Probability and Statistics

Subject Code : (BTCOC402)

Second Year Engineering

Semester IV - Computer Engineering / Computer Science and Engineering / Computer Science

(Dr. Babasaheb Ambedkar Technological University)

Strictly as per the New Syllabus of Dr. Babasaheb Ambedkar Technological University w.e.f. academic year 2018-2019

SANTOSH R. MITKARI
M.sc Mathematics
Assistant Professor
F.E. Coordinator
Department of Engineering Sciences and Allied Engineering,
Bharati Vidyapeeth's College of Engineering for Women, Pune – 411 043
Maharashtra, India

MAHESH V.GHOTKAR
M.sc Mathematics
Assistant Professor
Department of Engineering Sciences
JSPM'S Rajarshi Shahu College of Engg.
Tathawade, Pune - 33

BALASAHEB B.GADEKAR
M.sc Mathematics
Assistant Professor
Department of Engineering Sciences
JSPM'S Rajarshi Shahu College of Engg.
Tathawade, Pune - 33

VIDYARTHI BHANDAR
Books & Stationers, Xerox
Opp. J.N.E. College N-6, Cidco,
A'bad. Mob. 9764422777 BATUE6A

Tech-Max Publications, Pune
Innovation Throughout Engineering Division

SYLLABUS

[BTCOC402] Probability and Statistics

UNIT 1 — 6 hrs

Probability Theory : Definition of probability : Classical, empirical and axiomatic approach of probability, Addition theorem of probability, Multiplication theorem of probability, Bayes' theorem of inverse probability, Properties of probabilities with proofs, Examples. **(Refer chapter 1)**

UNIT 2 — 6 hrs

Random Variable and Mathematical Expectation : Random variables, Probability distributions, Probability mass function, Probability density function, Mathematical expectation, Join and marginal probability distributions, Properties of expectation and variance with proofs. **(Refer chapter 2)**

UNIT 3 — 6 hrs

Theoretical Probability Distributions : Binomial distribution, Poisson distribution, Normal distribution, Fitting of binomial distributions. Properties of binomial, Poisson and normal distributions, Relation between binomial and normal distributions, Relation between Poisson and normal distributions, Importance of normal distribution, Examples. **(Refer chapter 3)**

UNIT 4 — 6 hrs

Correlation : Introduction, Types of Correlation, Correlation and Causation, Methods of studying correlation, Karl Pearson's Correlation Coefficient, Spearman's Rank Correlation Coefficient, Properties of Karl Pearson's correlation coefficient and Spearman's rank correlation coefficient, probable Errors. **(Refer chapter 4)**

UNIT 5 — 6 hrs

Linear Regression Analysis : Introduction, Linear and Non-linear Regression, Lines of Regression, Derivations of Regression Lines of Y on X and X on Y ; Angle between the Regression Lines, Coefficient of Regression, Theorems on Regression Coefficient, Properties of Regression Coefficient. **(Refer chapter 5)**

UNIT 6 — 6 hrs

Applied Statistics : Curve Fitting by the Method of Least Square - Fitting of Straight Lines, Second degree parabolas and More General Curves, Test of Significance : Large Sample Test for single Proportion, Difference of Proportion, Single Mean, Difference of Means and Difference of Standard Deviations. **(Refer chapter 6)**

□□□

Table of Contents

UNIT I

Chapter 1 : Probability Theory 1-1 to 1-33

1.1 Introduction 1-1

1.2 Theory of Probability 1-1

✓ **Syllabus Topic :** Definition of Probability : Classical 1-3

1.2.1 Mathematical (Or Classical) Definition of Probability 1-3

✓ **Syllabus Topic :** Definition of Probability : Empirical 1-3

1.2.2 Statistical (Or Empirical) Definition of Probability 1-3

✓ **Syllabus Topic :** Definition of Probability Axiomatic Approach of Probability 1-5

1.2.3 Axioms 1-5

✓ **Syllabus Topic :** Properties of Probabilities with Proofs 1-5

1.3 Properties of Probabilities with Proofs 1-5

1.3.1 Probability of the Impossible Event is Zero, i.e., $P(\phi) = 0$ 1-5

1.3.2 Probability of the Complementary Event $\bar{A}$ is given by, $P(\bar{A}) = 1 - P(A)$ 1-5

1.3.3 For any Two Events A and B, $P(\bar{A} \cap B) = P(B) - P(A \cap B)$ 1-6

1.3.4 If $B \subset A$, Then
(i) $P(A \cap \bar{B}) = P(A) - P(B)$
(ii) $P(B) \le P(A)$ 1-6

1.3.5 $P(A \cap B) \le P(A)$ and $P(A \cap B) \le P(B)$ 1-6

1.4 Addition Theorem of Probabilities (Theorem of Total Probability) 1-6

1.4.1 If A, B and C are any Three Events, Then $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$ 1-7

1.4.2 If $A_1, A_2, ..., A_n$ are n Mutually Exclusive Events, then the Probability of the Happening of One of them is $P(A_1 \cup A_2 \cup ..... \cup A_n) = P(A_1 + A_1 + ... + A_n) = P(A_1) + P(A_2) + ... + P(A_n)$ 1-7

1.5 Conditional Probability 1-10

✓ **Syllabus Topic :** Multiplication Theorem on Probability 1-10

1.5.1 Multiplication Theorem on Probability 1-10

1.5.2 Multiplication Theorem for Independent Events 1-11

✓ **Syllabus Topic :** Baye's Theorem of Inverse Probability 1-20

1.6 Revising Prior Estimates of Probabilities 1-20

1.6.1 Baye's Theorem (Statement): 1-20

1.7 Miscellaneous Examples 1-23

➔ **Multiple Choice Questions**

UNIT II

Chapter 2 : Random Variable and Mathematical Expectation 2-1 to 2-24

2.1 Introduction 2-1

✓ **Syllabus Topic :** Random Variables 2-1

2.2 Random Variables 2-1

2.2.1 Types of Random Variables 2-2

✓ **Syllabus Topic :** Probability Distributions 2-3

2.3 Probability Distribution of a Random Variable 2-3

✓ **Syllabus Topic :** Probability Mass Function 2-3

2.4 Probability Mass Function 2-3

2.5 Probability Density Function 2-4
2.5.1 Properties of a Distribution Function 2-4
2.5.2 Properties of Probability Density Function 2-4
✓ **Syllabus Topic :** Mathematical Expectations 2-8
2.6 Mathematical Expectation 2-8
✓ **Syllabus Topic :** Joint Probability Distributions 2-9
2.7 Joint Probability Distribution 2-9
2.7.1 Properties of Joint Probability Distribution Function : 2-10
2.7.2 Joint Probability Density Function of A Two-Dimensional Random Variables : 2-10
2.7.3 Properties of Joint Probability Density Function 2-10
✓ **Syllabus Topic :** Marginal Probability Distribution 2-10
2.8 Marginal Probability Distribution Function 2-10
✓ **Syllabus Topic :** Properties of Expectations and Variance with Proof 2-17
2.9 Properties of Expectation and Variance 2-17
➔ **Multiple Choice Questions**

UNIT III

Chapter 3 : Probability Distribution **3-1 to 3-38**

✓ **Syllabus Topic :** Binomial Distribution 3-1
3.1 Binomial Distribution 3-1
3.1.1 Examples on Binomial Distribution 3-2
✓ **Syllabus Topic :** Poisson Distribution 3-8
3.2 Poisson's Distribution 3-8
3.2.1 Mean of Poisson's Distribution 3-8
3.2.2 Variance of Poisson's Distribution 3-8
3.2.3 Standard Deviation of Poisson's Distribution 3-9
3.2.4 Examples on Poisson's Distribution 3-9
✓ **Syllabus Topic :** Normal Distribution 3-18
3.3 Normal Distribution 3-18
3.3.1 Standard Form of the Normal Distribution 3-19
3.3.2 Steps to find Probability with different z 3-19
3.3.1 Examples on Normal Distribution 3-20
✓ **Syllabus Topic :** Fitting of binomial Distribution 3-25
3.4 Fitting of Binomial Distribution 3-25
✓ **Syllabus Topic :** Properties of Binomial Distribution 3-27
3.5 Properties of Binomial Distribution 3-27
3.5.1 Applications of Binomial Distribution : 3-28
✓ **Syllabus Topic :** Properties of Poisson Distribution 3-28
3.6 Properties of Poisson Distribution 3-28
✓ **Syllabus Topic :** Properties of normal Distribution 3-29
3.7 Properties of normal Distribution : 3-29
✓ **Syllabus Topic :** Relation between Binomial and Normal Distribution 3-31
3.8 Relation between Binomial and Normal Distribution 3-31
✓ **Syllabus Topic :** Relation between Poisson and Normal Distribution 3-31
3.9 Relation between Poisson and Normal Distribution 3-31
✓ **Syllabus Topic :** Importance of Normal Distribution 3-31
3.10 Importance of Normal distribution 3-31
➔ **Multiple Choice Questions**

UNIT IV

Chapter 4 : Correlation **4-1 to 4-20**

4.1 Introduction 4-1
4.1.1 Measures of Central Tendency 4-1
4.1.2 Arithmetic Mean (or Simply) Mean 4-2

4.1.3 Effect of Change the Origin and Scale 4-2
4.1.4 Dispersion 4-2
4.1.5 Standard Deviation (S.D) 4-2
4.1.6 Coefficient of Variation 4-3
✓ **Syllabus Topic :** Correlation and Causation 4-3
4.2 Correlation and Causation : 4-3
✓ **Syllabus Topic :** Types of Correlation 4-4
4.3 Types of Correlation 4-4
4.3.1 Covariance 4-4
4.3.1.1 Properties of Covariance 4-5
✓ **Syllabus Topic :** Methods of Studying Correlation 4-5
4.4 Methods of Studying Correlation : 4-5
4.4.1 Scatter Diagram Method : 4-5
✓ **Syllabus Topic :** Karl Pearson's Correlation Co-efficient 4-6
4.5 Karl-Pearson's Correlation 4-6
✓ **Syllabus Topic :** Properties of Karl Pearson's Correlation Coefficient 4-7
4.6 Properties of Karl Pearson's Correlation Coefficient 4-7
4.6.1 Working Rule for Finding Correlation Coefficient 4-7
✓ **Syllabus Topic :** Spearman's Rank Correlation Coefficient 4-13
4.7 Spearman's Rank Correlation Co-efficient 4-13
✓ **Syllabus Topic :** Properties of Spearman's Rank Correlation Co-efficient 14
4.8 Properties of Spearman's Rank Correlation Co-efficient 4-14
4.8.1 Working Rule to Solve the Problems on Spearman's Rank Correlation Co-efficient 4-14
✓ **Syllabus Topic :** Probable Errors 4-17
4.9 Probable Error of Karl Person's Coefficient of Correlation : 4-17
➔ **Multiple Choice Questions**

UNIT V

Chapter 5 : Linear Regression Analysis **5-1 to 5-19**

✓ **Syllabus Topic :** Introduction 5-1
5.1 Introduction 5-1
✓ **Syllabus Topic :** Linear Regression and Non- Linear Regression 5-1
5.2 Linear Regression 5-1
✓ **Syllabus Topic :** Lines Of Regression 5-1
5.3 Lines of Regression 5-1
✓ **Syllabus Topic :** Derivation of Lines of Regression of Y on X 5-2
5.3.1 Derivation of Lines of Regression of Y on X 5-2
✓ **Syllabus Topic :** Derivation of Lines of Regression of X on Y 5-3
5.3.2 Derivation of Lines of Regression of X on Y 5-3
✓ **Syllabus Topic :** Angle between the Regression Lines 5-5
5.4 Angle Between The Regression Lines 5-5
✓ **Syllabus Topic :** Co-efficient of Regression 5-5
5.5 Co-efficient of Regression 5-5
✓ **Syllabus Topic :** Theorems on Co-efficient of Regression 5-5
5.6 Theorems on Co-efficient of Regression 5-5
✓ **Syllabus Topic :** Properties of Regression Co-efficient 5-6
5.7 Properties of Regression Co-efficient 5-6
5.7.1 Working Rule to find Lines of Regression 5-6
➔ **Multiple Choice Questions**

UNIT VI

Chapter 6 : Applied Statistics 6-1 to 6-28

6.1 Introduction 6-1

6.1.1 General Rules (Laws) Reducible to Linear form (Straight Line) 6-1

6.1.2 Fitting a Straight Line by Graphical Method 6-2

✓ **Syllabus Topic :** Curve Fitting by the Method of Least Square 6-3

6.2 Curve Fitting by the Method of Least Square 6-3

6.2.1 Principle of Least Squares 6-3

✓ **Syllabus Topic :** Fitting of Straight Line by the Method of Least Squares 6-4

6.3 Fitting of Straight Line by the Method of Least Squares 6-4

6.3.1 There are different Methods for Finding the Constants m and c 6-4

✓ **Syllabus Topic :** Fitting of Second degree parabolas by the Method of Least Squares 6-9

6.4 Fitting of Parabola by the Method of Least Square 6-9

✓ **Syllabus Topic :** Fitting More General Curves by the Method of Least Squares 6-13

6.5 Fitting of Other Curves (Or Fitting Exponential Curve) ... 6-13

6.6 Testing of Hypothesis 6-15

✓ **Syllabus Topic :** Tests of Significance 6-16

6.7 Tests of Significance 6-16

✓ **Syllabus Topic :** Large Sample Test for Single Proportion, Difference of Proportion, Single Mean, Difference of Means and Difference of Standard Deviations 6-16

6.7.1 Tests of Significance For Large Samples 6-16

➔ **Multiple Choice Questions**

CHAPTER 1

UNIT 1

Probability Theory

Syllabus

Probability Theory : Definition of probability : Classical, empirical and axiomatic approach of probability, Addition theorem of probability, Multiplication theorem of probability, Bayes' theorem of inverse probability, Properties of probabilities with proofs, Examples.

1.1 Introduction

- Ever heard about a weather forecast at the end of a news bulletin on TV or read about the weather conditions of your city/country for the next few days in any newspaper? They specifically use the term "probability."
- We are going to learn a few basic concepts, probability formulas involved to calculate the probability for different types of situations.
- Probability theory had its origin in the analysis of certain games of chance that were popular in the seventeenth century.It has since found applications in many branches of Science and Technology and this extensive application makes it an important branch of study.
- Probabilitytheory, as a matter of fact,is a study of random or unpredictable experiments and is helpful in investigating the important features of these random experiments.
- Probability is the measure of the likelihood of an event. The basic mathematics of probability theory started with games of chance, but it can be applied to many situations, from weather forecasting to politics.
- Probabilities range from 0 (no likelihood) to 1 (certainty), and are expressed as rational numbers.
- From the collective information, the study of chance or uncertainty i.e. the analyzing and interpretating such data can be possible using theory of probability.

1.2 Theory of Probability

Here we define and explain certain terms which are used frequently.

1. **Die :**
   It is a small cube. Dots like

   marked on its faces.Plural of the die is dice. On throwing a die,the outcome is the number of dots on its upper face.

2. **Card :** A pack of cards consists of four suits i.e Spades , Hearts , Diamonds and Clubs.Each suit consist of 13 cards.Nine cards numbered 2 , 3 ,4 ,5....10,an Ace, a King , a queen and a Jack.Colours of Spades and Clubs are black and that of Hearts and Diamonds is red.Aces , Kings , Queens and Jacks are known as face cards

(a) Experiment :

- An experiment is an activity that results in one and only one outcome out of a set of disjoint outcomes where an outcome cannot be predicted with certainty.
- Consider an experiment of tossing a die.There are six outcomes and we do not know which outcome will come upwhen we throw.

  (a) **Trial and event.** Let an experiment be repeated under essentially the same conditions and let it result in any one of the several possible outcomes. Then, the experiment is called a trial and the possible outcomes are known as events or cases.

For example :

(i) Tossing of a coin is a trial and the turning up of head or tail is an event.

(ii) Throwing a die is a trial and getting 1 or 2 or 3 or 4 or 5 or 6 is an event.

(b) Exhaustive events or Sample space. The total number of all possible outcomes in any trial is known as exhaustive events or exhaustive cases.or Sample space.It is denoted by S.

For Example :

If a coin is tossed, the possible outcomes are H (Head) and T (Tail).

Thus, S = (H, T)

When a Die is tossed then any one number of dots from 1 , 2 , ...6 appear on upper face.

Here S = (1 , 2 , 3 ,4 ,5 , 6)

- **Sample point.** The elements of S, the sample space, are called sample points.
- For example, if a coin is tossed and H and T denote 'Head' and 'Tail' respectively, then S = (H, T).
- The two sample points are H and T.
- **Finite sample space.** If the number of sample points in a sample space is finite, we call it a finite sample space. (In this chapter, we shall deal with finite sample spaces only)

(i) In throwing a die, there are 6 exhaustive cases, for any one of the six faces may turn up.

(ii) In throwing two dice, the exhaustive cases are $6 \times 6 = 6^2$ for any of the 6 numbers from 1 to 6 on one die can be associated with any of the 6 numbers on the other die.

In general, in throwing n dice, the exhaustive cases are 6^n.

(c) Favourable events or cases. The cases which entail the happening of an event are said to be favourable to the event. It is the total number of possible outcomes in which the specified event happens.

For example :

(i) When a die is thrown then

Sample space = { 1,2,3,4,5,6}

Consider the event A that the upper face of die shows a number which is multiple of 3.In this sample space the outcome favourable to given event are 3 and 6.

∴ Favourable number of cases = 2

Let B be event that we get an even number.

As the outcome in which event B occurs are 2,4,6.

