ii) the age of the wife when the age of the husband is* 30.

Solution

Let $a = 26$ and $b = 17$ be the assumed means of x and y series respectively.

$$d_x = x - a = x - 26$$
$$d_y = y - b = y - 17$$
$$n = 10$$

x	y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
25	18	–1	1	1	1	–1
22	15	–4	–2	16	4	8
28	20	2	3	4	9	6
26	17	0	0	0	0	0
35	22	9	5	81	25	45
20	14	–6	–3	36	9	18
22	16	–4	–1	16	1	4
40	21	14	4	196	16	56
20	15	–6	–2	36	4	12
18	14	–8	–3	64	9	24
$\sum x = 256$	$\sum y = 172$	$\sum d_x = -4$	$\sum d_y = 2$	$\sum d_x^2 = 450$	$\sum d_y^2 = 78$	$\sum d_x d_y = 172$

$$b_{yx} = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{n}}{\sum d_x^2 - \frac{\left(\sum d_x\right)^2}{n}}$$

$$= \frac{172 - \frac{(-4)(2)}{10}}{450 - \frac{(-4)^2}{10}}$$

$$= 0.385$$

$$b_{xy} = \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sum d_y^2 - \dfrac{\left(\sum d_y\right)^2}{n}}$$

$$= \frac{172 - \dfrac{(-4)(2)}{10}}{78 - \dfrac{(2)^2}{10}}$$

$$= 2.227$$

$$\bar{x} = \frac{\sum x}{n} = \frac{256}{10} = 25.6$$

$$\bar{y} = \frac{\sum y}{n} = \frac{172}{10} = 17.2$$

The equation of the regression line of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 17.2 = 0.385(x - 25.6)$$

$$y = 0.385x + 7.344$$

The equation of the regression line of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 25.6 = 2.227(y - 17.2)$$

$$x = 2.227y - 12.704$$

Estimated age of the husband when the age of the wife is 19 is

$$x = 2.227\ (19) - 12.704 = 29.601 \text{ or } 30 \text{ nearly}$$

Age of the husband = 30 years

Estimated age of the wife when the age of the husband is 30 is

$$y = 0.385\ (30) + 7.344 = 18.894 \text{ or } 19 \text{ nearly}$$

Age of the wife = 19 years

Example 15

From the following data, obtain the two regression lines and correlation coefficient.

Sales (x)	100	98	78	85	110	93	80
Purchase (y)	85	90	70	72	95	81	74

Solution

Let $a = 93$ and $b = 81$ be the assumed means of x and y series respectively.

$$d_x = x - a = x - 93$$
$$d_y = y - b = y - 91$$
$$n = 7$$

x	y	d_x	d_y	d_x^2	d_y^2	d_xd_y
100	85	7	4	49	16	28
98	90	5	9	25	81	45
78	70	–15	–11	225	121	165
85	72	–8	–9	64	81	72
110	95	17	14	289	196	238
93	81	0	0	0	0	0
80	74	–13	–7	169	49	91
$\sum x = 644$	$\sum y = 567$	$\sum d_x = -7$	$\sum d_y = 0$	$\sum d_x^2 = 821$	$\sum d_y^2 = 544$	$\sum d_xd_y = 639$

$$b_{yx} = \frac{\sum d_xd_y - \dfrac{\sum d_x \sum d_y}{n}}{\sum d_x^2 - \dfrac{\left(\sum d_x\right)^2}{n}}$$

$$= \frac{639 - \dfrac{(-7)(0)}{7}}{821 - \dfrac{(-7)^2}{7}}$$

$$= 0.785$$

$$b_{xy} = \frac{\sum d_xdy - \dfrac{\sum d_x \sum d_y}{n}}{\sum d_y^2 - \dfrac{\left(\sum d_y\right)^2}{n}}$$

$$= \frac{639 - \dfrac{(-7)(0)}{7}}{544 - \dfrac{(0)^2}{7}}$$

$$= 1.1746$$

$$\bar{x} = \frac{\sum x}{n} = \frac{644}{7} = 92$$
$$\bar{y} = \frac{\sum y}{n} = \frac{567}{7} = 81$$

The equation of regression line of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$
$$y - 81 = 0.785(x - 92)$$
$$y = 0.785x + 8.78$$

The equation of regression line of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$
$$x - 92 = 1.1746(y - 81)$$
$$x = 1.1746 - 3.1426$$
$$r = \sqrt{b_{yx}\, b_{xy}} = \sqrt{(0.785)(1.1746)} = 0.9602$$

EXERCISE 4.3

1. The following are the lines of regression $4y = x + 38$ and $9y = x + 288$. Estimate y when $x = 99$ and x when $y = 30$. Also, find the means of x and y.

   [Ans.: $y = 43$, $x = 82$, $\bar{x} = 162$, $\bar{y} = 50$ **]**

2. The equations of the two lines of regression are $x = 19.13 - 0.87\,y$ and $y = 11.64 - 0.50\,x$. Find (i) the means of x and y, and (ii) the coefficient of correlation between x and y.

   [Ans.: $\bar{x} = 15.79$, $\bar{y} = 3.74$, (ii) $r = -0.66$, $b_{yx} = -0.5$, $b_{xy} = 0.87$ **]**

3. Given var(x) = 25. The equations of the two lines of regression are $5x - y = 22$ and $64\,x - 45\,y = 24$. Find (i) $\bar{x}$ and $\bar{y}$, (ii) r, and (iii) σ_y.