∴ Favourable number of cases = 3

(ii) When two dice are thrown simultaneously then

Sample space = { (1,1,) (1,2),(1,3),(1,4),(1,5),(1,6)
(2,1,), (2,2),(2,3),(2,4),(2,5),(2,6)
(3,1,), (3,2),(3,3),(3,4),(3,5),(3,6)
(4,1,), (4,2),(4,3),(4,4),(4,5),(4,6)
(5,1,), (5,2),(5,3),(5,4),(5,5),(5,6)
(6,1,), (6,2),(6,3),(6,4),(6,5),(6,6)}

the number of cases favourable to get a sum 6 is 5, viz, (1, 5); (5, 1); (2, 4); (4, 2); (3, 3).

∴ Favourable number of cases = 5.

(d) Mutually exclusive events. Events are said to be mutually exclusive or incompatible if the happening of any one of them precludes (i.e. rules out) the happening of all others, i.e., if no two or more than two of them can happen simultaneously in the same trial.

For example :

(i) In tossing a coin, the events head and tail are mutually exclusive, since if the outcome is head, the possibility of getting tail in the same trial is ruled out.

(ii) In throwing a die, all the six faces numbered, 1, 2, 3, 4, 5, 6 are mutually exclusive since any outcome rules out the possibility of getting any other.

(e) Equally likely events. Events are said to be equally likely if there is no reason to expect any one in preference to any other.

For example :

(i) When a card is drawn from a well shuffled pack, any card may appear in the draw so that the 52 different cases are equally likely.

(ii) In throwing a die, all the six faces are equally likely to come.

(f) Independent and dependent events. Two or more events are said to be independent if the happening or non-happening of any one does not depend (or is not affected) by the happening or non-happening of other. Otherwise they are said to be dependent.

For example : If a card is drawn from a pack of well shuffled cards and replaced before drawing the second card, the result of the second draw is independent of the first draw. However, if the first card drawn is not replaced, then, the second draw is dependent on the first draw.

(g) Impossible event: An event which is not possible to happen in given experiment is called as impossible event.

For example :

When a die is tossed then getting number 7 is impossible event.

Syllabus Topic : Definition of Probability : Classical

1.2.1 Mathematical (Or Classical) Definition of Probability

- If a trial results in n exhaustive, mutually exclusive and equally likely cases and m of them are favourable to the happening of an event E, then the probability of happening of E is given by,

$$p \text{ or } P(E) = \frac{\text{Favourable number of cases}}{\text{Exhaustive number of cases}} = \frac{m}{n}$$

Note :

1. Since the number of cases favorable to happening of E is m and the exhaustive number of cases n, therefore number cases unfavorable to happening of E are n – m.
2. The probability that the event E will not happen is given by,

$$q \text{ or } P(\bar{E}) = \frac{\text{Unfavourable number of cases}}{\text{Exhaustive number of cases}} = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - p$$

$p + q = 1$ i.e. $P(E) + P(\bar{E}) = 1$

- Obviously, p and q are non-negative and cannot exceed unity, i.e., $0 \le p \le 1, 0 \le q \le 1$.

Note :

3. If $P(E) = 1$, E is called a certain event i.e., the chance of its happening is cent percent. If $P(E) = 0$, then E is an impossible event.
4. m cases are favourable to E and m cases are favourable to $\bar{E}$ (i.e., unfavourable to E), then exhaustive number of cases of cases = n + m.

$$P(E) = \frac{n}{n+m} \text{ and } P(\bar{E}) = \frac{m}{n+m}$$

We say that "odds in favour of E" are n : m and "odds against E" are m : n.

Syllabus Topic : Definition of Probability : Empirical

1.2.2 Statistical (Or Empirical) Definition of Probability

If in n trials, an event E happens m times, then the probability of happening of E is given by,

$$p = P(E) = \underset{n \to \infty}{\text{Lt}} \frac{m}{n}.$$

Illustrative Examples

Ex. 1.2.1 : Find the Probability of getting an even number on rolling a die once.

Soln. :

Step I: When a die is thrown either of six numbers viz., 1, 2, 3, 4, 5, 6 will appear on upper face of die

Step II : Sample Space (S) = {1, 2, 3, 4, 5, 6}. Here Event is that an even number appear on upper face

∴ (E) = {2, 4, 6}

Therefore, n (S) = 6 and n (E) = 3

Step III :

➔ **Using standard formula of probability**

$$\ldots \left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}} \right]$$

Here number of favourablecases = 3

Total number of cases = 6

$$\therefore P(A) = \frac{3}{6} = \frac{1}{2} = 0.5 \checkmark \qquad \ldots\text{Ans.}$$

Ex. 1.2.2 : Find the Probability of getting HEAD at least once on tossing two coins simultaneously.

Soln. :

Step I: When two coins tossed simultaneously the possible combinations of heads and tails will be as follows

(Head , Head),(Head, Tail),(Tail, Head),(Tail, Tail)

Step II : Sample Space (S) = {HH, HT, TH, TT}; where H denotes Head and T denotes Tail.

Step III : Here Event is that we get at least one head

∴ (E) = {HH, HT, TH}

Therefore, Therefore, n (S) = 4 and r. (E) = 3

Step IV :

➔ **Using standard formula of probability**

$$\ldots \left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}} \right]$$

Here number of favourable cases = 3

Total number of cases = 4

$$\therefore P(A) = \frac{3}{4} = 0.75 \checkmark \qquad \ldots\text{Ans.}$$

Ex. 1.2.3 : In a class of 10 students, 4 are boys and the rest are girls. Find the probability that a student selected will be girl.

Soln. :

Step I : In a class there are 10 students. Out of 10 students there are 4 boys and rest means 10 – 4 = 6 are girls.

Let A be event that a student selected is girl.

Step II :

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

There are six girls out 10 students

Here number of favourable cases =6

Total number of cases = 10

$\therefore\ P(A) = \frac{6}{10} = \frac{3}{5}$ ✓ ...Ans.

Ex. 1.2.4 : What is the chance that a (i) Non-leap year (ii) Leap year should have fifty three Sundays.?

Soln. :

Step I : (i) In a non leap year, there are 365 days. In a week containing all days have seven days. Therefore there are $\frac{365}{7} = 52.14$

i.e. there are 52 full weeks and some extra days left.

52 weeks contains $52 \times 7 = 364$ days. So one day will be left. That one extra day will be any one from Monday. Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday.

Here sample space = [Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday]

Step II : So in a leap year all seven days will be counted 52 times and any one extra day will be 53 times. Let A be event that a non leap year will contain 53 Sundays.

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 1
- Total number of cases = 7

$\therefore\ P(A) = \frac{1}{7}$

Step III : (ii) In a leap year, there are 366 days. As above there are 52 whole weeks and two extra days will be remained. Which will be any of following ways Monday-Tuesday, Tuesday–Wednesday,

Wednesday –Thursday, Thursday – Friday,

Friday– Saturday, Saturday – Sunday and Sunday– Monday.

Here sample space

= [Monday-Tuesday, Tuesday Wednesday, Wednesday –Thursday, Thursday – Friday, Friday– Saturday, Saturday – Sunday and Sunday– Monday.]

Step IV : Hence there are seven possible combinations. Out of these combinations, there are two combinations containing Sunday. In these two combinations Sunday will be repeated fifty three times. Let A be event that a leap year contains 53 Sundays.

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 2
- Total number of cases = 7

$\therefore\ P(A) = \frac{2}{7}$ ✓ ...Ans.

Ex. 1.2.5 : A bag contains 7 white, 6 red and 5 black balls. Two balls are drawn at random. Find the probability that they will both be white.

Soln. :

Step I : Total number of balls = 7 + 6 + 5 = 18

Out of 18 balls, 2 can be drawn in ${}^{18}C_2$ ways.

→ **Using standard formula :** $\ldots\left[{}^{n}C_r=\frac{n!}{r!\times(n-r)!}\right]$

Here n = 18 , r = 2

$${}^{18}C_2 = \frac{18!}{2!\times(18-2)!} = \frac{18!}{2!\times 16!}$$

$$\text{Now, } 18! = 1\times2\times3\times4\times\ldots\ldots15\times16\times17\times18$$

$$= 16!\times17\times18$$

$$= 18\times17\times16!$$

$$\therefore 18C_2 = \frac{18\times17\times16!}{2\times16!}$$

$$\therefore \text{Exhaustive number of cases} = {}^{18}C_2 = \frac{18\times17}{2\times1} = 153$$

Step II : Out of 7 white balls, 2 can be drawn in ${}^{7}C_2$ ways

→ **Using standard formula :** $\ldots\left[{}^{n}Cr=\frac{n!}{r!\times(n-r)!}\right]$

Here n = 7 , r = 2

$${}^{7}C_2 = \frac{7!}{2!\times(7-2)!} = \frac{7!}{2!\times5!}$$

$$\text{Now, } 7! = 1\times2\times3\times4\times5\times6\times7$$

$$= 5!\times6\times7$$

$$= 7\times6\times5!$$

$$\therefore {}^{7}C_2 = \frac{7\times6\times5!}{2\times5!} = \frac{7\times6}{2\times1} = 21 \text{ ways}$$

∴Favourable number of cases = 21

Step III :

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

Here number of favourable cases = 153

Total number of cases = 21

$P(A) = \frac{21}{153} = \frac{7}{51}$ ✓ ...Ans.

Ex. 1.2.6 : A letter is chosen at random from the word 'PROFESSOR' :

(i) What is the probability that it is a vowel?

(ii) What is the probability that it is a 'S'?

Soln. :

The word PROFESSOR contains in all 9 letters.

(i) The vowels are : a, e, i, o, u. three vowels O,E and O

- Here number of favourable cases = 3
- Total number of cases = 9

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

$P(A) = \frac{3}{9} = \frac{1}{3}$ ✓ ...Ans.

(ii) In given words the letter S appears 2 times

- Here number of favourable cases = 2
- Total number of cases = 9

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

$P(A) = \frac{2}{9}$ ✓ ...Ans.

Ex. 1.2.7 : Assume that a card selected is randomly from a deck of 52 playing cards. Find the probability that

(i) Card drawn is the king

(ii) Card drawn is a "Diamond".

Soln. : In a deck there are 52 playing cards.

- In a pack there are 4 king cards.
- Here number of favourable cases = 4
- Total number of cases = 52

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

$P(A) = \frac{4}{52} = \frac{1}{13}$ ✓ ...Ans.

(ii) In a pack there are 13 "Diamond" cards

- Here number of favourable cases = 13
- Total number of cases = 52

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

$P(A) = \frac{13}{52} = \frac{1}{4}$ ✓ ...Ans.

Syllabus Topic : Definition of Probability Axiomatic Approach of Probability

1.2.3 Axioms

(i) With each event E (i.e., a sample point) is associated a real number between 0 and 1, called the probability of that event and is denoted by P(E). Thus $0 \le P(E) \le 1$.

(ii) The sum of the probabilities of all simple (elementary) events constituting the sample space is 1. Thus P(S) = 1.

(iii) The probabilities of a compound event (i.e., an event made up of two or more sample events) is the sum of the probabilities of the simple events comprising the compound event.

Thus, if there are n equally likely possible outcomes of a random experiment, then the sample space S contains n sample points and the probability associated with each sample point is $\frac{1}{n}$.

[By Axiom (ii)]

Now, if an event E consists of m sample points, then the probability of E is,

$$P(E) = \frac{1}{n} + \frac{1}{n} + \ldots \text{ m times} = \frac{m}{n}$$

$$= \frac{\text{Number of sample points in E}}{\text{Number of sample points in S}}$$

This closely agrees with the classical definition of probability.

Syllabus Topic : Properties of Probabilities with Proofs

1.3 Properties of Probabilities with Proofs

1.3.1 Probability of the Impossible Event is Zero, i.e., $P(\phi) = 0$

Proof : Impossible event contains no sample point. As such, the sample space S and the impossible event ϕ are mutually exclusive.

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

$\Rightarrow P(S) + P(\phi) = P(S) \Rightarrow P(\phi) = P(S) - P(S) = 0$

Thus we have , $P(\phi) = 0$

1.3.2 Probability of the Complementary Event

$\bar{A}$ is given by, $P(\bar{A}) = 1 - P(A)$

Proof : $\bar{A}$ and A are disjoint events. Also $A \cup \bar{A} = S$

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

$$\Rightarrow P(A) + P(\bar{A}) = 1$$

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

1.3.3 For any Two Events A and B,

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

Proof : $\bar{A}\cap B$ is set of elements which belongs to set B but belongs to set A $[x : x\in B \text{ and } x\notin A]$

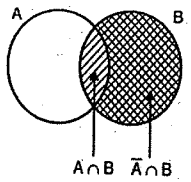


Fig. 1.3.1

From above Venn diagram we observe that $\bar{A}\cap B$ and $A\cap B$ are disjoint sets and also $\left(\bar{A}\cap B\right)\cup(A\cap B) = B$

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

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

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

Note : Similarly, it can be proved that,

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

1.3.4 If $B\subset A$, Then

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

(ii) $P(B)\le P(A)$

Proof

From Fig 1.3.2 we observe that when $B\subset A$, then B and $A\cap\bar{B}$ are disjoint and their union is A.

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

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

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

$$\Rightarrow P\left(A\cap\bar{B}\right) = P(A) - P(B) \qquad ...(1)$$

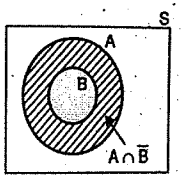


Fig. 1.3.2

Now, if E is any event,

then $0\le P(E)\le 1$, i.e., $P(E)\ge 0$

$$\therefore P\left(A\cap\bar{B}\right)\ge 0 \Rightarrow P(A) - P(B)\ge 0 \qquad \text{[Using (1)]}$$

$$\Rightarrow P(B)\le P(A)$$

1.3.5 $P(A\cap B)\le P(A)$ and $P(A\cap B)\le P(B)$

Proof

By (Section 1.3.2) $B\subset A \Rightarrow P(B)\le P(A)$

Since $(A\cap B)\subset A$ and $(A\cap B)\subset B$

$\therefore P(A\cap B)\le P(A)$ and $P(A\cap B)\le P(B)$

1.4 Addition Theorem of Probabilities (Theorem of Total Probability)

Statement

If A and B are any two events, then

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

i.e., $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Proof

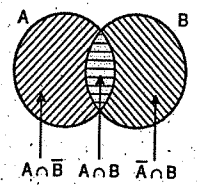


Fig. 1.4.1

– A and $\bar{A}\cap B$ are disjoint sets and their union is $A\cup B$.

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

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

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

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

$$\left[\because \bar{A}\cap B \text{ and } A\cap B \text{ are disjoint}\right]$$

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

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

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

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

Note 1 : If A and B are two mutually disjoint events, then $A\cap B = \phi$, so that $P(A\cap B) = P(\phi) = 0$.

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

Note 2 : $P(A\cup B)$ is also written as $P(A + B)$. Thus, for mutually disjoint events A and B,

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

$P(A\cap B)$ is also written as $P(AB)$

1.4.1 If A, B and C are any Three Events, Then $P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(B\cap C) - P(C\cap A) + P(A\cap B\cap C)$

Proof

Using the above result for two events, we have

$$P(A\cup B\cup C) = P[(A\cup B)\cup C]$$

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

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

(By Distributive law)

$$P(A) + P(B) + P(C) - P(A\cap B) - [P(A\cap C)] + P(B\cap C) - P[(A\cap C)\cap(B\cap C)] \text{ (By Art. 21.45)}$$

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

$$[\because (A\cap C)\cap(B\cap C) = A\cap B\cap C]$$

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

$$[\because A\cap C = C\cap A]$$

or $P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(CA) + P(ABC)$

1.4.2 $A_1, A_2, ..., A_n$ are n Mutually Exclusive Events, then the Probability of the Happening of One of them is

$$P(A_1\cup A_2\cup .....\cup A_n) = P(A_1 + A_1 + ... + A_n)$$
$$= P(A_1) + P(A_2) + ... + P(A_n)$$

Proof

– Let N be the total number of mutually exclusive, exhaustive and equally likely cases of which m_1 are favourable to A_1, m_2 are favourable to A_2 and so on.

– Probability of occurrence of event $A_1 = P(A_1) = \frac{m_1}{N}$

Probability of occurrence of event $A_2 = P(A_2) = \frac{m_2}{N}$

....(1)

Probability of occurrence of event $A_n = P(A_n) = \frac{m_n}{N}$

– The events being mutually exclusive and equally likely, the number of causes favourable to the event.

A_1 or A_2 or ... or A_n is $m_1 + m_2 + ...m_n$

$\therefore$ Probability of occurrence of one of the events $A_1, A_2, ..., A_n$ is $P(A_1 + A_2 + ... + A_n)$

$$= \frac{m_1 + m_2 + ... + m_n}{N} = \frac{m_1}{N} + \frac{m_2}{N} + ... + \frac{m_n}{N}$$

$$= P(A_1) + P(A_2) + ... + P(A_n) \text{ (Using (1))}$$

Ex. 1.4.1 : A card is drawn from a well-shuffled pack of playing cards. What is the probability that it is either a spade or an ace ?

Soln. :

Step I : Let A = the event of drawing a spade and B = the event of drawing an ace

A and B are not mutually exclusive

$(A\cap B)$ = the event of drawing the ace of spades

Step II: $P(A) = \frac{13}{52}, P(B) = \frac{4}{52}, P(A\cap B) = \frac{1}{52}$

Step III: Using standard formula of addition law of Probability

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

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13} \checkmark \qquad ...\text{Ans.}$$

Ex. 1.4.2 : In a triangular series , the probability of the Indian team winning match against Pakistan is 0.7 and that against Australia is 0.4.If probability of India winning both matches is 0.3.what is the probability that India win at least one match so that it can enter final?

Soln.:

Step I : Let A be event that India win match against Pakistan

$$\therefore P(A) = 0.7$$

Let B be event that India win match against Australia

$$\therefore P(A) = 0.4$$

probability of India win match against both Pakistan and Australia

$$\therefore P(A\cap B) = 0.3$$

Step II : Now the probability that India win at least one match is given by $P(A\cup B)$

As given events are not independent,

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

$$= 0.7 + 0.4 - 0.3 = 0.8 \checkmark \qquad ...\text{Ans.}$$

Ex. 1.4.3 : If a fair die is tossed once, find the probability of getting an odd or a number less than 4

Soln.:

Step I : Here the sample space S = (1,2,3,4,5,6)

Let A be event that an odd number appears.

There are three odd numbers 1,3,5

➔ **Using standard formula of probability**

$$...\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 3
  Total number of cases = 6

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

- Let B be event that a number less than 4 appear.
- There are three numbers 1,2,3 which are less than 4

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

- The number which are less than 4 and odd are 1 and 3
- $A \cap B$ is an event which gives number less than 4 and odd

$$\therefore P(A \cap B) = \frac{2}{6} = \frac{1}{3}$$

Step II : Now the probability getting an odd or a number less than 4 is given by $P(A \cup B)$

As given events are not independent,

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

$$= \frac{1}{2} + \frac{1}{2} - \frac{1}{3} = \frac{2}{3} \checkmark \quad \text{...Ans.}$$

Ex. 1.4.4 : 50 tickets are serially numbered 1 to 50.One ticket is drawn from these at random. Find the probability of its being multiple of 3 or 4

Soln.:

Step I : Here the sample space S = (1,2,3,4,5,6 .....50)

Let A be event that ticket number is multiple of 3

Sample points of A are

(3,6,9,12,15,18,21,24,27,30,33,36,39,42,45,48) = 16

➔ **Using standard formula of probability**

$$\left[\because P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases =16
- Total number of cases = 50

$$\therefore P(A) = \frac{16}{50}$$

- Let B be event that a ticket number is multiple of 4.

Sample points of B are (4,8,12,16,20,24,28,32,36,40,44,48) = 12

$$\therefore P(B) = \frac{12}{50}$$

- There are three numbers which are multiple of both 3 and 4
- Sample points of $A \cap B$ are (12,24,36,48) = 4

$$\therefore P(A \cap B) = \frac{4}{50}$$

Step II : Now the probability getting ticket number which is multiple of 3 or 4 is given by $P(A \cup B)$

As given events are not independent,

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

$$= \frac{16}{50} + \frac{12}{50} - \frac{4}{50} = \frac{24}{50} = \frac{12}{25} \checkmark \quad \text{...Ans.}$$

Ex. 1.4.5 : A student takes his examination in four subjects A, B, C, D. The chances that he qualify in the subjects are $\frac{4}{5}, \frac{3}{4}, \frac{5}{6}, \frac{2}{3}$ respectively. To qualify the examination he must pass in subject A and at least two other subjects. What is the probability that he qualifies ?

Soln. :

Step I : By given information

$$P(A) = \frac{4}{5}$$

- The probability that student qualify in subject A is P(A)

$$\therefore P(A) = \frac{4}{5}$$

- The probability that student qualify in subject B is P(B)

$$\therefore P(B) = \frac{3}{4}$$

- The probability that student does not quality in subject B is $P(\bar{B})$

$$P(\bar{B}) = 1 - P(B) = 1 - \frac{3}{4} = \frac{1}{4}$$

- The probability that student qualify in subject C is P(C)

$$P(C) = \frac{5}{6}$$

- The probability that student does not quality in subject C is $P(\bar{C})$

$$\therefore P(\bar{C}) = 1 - P(C) = 1 - \frac{5}{6} = \frac{1}{6}$$

- The probability that student qualifies in subject D is P(D)

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

- The probability that student does not qualify in subject D is $P(\bar{D})$

$$\therefore P(\bar{D}) = 1 - P(D) = 1 - \frac{2}{3} = \frac{1}{3}$$

Step II: The qualifying student need to pass in subject A and in any at least two other subjects which is given by,

(Qualifying all four subjects) or (Qualifying in subject A,B, C and not qualifying D) or

(Qualifying in subject A,B, D and not qualifying C)or (Qualifying in subject A,C, D and not qualifying B)

Step III: Using standard Addition formula of probability

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

i.e., P(A or B) = P(A) + P(B) – P(A and B)

Here all events are independent $P(A \cap B) = 0$

Required Probability

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

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

$$= \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdot \frac{2}{3} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdot \frac{1}{3} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{6} \cdot \frac{2}{3} + \frac{4}{5} \cdot \frac{1}{4} \cdot \frac{5}{6} \cdot \frac{2}{3}$$

$$= \frac{\cancel{4}}{\cancel{5}} \cdot \frac{\cancel{3}}{\cancel{4}} \cdot \frac{\cancel{5}}{\cancel{6}} \cdot \frac{2}{\cancel{3}} + \frac{\cancel{4}}{\cancel{5}} \cdot \frac{\cancel{3}}{\cancel{4}} \cdot \frac{1}{\cancel{3}} + \frac{\cancel{4}}{\cancel{5}} \cdot \frac{\cancel{3}}{\cancel{4}} \cdot \frac{1}{6} \cdot \frac{2}{\cancel{3}} + \frac{\cancel{4}}{\cancel{5}} \cdot \frac{1}{\cancel{4}} \cdot \frac{\cancel{5}}{6} \cdot \frac{2}{3}$$

$$= \frac{2}{6} + \frac{1}{6} + \frac{2}{30} + \frac{2}{18} = \frac{60 + 30 + 12 + 20}{180}$$

$$= \frac{122}{180} = \frac{61}{90} \checkmark \quad \text{...Ans.}$$

Ex. 1.4.6 : Four persons are chosen at random from a group containing 3 men,2 women and 4 children. Find the probability that exactly two of them will be children.

Soln. :

Step I : In a group there are 3 + 2 + 4 = 9 persons. The number of ways of choosing four persons out of 9 are

$$^9C_4 = \frac{91}{4! \times 5!} = \frac{9 \times 8 \times 7 \times 6 \times 5!}{4 \times 3 \times 2 \times 5!} = 126$$

Step II: Now in a required group containing four persons two of them must be children's . For this any one of

Following combinations is possible

(1) 2 children + one men + 1 women

(2) 2 children + 2 men + 0 women

(3) 2 children + 0 men + 2 women

Step III : Way I : The number of ways of having 2 children out of 4, 1 men out of 3 and 1 women out of 2 are $^4C_2 \times {}^3C_1 \times {}^2C_1$

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

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

$$= 6 \times 3 \times 2 = 36$$

➔ **Using standard formula of probability**

$$\left[\because P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases =36
- Total number of cases = 126

$$\therefore P(\text{Way I}) = \frac{36}{126}$$

Step III: Way II : The number of ways of having 2 children out of 4, 2 men out of 3 and 0 women out of 2 are $^4C_2 \times {}^3C_2 \times {}^2C_0$

$$= \frac{4!}{2! \times 2!} \times \frac{3!}{2!\,1!} \times \frac{2!}{0!\,2!}$$

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

$$= 6 \times 3 \times 1 = 18$$

➔ **Using standard formula of probability**

$$\left[\because P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 18
- Total number of cases = 126

$$\therefore P(\text{Way II}) = \frac{18}{126}$$

Step IV: Way III: The number of ways of getting 2 children out of 4, 0 men out of 3 and 2 women out of 2 are $^4C_2 \times {}^3C_0 \times {}^2C_2$

$$= \frac{4!}{2!\,2!} \times \frac{3!}{0!\,3!} \times \frac{2!}{2!\,0!}$$

$$= \frac{4 \times 3 \times 2!}{2! \times 2!} \times 1 \times 1 = 6 \times 1 \times 1 = 6$$

➔ **Using standard formula of probability**

$$\left[\because P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 6
- Total number of cases = 126

$$\therefore P(\text{Way III}) = \frac{6}{126}$$

Step V: The probability that a selection will contain 2 children will happen in any one of above way where all the three ways are independent.

∴ The probability that a selection will contain exactly 2 children's = P(Way I ∪ Way II ∪ Way III)

$$= P(\text{Way I}) + P(\text{Way II}) + P(\text{Way III})$$

$$= \frac{36}{126} + \frac{18}{126} + \frac{6}{126} = \frac{36 + 18 + 6}{126}$$

$$= \frac{60}{126} = \frac{10}{21} \checkmark \quad \text{...Ans.}$$

Ex. 1.4.7 : Three groups of children contain respectively 3 girls and 1 boy, 2 girls and 2 boys, 1 girl and 3 boys. One child is selected at random from each group. Find the chance of selecting 1 girl and 2 boys.

Soln. : From the given groups, the number of required

children's can be selected in following ways.

Step I :

Way I : 1 girl from first group, one boy from second and third group. There are 3 + 1 = 4 children in first group in which there are 3 girls. The probability of choice of one girl from first group $= \frac{3}{4}$

- In second group there are 2 boys and 2 girls. The probability of choice of one boy from second group is $= \frac{2}{4}$
- In third group there are 1 girl and 3 boys. Out of these four children's the probability of choice of one boy $= \frac{3}{4}$

∴ In this case the probability of choice of 1 girl and 2 boys

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

$$= \frac{3}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{9}{32}$$

Step II:

Way II : The selection may be a boy from first group, a girl from second group and a boy from third group.

In group, there are 3 girls and 1 boy, probability of selection of 1 boy

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

- In second group there are 2 girls and 2 boys. Probability of selection of 1 girl from second group $= \frac{2}{4}$
- In group, there are 1 girl and 3 boys. The probability of selection of one boy from third group $= \frac{3}{4}$. In this case the probability of choice of 1 girl and 2 boys

$$= \frac{1}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{1}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{3}{32}$$

Step III :

Way III : The selection may be a boy from first two groups and a girl from third group.

- In first group, there are 3 girls and 1 boy.
- The probability of selecting 1 boy from first group $= \frac{1}{4}$.
- In second group there are 2 girls and 2 boys. The probability of selecting 1 boy from second group $= \frac{2}{4}$
- In third group these are 1 girl and 3 boys. The probability of selecting girl from third group $= \frac{1}{4}$.
- Thus probability of selecting 1 girl and 2 boys in this way is given by

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

$$= \frac{1}{4} \times \frac{1}{4} \times \frac{1}{2} = \frac{1}{32}$$

Step IV : The probability of selecting 1 girl and 2 boy is possible in any one of above three cases where all the three ways are independent.

∴ The probability that a selection will contain exactly 1 girl and 2 boys = P(Way I ∪ Way II ∪ Way III)

$$= P(\text{Way I}) + P(\text{Way II}) + P(\text{Way III})$$

$$= \frac{9}{32} + \frac{3}{32} + \frac{1}{32} = \frac{13}{32} \checkmark \quad \text{...Ans.}$$

1.5 Conditional Probability

- The probability of the happening of an event E_1 when another event E_2 is known to have already happened is called conditional probability and is denoted by $P(E_1/E_2)$.
- **Mutually independent Events.** An event E_1 is said to be independent of an event E_2 if,

$$P(E_1/E_2) = P(E_1)$$

i.e., if the probability of happening of E_1 is independent of the happening of E_2.

Syllabus Topic : Multiplication Theorem on Probability

1.5.1 Multiplication Theorem on Probability

- Suppose you and your friends are playing with two dice. You note down the various possible outcomes of the sample space.
- When one of your friends throws the dice, he did not tell the result. Instead, he said that when you add the two numbers you will get an even number.
- Can you now guess the result of his throw? It is interesting to know how this additional information affects the probabilities of the possible outcomes. This is the basic concept of multiplication theorem on probability. Let us get more familiar with it.
- In conditional probability, we know that the probability of occurrence of some event is affected when some of the possible events have already occurred.
- When we know that a particular event B has occurred, then insfead of S, we concentrate on B for calculating the probability of occurrence of event A given B.
- Taking the above example of throwing of two dice, the possible outcomes are

$$S = \{(x, y): x, y = 1, 2, 3, 4, 5, 6\}.$$

- There are 36 elements in the sample space S. The probability of occurrence of any of the possible outcome is $P(Ei) = \frac{1}{36}$.
- We don't know the result of the throw of the dice by the friend. However, we have the information that the sum of the numbers is even. Let us find how this information is affecting the probability of the outcome.

- Event A shows the sum of the numbers is an even number.
- A = {(1, 1), (1, 3), (1, 5), (2, 2), (2, 4), (2, 6), (3, 1), (3, 3), (3, 5), (4, 2), (4, 4), (4, 6), (5, 1), (5, 3), (5, 5), (6, 2), (6, 4), (6, 6)}. This means that out of 36 outcomes these 18 outcomes are now only possible and the remaining are not.
- The probability for each of these outcomes is $P(A \mid Ei) = \frac{18}{36} = \frac{1}{2}$.
- This example shows that some additional information may change the probability of the happening of some event.

Theorem 1

For two events A and B,

$$P(A \cap B) = P(A)\, P(B \mid A), P(A) > 0.$$

or, $$P(A \cap B) = P(B)\, (A \mid B), P(B) > 0.$$

Proof : From the concept of conditional probability, we have $P(A \mid B) = P(A \cap B)/P(B)$.

Re-writing the above, we have, $P(A \cap B) = P(B)\, P(A \mid B)$.

Similarly,

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

$$\Rightarrow P(A \cap B) = P(A)\, P(B \mid A).$$

- he mathematical theorem on probability shows that the probability of the simultaneous occurrence of two events A and B is equal to the product of the probability of one of these events and the conditional probability of the other, given that the first one has occurred.

Theorem 2

For two events A and B such that P(B) > 0, $P(A \mid B) \le P(A)$.

Proof:

- Let S be sample space
- It is obvious that the number of common outcomes in A and B is either less or equal to the number of outcomes in any of the event.

$$n(A \cap B) \le n(A) \quad \text{... (i),}$$

$$\text{and, } n(B) \le n(S) \quad \text{... (ii)}$$

Dividing (i) and (ii), we get,

$$n(A \cap B)/n(B) \le n(A)/n(S)$$

$$\Rightarrow P(A \mid B) \le P(A).$$

1.5.2 Multiplication Theorem for Independent Events

- The multiplication theorem on probability for dependent events can be extended for the independent events.
- From the theorem, we have, $P(A \cap B) = P(A)\, P(B \mid A)$. If the events A and B are independent, then, $P(B \mid A) = P(B)$. The above theorem reduces to

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

- This shows that the probability that both of these occur simultaneously is the product of their respective probabilities.

(A) Extension of Multiplication Theorem of Probability to n Events

For n events A1, A2, ... , An, we have

$$P(A1 \cap A2 \cap \ldots \cap An) = P(A1)\, P(A2 \mid A1)\, P(A3 \mid A1 \cap A2) \ldots \times P(An \mid A1 \cap A2 \cap \ldots \cap An\text{-}1)$$

(B) Extension of Multiplication Theorem of Probability to n Independent Events

- For n independent events, the multiplication theorem reduces to

$$P(A1 \cap A2 \cap \ldots \cap An) = P(A1)\, P(A2) \ldots P(An).$$

Solved Examples

Ex 1.5.1 : A box contains 5 black, 7 red and 6 green balls. Three balls are drawn from this box one after the other without replacement. What is the probability that the three balls are

(i) all black balls

(ii) of different colors

(iii) two black and one green black.

Soln. :

Step I : The total number of the balls in the box is 5 + 7 + 6 = 18. Let events

B : drawing black balls.

R : drawing red balls.

G : drawing green balls.

- The balls are drawn without replacement one by one. For the first draw, there are 18 balls to choose from.
- The number of balls gets lessened by 1 for the second draw i.e., 18 – 1 = 17 and 16 for the third draw.

Step II : Probability that the three balls are all black

There are 5 black balls out of 18

In the first draw, the probability of getting black ball,

$$P(B1) = \frac{5}{18}$$

- In the second draw , there will 4 black balls out of 17 balls
- In the second draw, the probability of getting black ball,

$$P(B2 \mid B1) = \frac{4}{17}$$

- In the third draw , there will 3 black balls out of 16 balls
- In the third draw, the probability of getting black ball ,

$$P(B3 \mid B1 \cap B2) = \frac{3}{16}$$

Step III : The probability that the three balls are all black in colour= $P(B1) \times P(B2 \mid B1) \times P(B3 \mid B1 \cap B2)$

$$= \frac{5}{18} \times \frac{4}{17} \times \frac{3}{16} = \frac{5}{408}$$

Step IV : The probability that the three balls are all different in colour

- The three balls drawn of different colours is obtained by getting one black , one red and one green ball
- There are 5 black balls out of 18
- In the first draw, the probability of getting one black ball ,

$$P(B1) = \frac{5}{18}$$

- After drawing one black ball , there will remain 17 balls out of which 7 are red.
- In the second draw, the probability of getting red ball ,

$$P(R1 \mid B1) = \frac{7}{17}$$

- After drawing one black and one red ball in first two draws , there will remain 16 balls out of which 6 are green.
- In the third draw, the probability of getting green ball ,

$$P(G1 \mid B1 \cap R1) = \frac{6}{16}$$

Step V: The probability that the three balls are all different in colour = $P(B1) \times P(R1 \mid B1) \times P(G1 \mid B1 \cap R1)$

$$= \frac{5}{18} \times \frac{7}{17} \times \frac{6}{16} = \frac{35}{816}$$

Step VI: Probability that two black and one green balls

Now out of three balls drawn , two are black and one is green.