   [Ans.: $\bar{x} = 6$, $\bar{y} = 8$, (ii) $r = 1.87$ (iii) $\sigma_y = 0.2$ **]**

4. In a partially destroyed laboratory record of analysis of correlation data the following results are legible. Variance = 9, the equations of the lines of regression $4x - 5y + 33 = 0$, $20\,x - 9\,y - 107 = 0$. Find (i) the mean values of x and y, (ii) the standard deviation of y, and (iii) the coefficient of correlation between x and y

   [Ans.: (i) $\bar{x} = 13$, $\bar{y} = 17$, (ii) $\sigma_y = 4$, (iii) $r = 0.6$ **]**

5. From a sample of 200 pairs of observation, the following quantities were calculated:

$$\sum x = 11.34, \sum y = 20.78, \sum x^2 = 12.16, \sum y^2 = 84.96, \sum xy = 22.13$$

From the above data, show how to compute the coefficients of the equation $y = a + bx$.

[Ans.: $a = 0.0005, b = 1.82$ **]**

6. In the estimation of regression equations of two variables x and y, the following results were obtained:

$$\bar{x} = 90, \bar{y} = 70, n = 10, \Sigma(x - \bar{x})^2 = 6360, \Sigma(y - \bar{y})^2 = 2860$$
$$\Sigma(x - \bar{x})(y - \bar{y}) = 3900$$

Obtain the two lines of regression.

[Ans.: $x = 1.361\, y - 5.27$, $y = 0.613\, x + 14.812$**]**

7. Find the likely production corresponding to a rainfall of 40 cm from the following data:

	Rainfall (in cm)	Output (in quintals)
mean	30	50
SD	5	10
$r = 0.8$		

[Ans.: 66 quintals**]**

8. The following table gives the age of a car of a certain make and annual maintenance cost. Obtain the equation of the line of regression of cost on age.

Age of a car	2	4	6	8
Maintenance	1	2	2.5	3

[Ans.: $x = 0.325\, y + 0.5$**]**

9. Obtain the equation of the line of regression of y on x from the following data and estimate y for $x = 73$.

x	70	72	74	76	78	80
y	163	170	179	188	196	220

[Ans.: $y = 5.31\, x - 212.57$, $y = 175.37$**]**

10. The heights in cm of fathers (x) and of the eldest sons (y) are given below:

x	165	160	170	163	173	158	178	168	173	170	175	180
y	173	168	173	165	175	168	173	165	180	170	173	178

Estimate the height of the eldest son if the height of the father is 172 cm and the height of the father if the height of the eldest son is 173 cm. Also, find the coefficient of correlation between the heights of fathers and sons.

[Ans.: (i) $y = 1.016\,x - 5.123$ (ii) $x = 0.476\,y + 98.98$

(iii) 169.97, 173.45 (iv) $r = 0.696$**]**

11. Find (i) the lines of regression, and (ii) coefficient of correlation for the following data:

x	65	66	67	67	68	69	70	72
y	67	68	65	66	72	72	69	71

[Ans.: (i) $y = 19.64 + 0.72\,x$, $x = 33.29 + 0.5\,y$, (ii) $r = 0.604$**]**

12. Find the line of regression for the following data and estimate y corresponding to $x = 15.5$.

x	10	12	13	16	17	20	25
y	19	22	24	27	29	33	37

[Ans.: $y = 1.21x + 7.71$, $y = 26.465$**]**

13. The following data give the heights in inches (x) and weights in lbs (y) of a random sample of 10 students:

x	61	68	68	64	65	70	63	62	64	67
y	112	123	130	115	110	125	100	113	116	126

Estimate the weight of a student of height 59 inches.

[Ans.: 126.4 lbs**]**

14. Find the regression equations of y on x from the data given below taking deviations from actual mean of x and y.

Price in rupees (x)	10	12	13	12	16	15
Demand (y)	40	38	43	45	37	43

Estimate the demand when the price is ₹20.

[Ans.: $y = -0.25\,x + 44.25$, $y = 39.25$**]**

Points to Remember

Karl Pearson's Coefficient of Correlation

(i) $$r = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}$$

(ii) $$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}}$$

(iii) $$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}}$$

(iv) $$r = \frac{\sum d_x d_y - \frac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \frac{\left(\sum d_x\right)^2}{n}}\sqrt{\sum d_y^2 - \frac{\left(\sum d_y\right)^2}{n}}}$$

Spearman's Rank Correlation Coefficient

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

Spearman's Rank Correlation Coefficient for Tied Ranks

$$r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots\right]}{n(n^2-1)}$$

Lines of Regression

Line of Regression of *y* on *x*

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

It is also written as

$$y = a + bx$$

Line of Regression of *x* on *y*

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

It is also written as

$$x = a + by$$

Regression Coefficients

$$b_{yx} = r\,\frac{\sigma_y}{\sigma_x}$$

$$b_{xy} = r\,\frac{\sigma_x}{\sigma_y}$$

Expressions for Regression Coefficients

(i) $$b_{yx} = \frac{\sum (x-\overline{x})(y-\overline{y})}{\sum (x-\overline{x})^2}$$

and $$b_{xy} = \frac{\sum (x-\overline{x})(y-\overline{y})}{\sum (y-\overline{y})^2}$$

(ii) $$b_{yx} = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}$$

and $$b_{xy} = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}$$

(iii) $$b_{yx} = \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_x^2 - \dfrac{\left(\sum d_x\right)^2}{n}}}$$

and $$b_{xy} = \frac{\sum d_x d_y - \dfrac{\sum d_x \sum d_y}{n}}{\sqrt{\sum d_y^2 - \dfrac{\left(\sum d_y\right)^2}{n}}}$$