- There are 5 black balls out of 18
- In the first draw, the probability of getting one black ball ,

$$P(B1) = \frac{5}{18}$$

- After drawing one black ball , there will remain 17 balls out of which 4 are black
- In the second draw, the probability of getting black ball

$$P(B2 \mid B1) = \frac{4}{17}$$

- After drawing two black balls in first two draws , there will remain 16 balls out of which 6 are green.
- In the third draw, the probability of getting green ball ,

$$P(G1 \mid B1 \cap B2) = \frac{6}{16}$$

Step VII : The probability that the three balls are all different in colour = $P(B1) \times P(R1 \mid B1) \times P(G1 \mid B1 \cap R1)$

$$= \frac{5}{18} \times \frac{4}{17} \times \frac{6}{16} = \frac{5}{204} \checkmark \quad \text{...Ans.}$$

Ex. 1.5.2 : In a given race, the odds in favour of four horses A, B, C, D are 1 : 3, 1 : 4, 1 : 5, 1 : 6 respectively. Assuming that a dead heat is impossible; find the chance that one of them wins the race.

Soln. :

Step I: Let A be event that horse A wins the race

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

- Let B be event that horse B wins the race

$$\therefore P(B) = \frac{1}{5}$$

- Let C be event that horse C wins the race

$$\therefore P(C) = \frac{1}{6}$$

- Let D be event that horse D wins the race

$$\therefore P(D) = \frac{1}{7}$$

Step II: Since a dead heat (in which all the four horses cover same distance in same time) is not possible, the events are mutually exclusive, by addition law of probability that any one of the four horses wins the race is $P(A \cup B \cup C \cup D)$ which is given by

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

$$= \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} = \frac{319}{420} \checkmark \quad \text{...Ans.}$$

Ex. 1.5.3 : Six cards are drawn at random from a pack of 52 cards. What is the probability that 3 will be red and 3 black?

Soln. :

Step I : 6 cards from a pack of 52 cards can be drawn in $52C_6$ ways

➔ **Using standard formula** $\left[{}^nC_r = \frac{n!}{r! \times (n-r)!} \right]$

Here $n = 52, r = 6$

$${}^{52}C_2 = \frac{52!}{6! \times (52-6)!} = \frac{52!}{6! \times 46!}$$

Now,

$$52! = \underline{1 \times 2 \times 3 \times 4 \times \ldots\ldots 45 \times 46} \times 47 \times 48 \times 49 \times 50 \times 51 \times 52$$

$$= 46! \times 47 \times 48 \times 49 \times 50 \times 51 \times 52$$

$$= 52 \times 51 \times 50 \times 49 \times 48 \times 47 \times 46!$$

$$\therefore {}^{52}C_2 = \frac{52 \times 51 \times 50 \times 49 \times 48 \times 47 \times 46!}{6 \times 5 \times 4 \times 3 \times 2 \times 46!}$$

$$= 52 \times 17 \times 10 \times 47 \times 49$$

Step II: In a pack of 52 cards there are 13 diamond and 13 heart, red cards. That is there are 26 red cards. Out of 26 red 3 can be drawn in ${}^{26}C_3$ ways.

➔ **Using standard formula** $\left[{}^nC_r = \frac{n!}{r! \times (n-r)!} \right]$

Here $n = 26, r = 3$

$${}^{26}C_3 = \frac{26!}{3! \times (26-3)!} = \frac{26!}{3! \times 23!}$$

Now,

$$26! = \underline{1 \times 2 \times 3 \times 4 \times \ldots\ldots 22 \times 23} \times 24 \times 25 \times 26$$

$$= 23! \times 24 \times 25 \times 26$$

$$= 26 \times 25 \times 24 \times 23!$$

$$\therefore {}^{26}C_3 = \frac{26 \times 25 \times 24 \times 23!}{3 \times 2 \times 23!} = 26 \times 25 \times 4$$

Step III : Similarly there are 13 spade and 13 club black cards i.e. there are 26 black cards out of 26 black 3 can be drawn in ${}^{26}C_3$ ways.

$$26C_3 = \frac{26!}{3! \times 23!} = \frac{26 \times 25 \times 24 \times 23!}{3 \times 3 \times 23!} = 26 \times 25 \times 4$$

Step IV : The number of ways of getting three red and three black cards are ${}^{26}C_3 \times {}^{26}C_3$

➔ **Using standard formula of probability**

$$\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}} \right]$$

- Here number of favourable cases

$$= 26 \times 25 \times 4)(26 \times 25 \times 4)$$

- Total number of cases = $52 \times 17 \times 10 \times 47 \times 49$

Step V : The probability that out of six cards 3 will be red and 3 will be black is

$$= \frac{(26 \times 25 \times 4)(26 \times 25 \times 4)}{52 \times 17 \times 10 \times 47 \times 49}$$

$$= \frac{26 \times 25 \times 4 \times 26 \times 25 \times 4}{52 \times 17 \times 10 \times 47 \times 49}$$

$$= \frac{5 \times 4 \times 26 \times 25}{17 \times 49 \times 47}$$

$$= \frac{13000}{39151} \checkmark \quad \text{...Ans.}$$

Ex. 1.5.4 : Four cards are drawn from a pack of cards. Find the probability that (i) all are diamonds, (ii) there is one card of each suit, and (iii) there are two spades and two hearts.

Soln. :

Step I: To find the probability that all four cards are diamond4 cards can be drawn from a pack of 52 cards in ${}^{52}C_4$ ways.

➔ **Using standard formula** $\left[{}^nC_r = \frac{n!}{r! \times (n-r)!} \right]$

Here $n = 52, r = 4$

$${}^{52}C_4 = \frac{52!}{4! \times (52-4)!} = \frac{52!}{4! \times 48!}$$

Now,

$$52! = \underline{1 \times 2 \times 3 \times 4 \times \ldots\ldots 47 \times 48} \times 49 \times 50 \times 51 \times 52$$

$$= 48! \times 49 \times 50 \times 51 \times 52$$

$$= 52 \times 51 \times 50 \times 49 \times 48!$$

$$\therefore {}^{52}C_4 = \frac{52 \times 51 \times 50 \times 49 \times 48!}{4! \times 48!}$$

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

Step II : (i) There are 13 diamonds in the pack and 4 can be drawn out of them in ${}^{13}C_4$ ways.

➔ **Using standard formula** $\left[{}^nC_r = \frac{n!}{r! \times (n-r)!} \right]$

Here $n = 13, r = 4$

$${}^{13}C_4 = \frac{13!}{4! \times (13-4)!} = \frac{13!}{4! \times 9!}$$

Now,

$$13! = \underline{1 \times 2 \times 3 \times 4 \times \ldots\ldots 8 \times 9} \times 10 \times 11 \times 12 \times 13$$

$$= 9! \times 10 \times 11 \times 12 \times 13$$

$$= 13 \times 12 \times 11 \times 10 \times 9!$$

$$\therefore {}^{13}C_4 = \frac{13 \times 12 \times 11 \times 10 \times 9!}{4! \times 9!}$$

$$= \frac{13 \times 12 \times 11 \times 10}{4 \times 3 \times 2 \times 1} = 715$$

Step III :

➔ **Using standard formula of probability**

$$\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}} \right]$$

- Here number of favourable cases = 715
- Total number of cases = 270725

$$P(A) = \frac{715}{270725} = \frac{143}{54145} = \frac{11}{4165}$$

Step IV : To find the probability that four cards drawn, one card is from each suit

(ii) There are 4 suits, each containing 13 cards.

1 card from given suit can be drawn in ${}^{13}C_1$ ways

Thus one card from each of four suits can be drawn in ${}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1$ ways

Step V : $\therefore$ Favourable number of cases

$$= {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1$$

➔ **Using standard formula** ...[${}^nC_1 = n$]

$${}^{13}C_1 = 13$$

$\therefore$ Favourable number of cases

$$= 13 \times 13 \times 13 \times 13$$

Step VI :

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = $13\times13\times13\times13$

– Total number of cases = 270725

$$P(A) = \frac{13\times13\times13\times13}{270725}=\frac{2197}{20825}$$

Step VII: To find the probability that four cards drawn, two are spade and two are hearts

(iii) 2 spades out of 13 can be drawn in $^{13}C_2$ ways.

2 hearts out of 13 can be drawn in $^{13}C_2$ ways

$\therefore$ Favourable number of cases = $^{13}C_2\times{}^{13}C_2$

→ **Using standard formula** $\ldots\left[^{n}C_r=\frac{n!}{r!\times(n-r)!}\right]$

Here n = 13 , r = 2

$$^{13}C_2=\frac{13!}{2!\times(13-2)!}=\frac{13!}{2!\times11!}$$

Now,

$13! = \underline{1\times2\times3\times4\times\ldots\ldots8\times9\times10\times11}\times12\times13$

$= 11!\times12\times13$

$= 13\times12\times11!$

$$\therefore {}^{13}C_2 = \frac{13\times12\times11!}{2!\times11!}$$

$$=\frac{13\times12}{2\times1}=78$$

$\therefore {}^{13}C_2 = 78$

$\therefore$ Favourable number of cases = 78×78

Step VIII :

→ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 78×78

– Here number of cases = 270725

$$P(A) = \frac{78\times78}{270725}=\frac{468}{20825}\checkmark \qquad \ldots\text{Ans.}$$

Ex. 1.5.5 : A problem in mechanics is given to three students A, B, C whose chances of solving it are $\frac{1}{2},\frac{1}{3},\frac{1}{4}$ respectively. What is the probability that the problem will be solved ?

Soln. :

Step I : Let A be event that problem is solved by A

$$\therefore P(A) = \frac{1}{2}$$

– Let B be event that problem is solved by B

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

– Let C be event that problem is solved by C

$$\therefore P(C) = \frac{1}{4}$$

Step II :

– By definition of complementary event, The probability that A will not be solved by A ,

$$P(\bar{A}) = 1-P(A) = 1-\frac{1}{2}=\frac{1}{2}$$

– The probability that B will not be solved by B ,

$$P(\bar{B}) = 1-P(B) = 1-\frac{1}{3}=\frac{2}{3}$$

– The probability that C will not be solved by C ,

$$P(\bar{C}) = 1-P(C) = 1-\frac{1}{4}=\frac{3}{4}$$

Step III : $\therefore$ The probability that the problem is not solved by any of them

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

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

– Hence the probability that the problem is solved by at least one of them

$$= 1-P(\bar{A}\cap\bar{B}\cap\bar{C}) = 1-\frac{1}{4}=\frac{3}{4}\checkmark \qquad \ldots\text{Ans.}$$

Ex. 1.5.6 : The odds that a book will be favourably reviewed by three independent critics are 5 to 2, 4 to 3 and 3 to 4 respectively. What is the probability that, of the three reviews, a majority will be favourable ?

Soln. :

Step I : Let the three critics be A, B, C.

– The probability that the book being favourably reviewed by A, $P(A)=\frac{5}{7}$

– The probability that the book is not favourably reviewed by A, $P(\bar{A}) = 1-P(A) = 1-\frac{5}{7}=\frac{2}{7}$

– The probability that the book being favourably reviewed by B, $P(B)=\frac{4}{7}$

– The probability that the book is not favourably reviewed by B, $P(\bar{B}) = 1-P(B) = 1-\frac{4}{7}=\frac{3}{7}$

– The probability that the book being favourably reviewed by C, $P(C)=\frac{3}{7}$

– The probability that the book is not favourably reviewed by C, $P(\bar{C}) = 1-P(C) = 1-\frac{3}{7}=\frac{4}{7}$

Step II : A majority will be favourable if the reviews of at least two are favourable.

(i) All A, B, C all review favourably, the probability is,

$$P(A\cap B\cap C)=P(A)\,P(B)\,P(C)=\frac{5}{7}\times\frac{4}{7}\times\frac{3}{7}=\frac{60}{343}$$

(ii) If A, B review favourably and C reviews unfavourably, the probability is,

$$P(A\cap B\cap\bar{C}) = P(A)\,P(B)\,P(\bar{C}) = \frac{5}{7}\times\frac{4}{7}\times\frac{4}{7}=\frac{80}{343}$$

(iii) If A, C review favourably and B reviews unfavourably, probability is

$$P(A\cap\bar{B}\cap C) = P(A)\,P(\bar{B})\,P(C) = \frac{5}{7}\times\frac{3}{7}\times\frac{3}{7}=\frac{45}{343}$$

(iv) If B, C review favourbly and Areviewsunfavourably, the probability is,

$$P(\bar{A}\cap B\cap C) = P(\bar{A})\,P(B)\,P(C) = \frac{2}{7}\times\frac{4}{7}\times\frac{3}{7}=\frac{24}{343}$$

Step III : Since these are mutually exclusive events, the probability that a majority will be favourable is

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

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

$$=\frac{60}{343}+\frac{80}{343}+\frac{45}{343}+\frac{24}{343}=\frac{209}{343}\checkmark \qquad \ldots\text{Ans.}$$

Ex. 1.5.7 : A can hit a target 4 times in 5 shots; B 3 times in 4 shots; C twice in 3 shots. They fire a volley. What is the probability that at least two shots hit ?

Soln. :

Step I : Probability of A's hitting the target $P(A)=\frac{4}{5}$

Probability of A's not hitting the target $P(\bar{A}) = 1-P(A)$

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

Probability of B's hitting the target $P(B)=\frac{3}{4}$

– Probability of B 's not hitting the target $P(\bar{B}) = 1-P(B)$

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

Probability of C's hitting the target $P(C)=\frac{2}{3}$

– Probability of C 's not hitting the target $P(\bar{C}) = 1-P(C)$

$$=\left(1-\frac{2}{3}\right)=\frac{1}{3}$$

All above events are independent

Step II : To hit target by at least two, the following are possibilities

(i) All A, B, C hit the target, the probability for which is

$$P(A\cap B\cap C) = P(A)\,P(B)\,P(C) = \frac{4}{5}\times\frac{3}{4}\times\frac{2}{3}=\frac{24}{60}$$

(ii) A, B hit the target and C misses it, the probability for which is,

$$P(A\cap B\cap\bar{C}) = P(A)\,P(B)\,P(\bar{C}) = \frac{4}{5}\times\frac{3}{4}\times\frac{1}{3}=\frac{12}{60}$$

(iii) A, C hit the target and B misses it, the probability for which is,

$$P(A\cap\bar{B}\cap C) = P(A)\,P(\bar{B})\,P(C) = \frac{4}{5}\times\frac{1}{4}\times\frac{2}{3}=\frac{8}{60}$$

(iv) B, C hit the target and A misses it, the probability for which is,

$$P(\bar{A}\cap B\cap C) = P(\bar{A})\,P(B)\,P(C) = \frac{1}{5}\times\frac{3}{4}\times\frac{2}{3}=\frac{6}{60}$$

Step III : Since these are mutually exclusive events, required

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

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

$$=\frac{24}{60}+\frac{12}{60}+\frac{8}{60}+\frac{6}{60}=\frac{50}{60}=\frac{5}{6}\checkmark \qquad \ldots\text{Ans.}$$

Ex. 1.5.8 : A can hit the target 1 out of 4 times. B can hit the target 2 out of 3 times. C can hit the target 3 out of 4 times. Find the probability that at least two hit the target.

Soln. :

Step I : Probability of A's hitting the target $P(A) = \frac{1}{4}$

Probability of A's not hitting the target $P(\bar{A}) = 1-P(A)$

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

– Probability of B's hitting the target $P(B)=\frac{2}{3}$

– Probability of B 's not hitting the target $P(\bar{B}) = 1-P(B)$

$$=1-\frac{2}{3}=\frac{1}{3}$$

– Probability of C's hitting the target $P(C)=\frac{3}{4}$

– Probability of C 's not hitting the target $P(\bar{C}) = 1-P(C)$

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

– All above events are independent

Step II : To hit target by at least two the following are possibilities

(i) All A, B, C hit the target, the probability for which is

$$P(A \cap B \cap C) = P(A)\,P(B)\,P(C) = \frac{1}{4}\times\frac{2}{3}\times\frac{3}{4} = \frac{1}{8}$$

(ii) A, B hit the target and C misses it, the probability for which is,

$$P\left(A \cap B \cap \bar{C}\right) = P(A)\,P(B)\,P\left(\bar{C}\right) = \frac{1}{4}\times\frac{2}{3}\times\frac{1}{4} = \frac{1}{24}$$

(iii) A, C hit the target and B misses it, the probability for which is,

$$P\left(A \cap \bar{B} \cap C\right) = P(A)\,P\left(\bar{B}\right)\,P(C) = \frac{1}{4}\times\frac{1}{3}\times\frac{3}{4} = \frac{1}{16}$$

(iv) B, C hit the target and A misses it, the probability for which is,

$$P\left(\bar{A} \cap B \cap C\right) = P\left(\bar{A}\right)\,P(B)\,P(C) = \frac{3}{4}\times\frac{2}{3}\times\frac{3}{4} = \frac{3}{8}$$

Step III : Since these are mutually exclusive events, required

$$\text{probability} = P(A \cap B \cap C) + P\left(A \cap B \cap \bar{C}\right) + P\left(A \cap \bar{B} \cap C\right) + P\left(\bar{A} \cap B \cap C\right)$$

$$= \frac{1}{8} + \frac{1}{24} + \frac{1}{16} + \frac{3}{8} = \frac{29}{48} = 0.6041 \checkmark \quad \text{...Ans.}$$

Ex. 1.5.9 : A has 2 shares in a lottery in which there are 3 prizes and 5 blanks; B has 3 shares in a lottery in which there are 4 prizes and 6 blanks. Show that A's chance of success is to B's as 27 : 35.

Soln. :

Step I : A can draw two tickets (out of 3 + 5 = 8) in ${}^{8}C_3$ ways

➔ **Using standard formula** $\left[{}^{n}C_r = \frac{n!}{r! \times (n-r)!}\right]$

Here n = 8 , r = 3

$${}^{8}C_3 = \frac{8!}{3! \times (8-3)!} = \frac{8!}{3! \times 5!}$$

Now,

$$8! = 1\times2\times3\times4\times5\times6\times7\times8$$
$$= 5!\times6\times7\times8$$
$$= 8\times7\times6\times5!$$
$$\therefore {}^{8}C_3 = \frac{8\times7\times6\times5!}{3!\times5!}$$
$$= \frac{8\times7\times6}{3\times2\times1} = 56$$

Step II : A will get the blanks in ${}^{5}C_2$ ways.

➔ **Using standard formula** $\left[{}^{n}C_r = \frac{n!}{r! \times (n-r)!}\right]$

Here n = 5 , r = 2

$${}^{5}C_2 = \frac{5!}{2!\times(5-2)!} = \frac{5!}{2!\times3!}$$

Now,

$$5! = 1\times2\times3\times4\times5$$
$$= 3!\times4\times5$$
$$= 5\times4\times3!$$
$$\therefore {}^{5}C_2 = \frac{5\times4\times3!}{2!\times3!}$$
$$= \frac{5\times4}{2\times1} = 10$$

∴ A can win a prize in 56 – 10 = 46 ways

Step III :

➔ **Using standard formula of probability**

$$\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 56

– Here number of cases = 46

$$P(A) = \frac{46}{56} = \frac{23}{28}$$

– Hence A's chance of success $= \frac{23}{28}$

Step IV : B can draw 3 tickets in ${}^{10}C_3$ ways

➔ **Using standard formula** $\left[{}^{n}C_r = \frac{n!}{r! \times (n-r)!}\right]$

Here n = 10 , r = 3

$${}^{10}C_3 = \frac{10!}{3!\times(10-3)!} = \frac{10!}{3!\times7!}$$

Now,

$$10! = 1\times2\times3\times4\times5\times6\times7\times8\times9\times10$$
$$= 7!\times7\times8\times9\times10$$
$$= 10\times9\times8\times7!$$
$$\therefore {}^{10}C_3 = \frac{10\times9\times8\times7!}{3!\times7!}$$
$$= \frac{10\times9\times8}{3\times2\times1} = 120$$

– B will get all blanks in ${}^{6}C_3$ ways.

Step V:

➔ **Using standard formula** $\left[{}^{n}C_r = \frac{n!}{r! \times (n-r)!}\right]$

Here n = 6 , r = 3

$${}^{6}C_3 = \frac{6!}{3!\times(6-3)!} = \frac{6!}{3!\times3!}$$

Now,

$$6! = 1\times2\times3\times4\times5\times6$$
$$= 3!\times4\times5\times6$$
$$= 6\times5\times4\times3!$$
$$\therefore {}^{6}C_3 = \frac{6\times5\times4\times3!}{3!\times3!}$$
$$= \frac{6\times5\times4}{3\times2\times1} = 20$$

∴ B can win a prize in 120 – 20 – 100 ways.

Step III:

➔ **Using standard formula of probability**

$$\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 100

– Here number of cases = 120

$$P(B) = \frac{100}{120} = \frac{5}{6}$$

– Hence B's chance of success $= \frac{5}{6}$

Step VI:

∴ A's chance : B's chance $= \frac{23}{28} : \frac{5}{6} = 69 : 70 \checkmark$ ...Ans.

Ex. 1.5.10 : A and B throw alternately with a single die, A having the first throw. The person who first throws ace is to win. What are their respective chances of winning ?

Soln. :

Step I: The chance of throwing an ace with a single die $= \frac{1}{6}$.

By definition of complement event, the chance of not throwing an ace with a singled i.e. $= 1 - \frac{1}{6} = \frac{5}{6}$.

Step II: If A is to win, he should throw an ace in the first or third or fifth, ..., throws.

Step III: If B is to win, he should throw an ace in the second or fourth or sixth,....throw.

The chances that an ace is thrown in the first, second, third, ..., throws are

$$\frac{1}{6}, \frac{5}{6}, \frac{1}{6}, \frac{5}{6}, \frac{5}{6}, \frac{1}{6}, \frac{5}{6}, \frac{5}{6}, \frac{5}{6}, \frac{1}{6} \ldots\ldots$$

$$\text{or } \frac{1}{6}, \frac{5}{6}, \frac{1}{6}, \left(\frac{5}{6}\right)^2, \frac{1}{6}, \left(\frac{5}{6}\right)^3, \frac{1}{6}, \ldots\ldots$$

Step V: ∴ A's chance $= \frac{1}{6} + \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^4 \cdot \frac{1}{6} + \ldots\ldots$

– Using standard formula of sum of infinitely many terms of geometric progression

– If $a + ar + ar^2 + ar^3 + ar^4$ .....is G.P with first term a and common ratio r then sum of infinitely many terms is given by $S_n = \frac{a}{1-r}$ where

$$\text{Here } a = \frac{1}{6},\ r = \left(\frac{5}{6}\right)^2$$

$$S_n = \frac{\frac{1}{6}}{1 - \left(\frac{5}{6}\right)^2} = \frac{6}{11}$$

∴ A's chance of winning $\frac{6}{11}$

– Hence B's chance of winning $= 1 - \frac{6}{11} = \frac{5}{11} \checkmark$ ...Ar

Ex. 1.5.11 : Cards are drawn one by one from a we shuffled pack until an ace appears. Show th the probability that exactly n cards are dea before the first ace appears $\frac{4(51-n)(50-n)(49-n)}{52.51.50.49}$

Soln. :

Let A be the event of drawing n non-ace cards and the event of drawing an ace in the $(n + 1)^{th}$ draw.

Consider the event A

– n cards can be drawn out of 52 cards in ${}^{52}C_n$ ways.

⇒ Exhaustive cases = ${}^{52}C_n$

– n non-ace cards can be drawn out of 52 cards ${}^{48}C_n$ ways

⇒ Favourable cases = ${}^{48}C_n$

$$\therefore P(A) = \frac{{}^{48}C_n}{{}^{52}C_n} = \frac{48!}{(48-n)!\,n!} \times \frac{(52-n)!\,(n)!}{52!}$$

$$= \frac{48!\,(52-n)(51-n)(50-n)(49-n)(48-n)!}{(48-n)!.\,52.\,51.\,50.\,49.\,(48)!}$$

$$= \frac{(52-n)(51-n)(50-n)(49-n)}{52.\,51.\,50.\,49}$$

Consider the event B

– n cards have already been drawn in the first n draws.

Exhaustive cases = ${}^{52-n}C_1 = 52 - n$;

Favourable cases = ${}^{4}C_1 = 4$

$$\therefore P(B/A) = \frac{4}{52-n}$$

– Required Probability = P(A) · P(B/A)

$$= \frac{(52-n)(51-n)(50-n)(49-n)}{52.\,51.\,50.\,49} \times \frac{4}{52-n}$$

$$= \frac{4(51-n)(50-n)(49-n)}{52.\,51.\,50.\,49} \checkmark \quad \text{...A}$$

Ex. 1.5.12 : An urn contains 10 white and 3 black ba while another urn contains 3 white and 5 bl balls. Two balls are drawn from the first urn a put into the second urn and then a ball is dra from the latter. What is the probability that i a white ball ?

Soln. :

Step I : The first urn contains 10 white , 3 back = 13 ball The two balls drawn from the first urn may be

(i) both white

(ii) both black

(iii) one white and one black

Step II: Let A be event that balls drawn are both white

An urn contains 13 balls, out of which two balls can be drawn in $^{13}C_3$ ways.

→ **Using standard formula** $\left[\ldots {}^nC_r = \frac{n!}{r! \times (n-r)!}\right]$

Here n = 13 , r = 2

$$^{13}C_2 = \frac{13!}{2! \times (13-2)!} = \frac{13!}{2! \times 11!}$$

Now,

$$13! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \times 11 \times 12 \times 13$$
$$= 11! \times 2 \times 13$$
$$= 13 \times 12 \times 11!$$

$$\therefore {}^{13}C_2 = \frac{13 \times 12 \times 11!}{2! \times 11!}$$
$$= \frac{13 \times 12}{2 \times 1} = 78$$

- Therefore 2 balls can be drawn in 78 ways.
- Now 2 white balls from 10 white balls can be drawn in $^{10}C_2$ ways.

→ **Using standard formula**

$$\left[\ldots {}^nC_r = \frac{n!}{r! \times (n-r)!}\right]$$

Here n = 10 , r = 2

$$^{10}C_2 = \frac{10!}{2! \times (10-2)!} = \frac{10!}{2! \times 8!}$$

Now,

$$10! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10$$
$$= 8! \times 9 \times 10$$
$$= 10 \times 9 \times 8!$$

$$\therefore {}^{10}C_2 = \frac{10 \times 9 \times 8!}{2! \times 8!}$$
$$= \frac{10 \times 9}{2 \times 1} = 45$$

- Therefore 2 white balls can be drawn in 45 ways.
- The probability of drawing white ball from first urn

→ **Using standard formula**

$$\left[\ldots P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 45
- Here number of cases = 78

$$P(A) = \frac{^{10}C_2}{^{13}C_2} = \frac{45}{78} = \frac{15}{26}$$

Step III: Let B be event that balls drawn are both black

- An urn contains 13 balls, out of which two balls can be drawn in $^{13}C_3$ ways.

$$\therefore {}^{13}C_2 = 78$$

- Therefore 2 balls can be drawn in 78 ways.
- Now 2 black balls from 3black balls can be drawn in $3C_2$ ways

→ **Using standard formula** $\left[\ldots {}^nC_r = \frac{n!}{r! \times (n-r)!}\right]$

Here n = 3 , r = 2

$$^3C_2 = \frac{3!}{2! \times (3-2)!} = \frac{3!}{2! \times 1!} = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} = 3$$

- Therefore 2 black balls can be drawn in 3 ways.
- The probability of drawing black ball from first urn

→ **Using standard formula**

$$\left[\ldots P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 3.
- Here total number of cases = 78

$$P(B) = \frac{^3C_2}{^{13}C_2} = \frac{3}{78} = \frac{1}{26}$$

Step IV : Let C be event that balls drawn are one white and one black

- An urn contains 13 balls, out of which two balls can be drawn in $^{13}C_3$ ways.

→ **Using standard formula** $\left[\ldots {}^nC_r = \frac{n!}{r! \times (n-r)!}\right]$

Here n = 13 , r = 2

$$^{13}C_2 = 78$$

- Therefore 2 balls can be drawn in 78 ways.
- Now 1 white ball from 10 white balls can be drawn in $^{10}C_1$ ways

→ **Using standard formula** $\ldots[{}^nC_1 = n]$

Here n = 10

$$^{10}C_1 = 10$$

- Now 1 black ball from 3 black balls can be drawn in $^3C_1$ ways

→ **Using standard formula** $\ldots[{}^nC_1 = n]$

Here n = 3

$$^3C_1 = 3$$

- Hence one white and one black ball can be drawn in $^{10}C_1 \times {}^3C_1$ ways = 10 × 3 = 30 ways
- The probability of drawing one white and one black ball from first urn

→ **Using standard formula**

$$\left[\ldots P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 30
- Here number of cases = 78

$$P(C) = \frac{^{10}C_1 \times {}^3C_1}{^{13}C_2} = \frac{30}{78} = \frac{10}{26}$$

Step V: When two balls are transferred from first urn as per above order (i), (ii) and (iii) to second urn, the second urn will contain

(i) 5 white and 5 black balls

(ii) 3 white and 7 black balls

(iii) 4 white and 6 black balls.

Let W denote the event of drawing a white ball from the second urn in the three cases (i), (ii) and (iii).

Now, $P(W/A) = \frac{5}{10}, P(W/B) = \frac{3}{10}, P(W/C) = \frac{4}{10}$

$\therefore$ Required probability = $P(A) \cdot P(W/A) + P(B) \cdot P(W/B) + P(C) \cdot P(W/C)$

$$= \frac{15}{26} \cdot \frac{5}{10} + \frac{1}{26} \cdot \frac{3}{10} + \frac{10}{26} \cdot \frac{4}{10}$$
$$= \frac{75+3+40}{260} = \frac{118}{260} = \frac{59}{130} \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.5.13 : A card is drawn from an ordinary pack and a gambler bets that it is a spade or an ace. What is the probability that gambler lost the bet?

Soln. :

Step I : In a pack of cards, there are 52 cards. Out of these 52 cards, one card can be drawn in $52C_1$ ways.

→ **Using standard formula** $\ldots[{}^nC_1 = n]$

Here n = 52

$$^{52}C_1 = 52$$

- Let A be event that card drawn is spade.
- In a pack of 52, there are 13 spade cards.

→ **Using standard formula of probability**

$$\left[\ldots P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 13
- Total number of cases = 13

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

Step II: Let B be event that a card drawn is an ace.

In a pack of 52, there are 4 cards.

→ **Using standard formula of probability**

$$\left[\ldots P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 4
- Total number of cases = 52

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

Step III : The probability that a card drawn is spade or an ace is given by $P(A \cup B)$

- Since event A and B are dependent events i.e. a card drawn which is spade may be an ace card or a card drawn which is an ace may be spade.
- In a pack there are 52 cards there is one card which is spade and ace
- The probability of a card is spade and ace is $P(A \cap B)$

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

- Therefore by addition theorem of probability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

Step IV: If a card drawn is spade or ace then gambler will win the bet.

- The probability that gambler will loose the bet is given by

$$P(A \cup B) = 1 - P(A \cup B)$$
$$= 1 - \frac{4}{13}$$
$$= \frac{13-4}{13} = \frac{9}{13} \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.5.14 : A student takes his examination in four subjects A, B, C, D. The chances that he qualify in the subjects are $\frac{4}{5}, \frac{3}{4}, \frac{5}{6}, \frac{2}{3}$ respectively. To qualify the examination he must pass in subject A and at least two other subjects. What is the probability that he qualifies ?

Soln. :

Step I :

- The probability that student qualify in subject A is P(A)

$$\therefore P(A) = \frac{4}{5}$$

- The probability that student qualify in subject B is P(B)

$$\therefore P(B) = \frac{3}{4}$$

- The probability that student does not quality in subject B is $P(\bar{B})$

$$\therefore P(\bar{B}) = 1 - P(B) = 1 - \frac{3}{4} = \frac{1}{4}$$

– The probability that student qualify in subject C is P(C)

$\therefore \quad P(C) = \frac{5}{6}$

– The probability that student does not quality in subject

$\therefore$ C is $P(\bar{C})$

$\therefore \quad P(\bar{C}) = 1 - P(C) = 1 - \frac{5}{6} = \frac{1}{6}$

– The probability that student qualifies in subject D is P(D)

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

– The probability that student does not qualify in subject D is $P(\bar{D})$

$\therefore P(\bar{D}) = 1 - P(D) = 1 - \frac{2}{3} = \frac{1}{3}$

Step II: The quality student need to pass in subject A and any two other subjects which is given by

(i) Q1 : Qualify in subjects A and all of B , C . D

$P(Q1) = P(A \cap B \cap C \cap D)$

(ii) Q2 :Qualify in subjects A and B , C and not qualify in D

$P(Q2) = P(A \cap B \cap C \cap \bar{D})$

(iii) Q3 :Qualify in subjects A and B , D and not qualify in C

$P(Q3) = P(A \cap B \cap \bar{C} \cap D)$

(iv) Q4 :Qualify in subjects A and C , D and not qualify in B

$P(Q4) = P(A \cap \bar{B} \cap C \cap D)$

Step III: The student can qualify by any one of above way, probability of which is given by

$P(Q1 \cup Q2 \cup Q3 \cup Q4)$

Since all above possibilities are independent

$$P(Q1 \cup Q2 \cup Q3 \cup Q4) = P(Q1) + P(Q2) + P(Q3) + P(Q4)$$

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

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

$$= \frac{4}{5}\frac{3}{4}\frac{5}{6}\frac{2}{3} + \frac{4}{5}\frac{3}{4}\frac{5}{6}\frac{1}{3} + \frac{4}{5}\frac{3}{4}\frac{1}{6}\frac{2}{3} + \frac{4}{5}\frac{1}{4}\frac{5}{6}\frac{2}{3}$$

$$= \frac{\cancel{4}}{\cancel{5}}\frac{\cancel{3}}{\cancel{4}}\frac{\cancel{5}}{\cancel{6}}\frac{2}{\cancel{3}} + \frac{\cancel{4}}{\cancel{5}}\frac{\cancel{3}}{\cancel{4}}\frac{\cancel{5}}{\cancel{6}}\frac{1}{\cancel{3}}$$

$$+ \frac{\cancel{4}}{5}\frac{\cancel{3}}{\cancel{4}}\frac{1}{6}\frac{2}{\cancel{3}} + \frac{\cancel{4}}{\cancel{5}}\frac{1}{\cancel{4}}\frac{\cancel{5}}{6}\frac{2}{3} = \frac{2}{6} + \frac{1}{6} + \frac{2}{30} + \frac{2}{18}$$

$$= \frac{60 + 30 + 12 + 20}{180} = \frac{122}{180} = \frac{61}{90} \checkmark \qquad \text{...Ans.}$$

Syllabus Topic : Baye's Theorem of Inverse Probability

1.6 Revising Prior Estimates of Probabilities

☞ **Baye's Theorem**

– Sometimes the probability of certain event is calculated by using available information. However, subsequently, some additional information becomes available. This additional information will affect in change of prior estimate of probability.

– The new probabilities are known as revised or posterior probabilities.

– The origin of concept of obtaining posterior probabilities with limited information is attributed to Reverend Thomas Bayes and basic formula for conditional probability under dependence

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

1.6.1 Baye's Theorem (Statement):

– If $A_1, A_2, A_3, \ldots, A_n$ are mutually exclusive and exhaustive events and B is any event that occurs with $A_1, A_2, A_3, \ldots A_n$ then

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2) + \ldots + P(A_n) \cdot P(B \mid A_n)}$$

☞ **Proof**

– Let S – be the sample space.

– $A_1, A_2, A_3 \ldots A_n$ are subsets of S and mutually exclusive and exhaustive events.

– Let B – be any elementary mutually exclusive and exhaustive event occurs with $A_1, A_2, A_3, \ldots A_n$

– Since, by multiplication law of probability.

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{P(B)} \qquad \ldots(1)$$

where, $P(B) = P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2)$.

– But the event B occurs with $A_1, A_2, A_3 \ldots A_n$, then by addition law of probability, we have

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

Note : The probability of $A_1, A_2, A_3 \ldots A_n$ i.e. $p(A_i)$ $i = 1, 2, 3, \ldots n$ are called prior probability while the probabilities $p(B \mid A_i)$ are called posterior probability.

Ex. 1.6.1 : It is assumed that a car, on a particular root departs its place on time with probability 0.87. The car arrives at its destination on time (irrespective of its departure time) with probability 0.91. Also, the probability that a car arrives on time, given that its departs on time is 0.93. Find the probability that a car departed on time, given that it arrives on time.

Soln. :

Step I : Let, A-be the event, car arrives on time

B-be the event, car depart on time

Given : $P(A) = 0.91, P(B) = 0.87$

$P(A \mid B) = 0.93$

Step II:

$$\therefore \quad P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \mid B) \cdot P(B)}{P(A)}$$

$$= \frac{(0.93) \cdot (0.87)}{0.91} = 0.88 \checkmark \qquad \text{...Ans.}$$

Ex.1.6.2 : A manufacturing firm receives shipment of machine parts from two suppliers A and B. Currently 65 per cent of parts are purchased from supplier A and the remaining from supplier B .The past record shows that 2 per cent of the parts supplied by A are defective, whereas 5 per cent of the parts supplied by B are found defective. On a particular day a defective the machine breaks down because a defective part is fitted to it.

Given the information that the part was bad, using Bayes' theorem find the probability it was supplied by supplier B.

Soln. :

Step I : Let A be event that a part is supplied by Supplier A and B be event that a part is supplied by Supplier B.

$P(A) = 0.65$

$P(B) = 0.35$

– $P(D \mid A)$ be the probability that a part supplied by A is defective

$P(D \mid A) = 0.02$

– $P(D \mid B)$ be the probability that a part supplied by B is defective

$P(D \mid B) = 0.05$

Step II: $P(A)\, P(D \mid A) = 0.65 \times 0.02 = 0.0130$

$P(B)\, P(D \mid B) = 0.35 \times 0.05 = 0.0175$

Step III: The probability part supplied by supplier B is defective

$$= \frac{P(B) \cdot P(D \mid B)}{P(A) \cdot P(D \mid A) + P(B) \cdot P(D \mid B)}$$

$$= \frac{0.35 \times 0.05}{0.65 \times 0.02 + 0.35 \times 0.05}$$

$$= \frac{0.0175}{0.0130 + 0.0175} = \frac{0.0175}{0.0305} = 0.57 \checkmark \qquad \text{...Ans.}$$

Ex.1.6.3 : If a machine is set up correctly it produce 90 per cent good items; if it is incorrectly set up then it produces 10 per cent of good items. Chances for a setting to be correct and incorrect are in the ratio 7:3.After a setting is made the first two items produced are found to be good items. What is the chance that the setting was correct ?

Soln. :

Step I : Let A be event that a machine is set up correctly and B be event that a machine is not set up correctly.

$P(A) = \frac{7}{10} = 0.7$

$P(B) = \frac{3}{10} = 0.3$

– Let E be event that the first two items are good

– $P(E \mid A)$ be the probability that a item is good and produced by machine A

$P(E \mid A) = 0.9$

– $P(E \mid B)$ be the probability that a item is good and produced by machine B

$P(E \mid A) = 0.1$

Step II: $P(A)\, P(E \mid A) = 0.7 \times 0.9 = 0.63$

$P(B)\, P(E \mid B) = 0.3 \times 0.1 = 0.03$

Step III: The probability the machine is set up correctly

$$= \frac{P(A) \cdot P(E \mid A)}{P(A) \cdot P(E \mid A) + P(B) \cdot P(E \mid B)}$$

$$= \frac{0.7 \times 0.9}{0.7 \times 0.9 + 0.3 \times 0.1}$$

$$= \frac{0.63}{0.63 + 0.03} = \frac{0.63}{0.66} = 0.95 \checkmark \qquad \text{...Ans.}$$

The probability that machine is set up correctly is 0.95

Ex. 1.6.4: Suppose we have two machines I and II , which are used in the manufacture of shoes. Let E_1 be the event of shoes produced by machine I and E_2 be the event of shoes produced by machine II. Machine I produces 60 per cent of the shoes and machine II 40 per cent. It is also reported that 10 per cent of the shoes produced by machine I are defective as against the 20 per cent by machine II. What is the probability that a non-defective shoe was manufactured by machine I ?

Soln. :

Step I: Let E_1 be event that the shoes are produced by machine I and E_2 be event that the shoes are produced by machine I

$P(E_1) = 0.6$

$P(E_2) = 0.4$

– $P(N \mid E_1)$ be the probability that the shoes produced by machine I are non-defective

- As 10 per cent of the shoes produced by machine I are defective
- The percentage of non-defective shoes produced by machine I are $100 - 10 = 90$

$$P(N \mid E_1) = 0.9$$

- $P(N \mid E_2)$ be the probability that the shoes produced by machine II are defective
- As 20 per cent of the shoes produced by machine II are defective
- The percentage of non-defective shoes produced by machine I are $100 - 20 = 80$

$$P(N \mid E_2) = 0.8$$

Step II : $P(E_1)P(N \mid E_1) = 0.6 \times 0.9 = 0.54$

$P(E_2)\,P(N \mid E_2) = 0.4 \times 0.8 = 0.32$

Step III: The probability machine II will produce non-defective shoes

$$= \frac{P(E_2)\cdot P(N \mid E_2)}{P(E_1)\,P(N \mid E_1) + P(E_2)\cdot P(N \mid E_2)}$$

$$= \frac{0.4 \times 0.8}{(0.6 \times 0.9) + (0.4 \times 0.8)}$$

$$= \frac{0.32}{0.54 + 0.32} = \frac{0.32}{0.86} = 0.37 \checkmark \qquad \text{...Ans.}$$

The probability machine II will produce non-defective shoes is 0.37.

Ex.1.6.5 : A manufacturing firm is engaged in the production of steel pipes in its three plants with a daily production of 1000 , 1500 and 2500 units respectively. According to the past experience , it is known that the fraction of defective pipes produced by the three plants are respectively 0.04, 0.09 and 0.07.If a pipe is selected from a day's total production and found to be defective, find out (a)from which plant the defective pipe has come ? (b) What is the probability that it has come from the second plant?

Soln. :

Step I: Let A_1 be event that the pipe is manufactured by plant I ,

- A_2 be event that the pipe is manufactured by plant II ,
- A_3 be event that the pipe is manufactured by plant III
- In all plants total pipes produced are $1000 + 1500 + 2500 = 5000$
- Out of these 1000 pipes are produced in plant I

➔ **Using standard formula of probability**

$$\text{...}\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 1000
- Total number of cases = 5000

$$P(A_1) = \frac{1000}{5000} = 0.2$$

i.e. the probability that a pipe is manufactured in plant I,

$$P(A_1) = 0.2$$

- Similarly, the probability that a pipe is manufactured in plant II,

$$P(A_2) = \frac{1500}{5000} = 0.3$$

- Similarly, the probability that a pipe is manufactured in plant II,

$$P(A_2) = \frac{2500}{5000} = 0.2$$

- Let P(D) be probability that a pipe selected is defective
- $P(D \mid A_1)$ be the probability that the pipe manufactured in plant I is defective

$\therefore$ By given information ,$P(D \mid A_1) = 0.04$

- $P(D \mid A_2)$ be the probability that the pipe manufactured in plant II is defective

$\therefore$ By given information , $P(D \mid A_2) = 0.09$

- $P(D \mid A_3)$ be the probability that the pipe manufactured in plant III is defective

$\therefore$ By given information , $P(D \mid A_3) = 0.07$

Step II: $P(A_1)\,P(D \mid A_1) = 0.2 \times 0.04 = 0.008$

$P(A_2)\,P(D \mid A_2) = 0.3 \times 0.09 = 0.027$

$P(A_3)\,P(D \mid A_3) = 0.5 \times 0.07 = 0.035$

Step III: The posterior probabilities of all events are given by

1) Probability that pipe manufactured in plant I is defective

$$= \frac{P(A_1)\cdot P(D \mid A_1)}{P(A_1)\,P(D \mid A_1) + P(A_2)\cdot P(D \mid A_2) + P(A_3)\,P(D \mid A_3)}$$

$$= \frac{0.008}{(0.008) + (0.027) + (0.035)}$$

$$= \frac{0.008}{0.07} = 0.0114$$

2) Probability that pipe manufactured in plant II is defective

$$= \frac{P(A_2)\cdot P(D \mid A_2)}{P(A_1)\,P(D \mid A_1) + P(A_2)\cdot P(D \mid A_2) + P(A_3)\,P(D \mid A_3)}$$

$$= \frac{0.027}{(0.008) + (0.027) + (0.035)}$$

$$= \frac{0.027}{0.07} = 0.3857$$

3) Probability that pipe manufactured in plant III is defective

$$= \frac{P(A_3)\cdot P(D \mid A_3)}{P(A_1)\,P(D \mid A_1) + P(A_2)\cdot P(D \mid A_2) + P(A_3)\,P(D \mid A_3)}$$

$$= \frac{0.035}{(0.008) + (0.027) + (0.035)}$$

$$= \frac{0.035}{0.07} = 0.5 \checkmark \qquad \text{...Ans.}$$

Step IV : From above calculations we observe that most probably,

(a) defective pipe will be manufactured in plant III

(b) The probability that the defective pipe has come from plant II is 0.3857

1.7 Miscellaneous Examples

Ex. 1.7.1 : Professor X and Madam Y appear for an interview. The probability of professor X's selection is $\frac{1}{7}$ and that of Madam Y's selection is $\frac{1}{5}$. Find the probability that only one of them is selected. What is the probability that at least one of them is selected.

Soln. :

Step I: Given :

- Let P(X) be probability that professor X is selected in an interview

$$\therefore\ P(X) = \frac{1}{7}$$

- The probability that professor X is not selected in an interview is $P(\bar{X})$

$$P(\bar{X}) = 1 - P(X) = 1 - \frac{1}{7} = \frac{6}{7}$$

- Let P(Y) be probability that madam Y is selected in an interview

$$\therefore\ P(Y) = \frac{1}{5}$$

- The probability that madam Y is not selected in an interview is $P(\bar{Y})$

$$P(\bar{Y}) = 1 - P(Y) = 1 - \frac{1}{5} = \frac{4}{5}$$

Step II : As only one out of them is selected.

A event → If X is selected, Y is not selected

B event → If Y is selected, X is not selected

Step III :

➔ **Using standard multiplication theorem of probability of independent events A and B,**

$$\text{...}[P(A \cap B) = P(A)\,P(B)]$$

$$\left.\begin{aligned} P(A) &= P(X)\cdot P(\bar{Y}) = \frac{1}{7}\cdot\frac{4}{5} = \frac{4}{35} \\ P(B) &= P(Y)\cdot P(\bar{X}) = \frac{1}{5}\cdot\frac{6}{7} = \frac{6}{35} \end{aligned}\right\}$$

Step IV : Required probability (only one selected)

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

$$= P(A) + P(B) \qquad (\because P(A \cap B) = 0)$$

$$= \frac{4}{35} + \frac{6}{35} = \frac{10}{35} = \frac{2}{7}$$

Step V

(2) Probability that none of them is selected

$$P(\bar{X} \cap \bar{Y}) = P(\bar{X})\cdot P(\bar{Y}) = \frac{6}{7} \times \frac{4}{5} = \frac{24}{35} = q$$

$\therefore$ P(at least one selected) $= 1 - \frac{24}{35} = \frac{11}{35} \checkmark$ ...Ans.

$$(\because\ p + q = 1 \Rightarrow p = 1 - q)$$

Ex.1.7.2 : From a well shuffled pack of cards, three cards are drawn at random. Find the probability that they form a King, Queen, Jack combinations.

Soln. :

Step I : In a pack there are 52 cards.

➔ **Using standard formula** $\text{...}\left[{}^nC_r = \frac{n!}{r! \times (n-r)!}\right]$

Here $n = 52$, $r = 3$

$${}^{52}C_3 = \frac{52!}{3! \times (52-3)!} = \frac{52!}{3! \times 49!}$$

Now,

$$\begin{aligned} 52! &= 1 \times 2 \times 3 \times \ldots\ldots\ldots \times 46 \times 47 \times 48 \times 49 \times 50 \times 51 \times 52 \\ &= 49! \times 50 \times 51 \times 52 \\ &= 52 \times 51 \times 50 \times 49! \end{aligned}$$

$$52C_3 = \frac{52 \times 51 \times 50 \times 49!}{3 \times 2 \times 1 \times 49!} = \frac{52 \times 51 \times 50}{3 \times 2 \times 1}$$

$$= 52 \times 17 \times 25$$

Step II : In a complete pack, there are 4-King , 4-Queen, 4-Jack cards out of which 1 card can be drawn in ${}^4C_1$ ways

$\therefore$ The combination can be chosen in

$${}^4C_1 \times {}^4C_1 \times {}^4C_1 = \frac{4!}{3!}\cdot\frac{4!}{3!}\cdot\frac{4!}{3!} = 4 \times 4 \times 4 = 64$$

Step III : Probability that the three cards drawn are king, Queen and Jack P (King , Queen , Jack)

➔ **Using standard formula of probability**

$$\text{...}\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 64
- Total number of cases = 22100

$$P(\text{King , Queen , Jack}) = \frac{64}{22100}$$

$$= \frac{16}{5525} = 0.00289 \checkmark \qquad \text{...Ans.}$$

Ex. 1.7.3 : Among six books, there are two volume of one book. These books are arranged in a random order on a shelf. Find the probability that the two volumes are always kept together.

Soln. :

Step I : Six books are arranged in 6! ways

$n = 720$ ways

Step II : Let first four places for four books of different volumes and last two places for same volume in $4! \times 2! = 48$ ways.

Step III : Let two volume come together either at 1st and 2nd, 2nd and 3rd, 3rd and 4th, 4th and 5th or 5th and 6th position.

That can be happen in 48 ways.

$\therefore m = 48 \times 5 = 240$

Step IV :

$\therefore$ Probability $= \frac{240}{720} = \frac{1}{3}$ ✓ ...Ans.

That two volume are kept together

Ex. 1.7.4 : An urn contains 6 white and 8 red balls. Second urn contains 9 white and 10 red balls. One ball is drawn at random from the first urn and put it into the second urn without noticing its colour. A ball is then drawn at random from the second urn. What is the probability that it is red ?

Soln. :

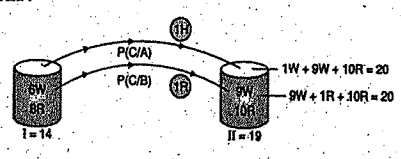


Fig. P. 1.7.1

Step I : There are two possibilities

Let A → be the event of "white ball is drawn from 1st urn and put into the 2nd urn.

B → be the event of red ball is drawn from 1st urn and put it into the 2nd urn.

C → be the event of red ball is drawn from 2nd urn.

A and B are mutually exclusive (dependent event).

Step II : For event A ,in first urn there are 6W + 8R = 14 balls out of which 6 balls are white.

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 6
- Total number of cases = 14

$$P(A) = \frac{6}{14}$$

- In this case now in urn2,there will be 9 white + 10 red + 1 white = 20 balls.
- Probabilty of drawing red ball from second urn when white ball is transformed from first urn is P(C | A)

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 20
- Total number of cases = 10

$$P(C|A) = \frac{1R + 9R}{1 \text{ white} + 9W + 10R} = \frac{10}{20} = \frac{1}{2}$$

- Event A and C are dependent

$$\therefore P(A \cap C) = P(A) \times P(C|A) = \frac{3}{7} \times \frac{1}{2} = \frac{3}{14}$$

Step III : For event B ,in urn 1 there are 6W + 8R = 14 balls out of which 8 balls are red.

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 8
- Total number of cases =14

$$P(B) = \frac{8}{14} = \frac{4}{7}$$

- In this case now in urn2,there will be 9 white + 10 red + 1 red = 20 balls.
- Probability of drawing red ball from second urn when red ball is transformed from first urn is P(C | B)

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 11
- Total number of cases = 20

$$P(C|B) = \frac{1R + 10R}{9W + 1R + 10R} = \frac{11}{20}$$

- Event A and C are dependent

$$\therefore P(B \cap C) = P(B) \times P(C|A) = \frac{4}{7} \times \frac{11}{20} = \frac{11}{35}$$

Step V : $\therefore$ Required probability

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

$$= \frac{3}{14} + \frac{11}{35} = \frac{259}{490} = 0.52857 \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.7.5 : A and B play a game of alternating tossing a coin, one who gets head first wins the game. Find the probability of B winning the game if A has start.

Soln. :

Step I : Given that who will get head first is winner and A has start first.

We have to find the probability of 'B' winning the game.

A → T, B → H T H

A → T, B → T, A → T, B → HTTTH

A → T, B → T, A → T, A → T, B → HTTTTTH

TTTTTTTH .....

Step II : $P(T) = \frac{1}{2}, P(H) = \frac{1}{2}$

$P(TH) = P(T) \cdot P(H)$

Independent events $= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = \frac{1}{2^2}$

$P(TTTH) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2^4}$

$P(TTTTTH) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2^6}$

Step III : $\therefore$ Required probability is $= \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \ldots;$

which is in G. P. with

Common ratio (r) $= \frac{1}{2^2} = \frac{1}{4}$ and $a = \frac{1}{2^2}$

Step IV :

$$\text{Required probability} = \frac{a}{1-r} = \frac{\left(\frac{1}{2^2}\right)}{1 - \left(\frac{1}{2^2}\right)}$$

$$= \frac{1}{3} = 0.02326 \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.7.6 : From a deck of 52 cards, two cards are drawn at random. Find the probability that

(i) Both are Hearts.

(ii) Both the cards are of different series.

Soln. :

Step I : Given, n = 52 , r = 2

${}^nC_r$ ways to draw cards $= {}^{52}C_2 = 1326$

Now, there are 13 hearts in a one pack and we have to select two hearts

$= {}^{13}C_2 = 78$

Step II :

(1) $\therefore P(\text{Both hearts}) = \frac{{}^{13}C_2}{{}^{52}C_2} = \frac{78}{1326} = \frac{39}{663}$

Step III :

(2) There are 4 suits of 13; we select 2 from 4 suits in ${}^4C_2$ ways

1 card from 1 suits (13 cards) $= {}^{13}C_1$

1 card from another suit $= {}^{13}C_1$

Step IV :

$$\therefore \text{Required prob.} = \frac{{}^4C_2 \times {}^{13}C_1 \times {}^{13}C_1}{{}^{52}C_2}$$

$$= \frac{13}{17} = 0.7647 \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.7.7 : There are six married couples in a garden. If two persons are chosen at random. Find the probability that (i) They are of different sex (ii) they are married to each other.

Soln. : Given

Step I : Six married couple. Six men and six women.

To select two persons.

(Number of ways $= {}^6C_1 \cdot {}^6C_1 = 6.6 = 36$)

Step II :

(i) P (they are of different sex)=two persons are selected out of 12

$$= \frac{{}^6C_1 \cdot {}^6C_1}{{}^{12}C_2} = \frac{36}{66} = \frac{6}{11} = 0.54545$$

Step III :

(ii) P (they are married to each other) =

$\therefore$ There are six married couple, we have to select only one pair ${}^6C_1 = 6$ out of ${}^{12}C_2$

$$\frac{{}^6C_1}{{}^{12}C_2} = \frac{6}{66} = \frac{1}{11} = 0.090 \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.7.8 : Two cards are drawn from a well shuffled pack of 52 cards. Find the probability that they are both kings if

(i) The first card drawn is replaced.

(ii) The first card is not replaced.

Soln. :

Case I : When card is drawn is replaced

Step I : Let Abe the event that a king card is drawn first.

B be event that a king card is drawn in second draw when card drawn in first draw is replaced

There are 52 cards out of which 4 cards are king cards.

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 52
- Total number of cases = 4

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

- After first draw if card is replaced that means a card drawn is again put into the pack so that in a pack there will be 52 cards. So drawing any card from this pack is independent that which card is drawn in first draw.
- Events A and B are independent
- As in previous case the probability that a king card is drawn

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

Step II :

➔ **Using standard multiplication theorem on probability of independent events**

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

Case II: If card drawn in first draw is not replaced

Step III : Let B be event that card drawn is king card when first card is not replaced.

When first card drawn is king and if it is not replaced then in pack there will be 52 – 1 = 51 cards.

As first card is king card so in this pack there will remain 4 – 1 = 3 king cards

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 51

– Total number of cases = 3

$$P(B/A) = \frac{3}{51}$$

– Event A and B are dependent.

$$P(A \cap B) = P(A_1) \cdot P(B/A) = \frac{4}{52} \cdot \frac{3}{51} = \frac{1}{13} \cdot \frac{1}{17} = \frac{1}{221} \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.7.9 : An envelope contains 6 tickets with numbers 1, 2, 3, 5, 6, 7. Another envelope contains 4 tickets with numbers 1, 3, 5, 7. An envelope is chosen at random and ticket is drawn from it. Find the probability that the ticket bears the numbers (i) 2 or 5 (ii) 2.

Soln. :

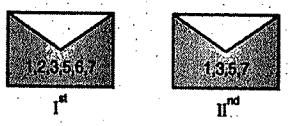


Fig. P. 6.1.11

Case I : Probability that ticket bears numbers 2 or 3

Step I : Let A → be the event of 1st envelope chosen

B → be the event of 2nd envelope chosen

– Probability of choosing the envelope $= \frac{1}{2}$

Step II : In first envelope there are 6 numbers out of which 2 or 3 are at 2 places

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 6

– Total number of cases = 2

P (envelop 1 and having number 2, 3) $= \frac{2}{6}$

– Since probability of choosing the envelope is $\frac{1}{2}$, therefore probability of getting 2 or 3 from first envelop, P(2 or 3)

– Similarly in second envelop there are 4 numbers out of which 2 or 3 is at only 1 place.

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 4

– Total number of cases = 1

P (envelop 2 and having number 2, 3A) $= \frac{1}{4}$

(i) P (2 or 5) = P (envelop 1 and having number 2, 3)

+ P (envelop 2 and having number 2, 3)

$$= \left(\frac{1}{2}\times\frac{2}{6}\right) + \left(\frac{1}{2}\times\frac{1}{4}\right) = \frac{1}{6} + \frac{1}{8} = \frac{14}{48} = \frac{7}{24}$$

Case II : Probability that ticket bears numbers 2

Step III :

(ii) P (2) = P(A and having number 2)

+ P(B and having number 2)

$$= \left(\frac{1}{2}\times\frac{1}{6}\right) + \left(\frac{1}{2}\times\frac{0}{6}\right) = \frac{1}{12} = 0.0833 \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.7.10 : A box contains 6 red balls, 4 white balls and 5 blue balls. Three balls are drawn successively from the box. Find the probability that they are drawn in the order red, white, blue if each ball is not replaced.

Soln. :

Step I :

Let A → be the event of red ball is drawn.

B → be the event of white ball is drawn when red ball is drawn.

C → be the event of blue ball is drawn when one red and one white ball is already drawn.

– In a box there are 6 red , 4 white and 5 blue balls i.e total

6 + 4 + 5 = 15 balls

– In first event , the probability a red ball drawn is P(A)

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 15

– Total number of cases = 6

$$P(A) = \frac{6}{15}$$

Step II: When red ball is drawn , then in a box there remains 15 – 1 = 14 balls

– In these 14 balls there are 4 white balls.

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 14

– Total number of cases = 4

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

Step III: In third event, after drawing one red and one white ball from box, there will remain 15 – 2 = 13 balls. Out of these 13 balls, there are 5 blue balls

➔ **Using standard formula of probability**

$$\ldots\left[P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

– Here number of favourable cases = 13

– Total number of cases = 5

$$P(C/A \cap B) = \frac{5}{13}$$

Step IV: All these events are dependent.

$$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/A \cap B) = \frac{6}{15}\frac{4}{14}\frac{5}{13} = \frac{4}{91} \checkmark \quad \ldots\text{Ans.}$$

Ex. 1.7.11 : A six faced die is so biased that it is two times to show an even number as an odd number, when thrown. It is thrown twice. What is the probability that the sum of the two numbers thrown is even.

Soln. :

Step I : Let us suppose that the probability of an odd number is P.

∴ The probability of an even number is 2P.

∴ P (1) + P (2) + P (3) + P (4) + P (5) + P (6)=1

⇒ P + 2P + P + 2P + P + 2P =1

$$\Rightarrow 9P = 1 \Rightarrow \boxed{P = \frac{1}{9}}$$

∴ Probability of odd number $= \frac{1}{9}$

and probability of an even number $= 2 \times \frac{1}{9} = \frac{2}{9}$

Step II : A die is thrown two times

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

Probability of getting sum of two number is even

Step III :

(i) Ist Row P (1,1) = P (1) P (1) $= \frac{1}{9}\frac{1}{9}$ = P (1,3) = P (1,5)

(ii) IInd Row P (2,2) = P (2,4) = P (2,6) $= 2\left(\frac{1}{9}\right) \cdot 2\left(\frac{1}{9}\right) = \frac{4}{81}$

(iii) IIIrd Row P (3,1) = P (3,3) = P (3,5) $= \frac{1}{9}\frac{1}{9} = \frac{1}{81}$

(iv) IVst Row P (4,2) = P(4,4) = P (4,6) $= \frac{4}{81}$

(v) Vth Row P (5,1) = P(5,3) = P (5,5) $= \frac{1}{81}$

(vi) VIth Row P (6,2) = P (6,4) = P (6,6) $= \frac{4}{81}$

Step IV:

P = P (1,1) + P (1,3) + P (1,5) + P (2,2) + P (2,4) + P (2,6)
+ P (3,1) + P (3,3) + P (3,5) + P (4,2) + P (4,4) + P (4,6)
+ P (5,1) + P (5,3) + P (5,5) + P (6,2) + P (6,4) + P (6,6)

$$= \left[\frac{1}{81}+\frac{1}{81}+\frac{1}{81}\right] + \left[\frac{4}{81}+\frac{4}{81}+\frac{4}{81}\right] + 3\left[\frac{1}{81}\right] + 3\left[\frac{4}{81}\right] + 3\left[\frac{1}{81}\right] + 3\left[\frac{4}{81}\right] = \frac{9}{81} + 9\frac{4}{81} = \frac{45}{81} = \frac{5}{9} \checkmark \ldots\text{Ans.}$$

Ex. 1.7.12 : An urn contains 8 white and 10 red balls. Second urn contains 11 white, 12 red balls. One ball is drawn at random from the first urn and put into the second urn without noticing its colour. A ball is then drawn at random from the second urn. What is the probability that it is white.

Soln. :

Step I :

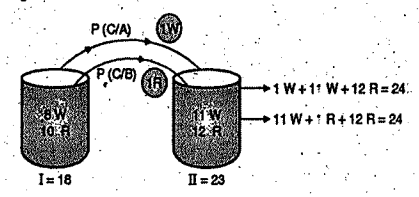


Fig. 1.7.2

Step I : There are two possibilities

Let A → be the event of white ball is drawn from 1st urn and put into the 2nd urn.

B → be the event of red ball is drawn from 1st urn and put it into the 2nd urn.

C → be the event of white ball is drawn from 2nd urn.

A and B are mutually exclusive (dependent event).

Step II: For event A ,in first urn there are 8W + 10R = 18 balls out of which 8 balls are white.

➔ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 8
- Total number of cases = 18

$$P(A) = \frac{8}{18} = \frac{4}{9}$$

- In this case now in urn2,there will be 11 white + 12 red + 1 white = 24 balls.
- Probabilty of drawing white ball from second urn when white ball is transformed from first urn is P(C|A)

➔ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 12
- Total number of cases = 24

$$P(C|A) = \frac{11W+1W}{11\text{ white}+12R+1W} = \frac{12}{24}=\frac{1}{2}$$

- Event A and C are dependent

$$\therefore P(A\cap C) = P(A)\times P(C|A)=\frac{4}{9}\cdot\frac{1}{2}=\frac{2}{9}$$

Step III: For event B in urn 1 there are 8W + 10R = 18 balls out of which 10 balls are red.

➔ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 10
- Total number of cases = 18

$$P(B)=\frac{10}{18} = \frac{5}{9}$$

- In this case now in urn2,there will be 11 white + 12 red + 1 red = 24 balls.
- Probability of drawing white ball from second urn when red ball is transformed from first urn is P(C|B)

➔ **Using standard formula of probability**

$$\ldots\left[P(A)=\frac{\text{Number of favourable cases}}{\text{Total number of cases}}\right]$$

- Here number of favourable cases = 11
- Total number of cases = 24

$$P(C|B) = \frac{11W}{11\text{ white}+12R+1W}=\frac{11}{24}$$

- Event B and C are dependent

$$\therefore P(B\cap C) = P(B)\times P(C|A)=\frac{5}{9}\cdot\frac{11}{24}=\frac{55}{216}$$

Step V: ∴ Required probability

$$P(C) = P(A\cap C)+P(B\cap C)$$
$$= \frac{2}{9}+\frac{55}{216}=\frac{103}{216}=0.4768\checkmark \quad \ldots\text{Ans.}$$

Ex. 1.7.13: According to past record of one day internationals between India and Pakistan. India has won 15 matches and lost 10. If they decide to play a series of 6 matches now, what is the probability of India winning the series? (Draw is ruled out.)

Soln.:

Step I: Past Record

$$P(\text{Probability of India win}) = \frac{15}{25}=0.6$$

$$P(\text{Probability of Pakistan win}) = \frac{10}{25}=0.4$$

Here, draw is ruled out.

Step II: ∴Now there are following events of India winning the series of 6 matches

(i) Four India win and 2 Pakistan.

(ii) Five India win and 1 Pakistan

(iii) All 6 matches win by India.

All these cases are Independent.

Step III:

1. India wins 4 matches and Pakistan 2

$$= [P(w)\cdot P(w)\cdot P(w)\cdot P(w)]\,{}^nC_r$$
$$= (0.6)\times(0.6)\times(0.6)\times(0.6)\times{}^6C_4$$
$$\therefore P(I) = 1.944$$

Step IV:

2. India wins 5 matches and Pakistan 1

$$= [P(w)\cdot P(w)\cdot P(w)\cdot P(w)\cdot P(w)]\times{}^nC_r$$
$$= [(0.6)\times(0.6)\times(0.6)\times(0.6)\times(0.6)]\times{}^6C_5$$
$$\therefore P(II) = 0.46656$$

Step V:

3. All matches win by India

$$= P(w)\cdot P(w)\cdot P(w)\cdot P(w)\cdot P(w)\cdot P(w)$$
$$\therefore P(III) = 0.04665$$

Step VI:

$$\text{Required probability} = P(1)+P(2)+P(3)$$
$$= 1.944+0.46656+0.04665$$
$$= 2.45721\checkmark \quad \ldots\text{Ans.}$$

Ex.1.7.14: A can hit a target 3 times in 5 shots B 2 times in 5 shots and C three times in 4 shots. All of then fire one shot each simultaneously at the target.

What is the probability that (i) 2 shots hit (Ii) At least two shots hit ?

Soln.:

- The probability that A can hit the target is given by

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

- The probability that B can hit the target is given by

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

- The probability that C can hit the target is given by

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

(i) The two shots can hit in following different ways.

(a) A and B hits and C does not hit.

(b) A and C hits and B does not hit

(c) B and C hits and A does not hit

Either of above three cases is possible at a time,

∴ The probability that 2 shots hit the target is given by

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

Since all are in dependent events.

$$\text{Required probability} = P(A)\cdot P(B)\cdot P(\bar{C})$$
$$+P(A)\cdot P(\bar{B})\cdot P(C)$$
$$+P(\bar{A})\cdot P(B)\cdot P(C)$$
$$=\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)\left(1-\frac{3}{4}\right)+\left(\frac{3}{5}\right)\left(1-\frac{2}{5}\right)\left(\frac{3}{4}\right)$$
$$+\left(1-\frac{3}{5}\right)\left(\frac{2}{5}\right)\left(\frac{3}{4}\right)$$
$$=\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)\left(\frac{1}{4}\right)+\left(\frac{3}{5}\right)\left(\frac{3}{5}\right)\left(\frac{3}{4}\right)+\left(\frac{2}{5}\right)\left(\frac{2}{5}\right)\left(\frac{3}{4}\right)$$
$$=\frac{6}{100}+\frac{27}{100}+\frac{12}{100}$$
$$=\frac{6+27+12}{100}$$
$$=\frac{45}{100}=\frac{9}{20}$$

(ii) The probability that at least two hits the target can be possible in following ways.

(a) Any two hits the target

(b) All three hits the target

In this case,

$$\text{Required probability} = P(\text{Any two hits the target})$$
$$+P(\text{All three hits the target})$$
$$= P(A)\cap P(B)\cap P(\bar{C})+P(A)\cap P(\bar{B})\cap P(C)$$
$$+P(\bar{A})\cap P(B)\cap P(C)+P(A)\cap P(B)\cap P(C)$$
$$=\frac{9}{20}+P(A)\cdot P(B)\cdot P(C)$$
$$=\frac{9}{20}+\left(\frac{3}{5}\cdot\frac{2}{5}\cdot\frac{3}{4}\right)$$
$$=\frac{9}{20}+\frac{18}{100}$$
$$=\frac{9}{20}\times\frac{5}{5}+\frac{18}{100}$$
$$=\frac{45}{100}+\frac{18}{100}$$
$$=\frac{63}{100}\checkmark \quad \ldots\text{Ans.}$$

Ex.1.7.15: The probability that machine A will be performing an usual function in 5 years is $\frac{1}{4}$. While the probability that machine B will still be operating usefully at the end of the same period is $\frac{1}{3}$.

Find the probability in the following cases that in 5 years time,

(i) Both machines will be performing an usual function.

(ii) Neither will be operating

(iii) Only machine B will be operating.

(iv) At least one of the machine will be operating.

Soln.:

- The probability that machine A will be operating in 5 years is

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

- The probability that machine B will be operating in 5 years is

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

- The probability that machine A will not work,

$$P(\bar{A}) = 1-P(A).$$
$$= 1-\frac{1}{4}=\frac{4-1}{4}=\frac{3}{4}$$

- The probability that machine B will not work,

$$P(\bar{B}) = 1-P(B)=1-\frac{1}{3}$$
$$=\frac{3-1}{3}=\frac{2}{3}$$

Now,

(i) The probability that both machine will be operating

$$P(A\cap B) = P(A)\cdot P(B)$$
$$=\left(\frac{1}{4}\right)\left(\frac{1}{3}\right)=\frac{1}{12}$$

(ii) The probability that neither of machine will work

= Complement of both machine will work

$$\therefore P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B})$$
$$= \left(\frac{3}{4}\right)\left(\frac{2}{3}\right)$$
$$= \frac{1}{2}$$

(iii) the probability that only machine B will be operating miens by default machine A is not operating is given by

$$P(\bar{A} \cap B) = P(\bar{A}) \cdot P(B) = \left(\frac{3}{4}\right)\left(\frac{1}{3}\right)$$
$$= \frac{1}{4}$$

(iv) The probability that at least are machine is working means either only A is working or only B is working or A and B both working.

$\therefore$ P (at least one machine is working)

$$= P(A \cap \bar{B}) + P(\bar{A} \cap B) + P(A \cap B)$$
$$= P(A) \cdot P(\bar{B}) + P(\bar{A}) \cdot P(B) + P(A) \cdot P(B)$$
$$= \left(\frac{1}{4}\right)\left(\frac{2}{3}\right) + \left(\frac{3}{4}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{3}\right)$$
$$= \frac{1}{\cancel{4}_2} \times \frac{\cancel{2}^1}{3} + \frac{\cancel{3}}{4} \cdot \frac{1}{\cancel{3}} + \frac{1}{4} \cdot \frac{1}{3}$$
$$= \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{4} + \frac{1}{12}$$
$$= \frac{1}{6} + \frac{1}{4} + \frac{1}{12} = \frac{2+3+1}{12} = \frac{6}{12} = \frac{1}{2} \checkmark \quad \text{...Ans.}$$

Ex. 1.7.16 : An urn contains 10 black and 10 white balls. Find the probability of drawing two balls of the same colour.

Soln. :

Step I : In an urn there are 10 black + 10 white = 20 balls. 2 balls from 20 can be drawn in $^{20}C_2$ ways.

$$\therefore {}^{20}C_2 = \frac{20!}{2!\,(20-2)!} = \frac{20!}{2!\,18!}$$
$$= \frac{20 \times 19 \times 18!}{2 \times 1 \times 18!}$$
$$= \frac{20 \times 19}{2} = 190$$

$\therefore$ Total number of ways = 1.90

Step II : Two balls drawn are of same colour are either both black or both white balls.

- Let A be event that both balls drawn are black.
- B be event that both balls drawn are white.,
- The required probability is given by P (a $\cup$ B).
- As both events are independent by addition law of probability of independent events,

$$P(A \cup B) = P(A) + P(B) \quad (\therefore P(A \cap B) = 0) \quad \text{...(i)}$$

Step III : The number of ways of drawing both black balls are $^{10}C_2$.

$$^{10}C_2 = \frac{10!}{2!\,(10-2)!} = \frac{10!}{2!\,8!} = \frac{10 \times 9 \times 8!}{2 \times 1 \times 8!}$$
$$= \frac{10 \times 9}{2} = 45$$

Step IV : Here $P(A) = \frac{45}{190}$

$$= \frac{9}{38}$$

Step V : The number of ways of drawing both white balls are $^{10}C_2$

$$^{10}C_2 = 45$$
$$\therefore P(B) = \frac{45}{190} = \frac{9}{38}$$

Step VI : From equation (i)

$$P(A \cup B) = \frac{9}{38} + \frac{9}{38} = \frac{18}{38} = \frac{9}{38} \checkmark \quad \text{...Ans.}$$

Ex. 1.7.17 : An urn contains nine balls, two of which are red, three blue and four black. Three balls are drawn from the urn at random.

What is the probability that.

(i) The three balls are of different colour?

(ii) The three balls are of same colour ?

Soln. :

Step I : An urn contains 2 red, 3 blue and 4 black = 9 balls. The three balls out of 9 can be drawn in $^9C_3$ ways.

$$\therefore {}^9C_3 = \frac{9!}{3!\,(9-3)!} = \frac{9!}{3!\,6!}$$
$$= \frac{9 \times 8 \times 7 \times 6!}{3 \times 2 \times 1 \times 6!} = \frac{9 \times 8 \times 7}{3 \times 2}$$
$$= \frac{3 \times 4 \times 7}{1 \times 1} = 84$$

Step II :

(i) The probability that three balls are of different colour are possible as follows.

- 1 red, 1 blue and 1 black ball.
- 1 red ball out of 2 red can be drawn in 2_{C_1} ways.

$$^2C_1 = 2$$

- 1 blue ball out of 3 can be drawn in 3_{C_1} ways.

$$^3C_1 = 3$$

- 1 black ball from 4 can be drawn in 4_{C_1} ways.

$$\therefore {}^4C_1 = 4$$

Step III : The number of ways of getting 1 red, 1 blue and 1 black ball are

$$^2C_1 \times {}^3C_1 \times {}^4C_1 = 2 \times 3 \times 4 = 24$$

Step IV : The probability of drawn three balls of different colour is

$$P(A) = \frac{24}{84} = \frac{2}{7}$$

Step V :

(ii) The probability that the three balls drawn are of same colour.

Now this possible only in two ways.

- Either three balls are blue or three balls are black. (since there are only two red balls)
- The three 3 blue balls from total 3 blue balls can be drawn in 3_{C_3} ways.

$$^3C_3 = 1.$$

- The probability of drawing 3 blue balls is

$$P(A) = \frac{7}{84}$$

Step VI : The three black balls from total 4 black balls can be drawn in 4_{C_3} ways.

$$^4C_3 = \frac{4!}{3!\,(4-3)!} = \frac{4!}{3!\,1!}$$
$$= \frac{4 \times 3!}{3! \times 1} = 4$$

- The probability of drawings 3 black balls is

$$P(B) = \frac{4}{84}$$

Step V : Since the probability of drawing 3 blue or 3 black balls are independent

$$\therefore P(A \cap B) = 0$$

Step VI : $P(A \cup B) = P(A) + P(B) - P(A \cap b)$

$$= \frac{1}{84} + \frac{4}{84} - 0 = \frac{1+4}{84}$$
$$= \frac{5}{84} \checkmark \quad \text{...Ans.}$$

Ex. 1.7.18 : A bag contain 4 white and 2 black balls and a second bag contains three of each colour. A bag is selected at random and a ball is then drawn at random from the bag chosen. What is the probability that the ball drawn is white ?

Soln. :

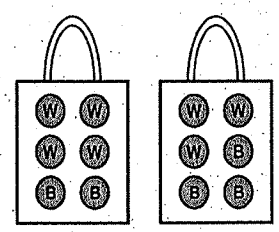


Fig. P. 1.7.18

Step I : Let A be event that a white ball is drawn from bag 1.

B be the event that a white ball is drawn from bag 2.

Step II : Since there are 4 white balls out of

$$(4+2) = 6 \text{ balls in bag 1}$$
$$\therefore P(A) = \frac{4}{6} = \frac{2}{3}$$

Step III : There are 3 white balls out of

$$(4+2) = 6 \text{ balls in bag 1}$$
$$P(B) = \frac{3}{6} = \frac{1}{2}$$

- Since also probability of selecting anyone bag from 2 bags is $\frac{1}{2}$
- $\therefore$ Since event of selecting white balls from any one bag are independent

$$P(A \cap B) = 0$$

- From addition law of probability of independent events,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= \frac{1}{2} \times \left(\frac{2}{3}\right) + \frac{1}{2} \times \left(\frac{1}{2}\right)$$
$$= \frac{1}{3} + \frac{1}{4}$$
$$= \frac{4+3}{12} = \frac{7}{12} \checkmark \quad \text{...Ans.}$$

Ex. 1.7.19 : A box A contains 2 white and 4 black balls Another box B contains 5 white and 7 black balls. A ball is transferred from the box A to the box B. Then a ball is drawn from the box B. Find the probability that it is white.

Soln. :

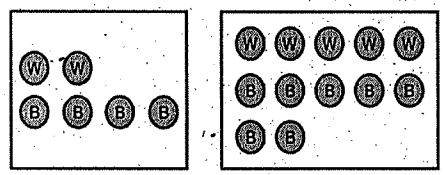


Fig. P. 1.7.19

Step I : Initially in Box A there are 2 White and 4 Black = 10 balls. In Box B there are 5 white and 7 black = 12 balls.

There are two possible cases.

- Case I : Ball transferred from Box A is white and
- Case II : Ball k transferred from Box A is Black.

Case I : A ball transferred is white.

- In this case the number of balls in box A will remain 1 + 4 = 5
- In box B, one white ball will get added

$\therefore$ Total number of balls in box B are 12 + 1 = 13.

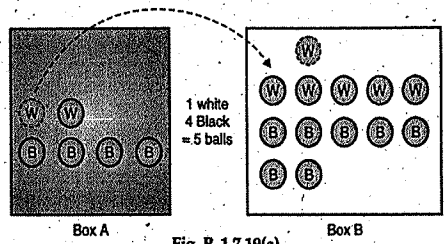


Fig. P. 1.7.19(a)

- Now probability of selecting white ball from box A $= \frac{2}{6}$.
- Probability of drawing white ball from box B $= \frac{6}{13}$.

$\therefore$ Probability of transfer of white ball from box A and then drawing white ball from box B $= \frac{2}{6} \times \frac{6}{13} = \frac{2}{13}$

Step II :

Case II : A ball transferred is black

- In this case the number of balls in box A will remain
- In box B, one black ball get added

$\therefore$ Total number of balls in box B are 12 + 1 = 13

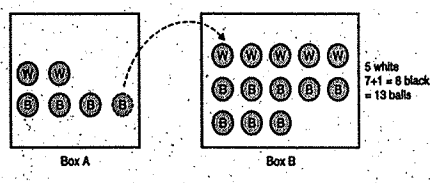


Fig. P. 1.7.19(b)

- The probability of selecting black ball from box A $= \frac{4}{6}$.
- The probability of selecting white ball from box B $= \frac{5}{13}$.

$\therefore$ The probability of transfer of black ball from box A and drawing white ball from box B.

$$B = \frac{4}{6} \times \frac{5}{13} = \frac{10}{39}$$

Step III : Either of above cases are is possible and both cases are independent.

$$\therefore \text{ Required probability } = \frac{2}{13} + \frac{10}{39}$$

$$= \frac{2 \times 3}{13 \times 3} + \frac{10}{39}$$

$$= \frac{6}{39} + \frac{10}{39}$$

$$= \frac{6+10}{39} = \frac{16}{39} \checkmark \quad \text{...Ans.}$$

Multiple Choice Questions

Q. 1 Assuming that a box contains in all ten balls of which six are green and four are red,what is the probability of drawing a red ball?

(a) 0.1 (b) 0.5 (c) 1 (d) 0.4

Ans. : (d)

Q. 2 What is the probability of getting an even number when a die is tossed?

(a) $\frac{1}{3}$ (b) $\frac{1}{2}$ (c) $\frac{1}{6}$ (d) 1

Ans. : (b)

Q. 3 What is the probability of getting more than 2 when a die is tossed?

(a) $\frac{4}{6}$ (b) $\frac{1}{2}$ (c) $\frac{2}{6}$ (d) $\frac{1}{3}$

Ans. : (a)

Q. 4 What is the probability of getting a total of 5 when a pair of die is tossed simultaneously?

(a) $\frac{1}{36}$ (b) $\frac{1}{12}$ (c) $\frac{4}{9}$ (d) $\frac{1}{9}$

Ans. : (d)

Q. 5 If probability that Narendra will solve the problem is $\frac{2}{3}$ and probability that Rahul will solve it is $\frac{3}{4}$. The probability that neither will solve the problem is

(a) $\frac{11}{12}$ (b) $\frac{1}{12}$ (c) $\frac{5}{12}$ (d) $\frac{6}{12}$

Ans. : (b)

Q. 6 If probability that Narendra will solve the problem is $\frac{2}{3}$ and probability that Rahul will solve it is $\frac{3}{4}$. The probability that the problem will be solved is

(a) $\frac{11}{12}$ (b) $\frac{1}{12}$ (c) $\frac{5}{12}$ (d) $\frac{6}{12}$

Ans. : A

Q. 7 A husband and wife appear in an interview for two vacancies in the same post. The probability of husband's selection is $\frac{1}{7}$ and that of wife's selection is $\frac{1}{5}$. What is the probability that both of them is selected is

(a) $\frac{2}{35}$ (b) $\frac{24}{35}$ (c) $\frac{2}{7}$ (d) $\frac{1}{35}$

Ans. : (d)

Q. 8 A husband and wife appear in an interview for two vacancies in the same post. The probability of

husband's selection is $\frac{1}{7}$ and that of wife's selection is $\frac{1}{5}$. Find the probability that neither of them is selected

(a) $\frac{2}{35}$ (b) $\frac{2}{7}$ (c) $\frac{24}{35}$ (d) $\frac{1}{35}$

Ans. : (c)

Q. 9 If $P(A \cap B) = 0$ then the two events A and B are said to be

(a) Dependent (b) Independent
(c) Equally likely (d) None of above

Ans. : (b)

Q. 10 Baye's theorem is useful in

(a) Computing conditional probabilities
(b) Computing sequential probabilities
(c) Revising prior probabilities (d) None of above

Ans. (c)

Q. 11 What is the probability of drawing an ace card from bunch of playing cards?

(a) $\frac{1}{10}$ (b) $\frac{1}{13}$ (c) $\frac{1}{52}$ (d) $\frac{17}{51}$

Ans. : (b)

Q. 12 An unbiased coin is tossed three times. The probability of getting at least two heads is

(a) $\frac{1}{2}$ (b) $\frac{3}{8}$ (c) $\frac{1}{8}$ (d) 1

Ans. : (a)

Q. 13 An unbiased coin is tossed three times. The probability of getting exactly two tails is

(a) $\frac{1}{2}$ (b) $\frac{2}{8}$ (c) $\frac{1}{8}$ (d) $\frac{3}{8}$

Ans. : (d)

Q. 14 The probability that machine A will be performing an usual function in 5 years is $\frac{1}{4}$. While the probability that machine B will still be operating usefully at the end of the same period is $\frac{1}{3}$.The probability that in next 5 year , no machine will be operating is

(a) $\frac{1}{2}$ (b) $\frac{3}{4}$ (c) $\frac{2}{3}$ (d) $\frac{1}{12}$

Ans. : (a)

Q. 15 The probability that machine A will be performing an usual function in 5 years is $\frac{1}{4}$. While the probability that machine B will still be operating usefully at the end of the same period is $\frac{1}{3}$.The probability that in next 5 year , only machine A will be operating is,

(a) $\frac{1}{4}$ (b) $\frac{1}{6}$ (c) $\frac{1}{2}$ (d) $\frac{1}{12}$

Ans. : (b)

Chapter Ends...

□□□

CHAPTER 2

UNIT II

Random Variable and Mathematical Expectation

Syllabus

Syllabus

Random variables, Probability distributions, Probability mass function, Probability density function, Mathematical expectation, Join and marginal probability distributions, Properties of expectation and variance with proofs.

2.1 Introduction

- As we know about the random experiments, sample space; probability of any event (E). Sample space is the set of all possible outcomes of a random experiment.

  $S_1 = \{0, 1\}$

  $S_2 = \{0, 1, 2, 3\}$,

  $S_3 = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$

  $S_4 = \{\text{Head, Tail}\}$

  $S_5 = \{\text{Success, failures}\}$,

  $S_6 = \{\text{black, white, blue}\}$.
- Thus, some experiments have sample spaces whose elements are numbers and some are not numerical value.
- However, a simple and easiest way to express the outcome of an experiment is, through random variable. Through the notion of random variables, we study the experiments whose outcomes may be described numerically.
- Random variables provide a unique approach through which we can express an outcome of an experiment in terms of numbers. Besides this, random variables also provide a more compact description of an experiment of sample space.
- For example, in the inspection of manufactured products, we are interested only in the total number of defective items, not in the types of defects i.e. it allows us to discard unimportant description. After define clearly random variables, we will consider the concept of probability mass function, mathematical expectation and their properties.

Syllabus Topic : Random Variables

2.2 Random Variables

- A random variable is the outcome or an experiment, expressed in terms of numerical value.
- We know, in most of the problems, the outcomes of an experiment is a number. e.g. the number of defective bolts, the salary of a person, number of day of week. That means, when the outcomes of an experiment is not a number, we can express these outcomes in numbers to denote.
  - (i) Head by 1 and tail by 0.
  - (ii) Yes by 1 and No by 0.
  - (iii) Monday by 1, Tuesday by 2, Wednesday by 3, ..., and Sunday by 7.
- So, in a probability theory, we first find the variable and value of the variable, which describe the outcomes of the experiments.
- In example 1, the toss of coin, the outcomes { H, T } is a non-numerical in nature. We can introduce a variable value 1 for head and variable value 0 for tail.
- In example 2, in the answers of the questions, variable takes the value 1 for yes and 0 for No.
- In example 3, the variable takes the value 1 for Monday, 2 for Tuesday, ... and 7 for Sunday. So, such variable is called **Random** variable.
- A random variable is a rule that assign a numerical value to each possible outcome of an experiment. Usually, random variables are denoted by capital letters X, Y, Z, ...etc. and the values of these random variables

by small letters x, y, z ... Random variable X (say) is a function which assigns a real number to the outcome of an experiment.
- Rather, we say, random variable X is a function whose domain is the sample space (S), and whose range is the set of all real numbers $\mathbb{R}$. The set of all values taken by X, called **image** of X. (images are the subset of the set of all real numbers).

Ex. 1 : In rolling of die, we observe the face value. Here, the outcome is numerical. Here, random variable X-denotes the face value of the rolled die.

Ex. 2 : In two coins experiment, the sample space (outcomes) S = { HH, HT, TH, TT } could assign a random variable, say, X = number of times getting head out of 2 tosses.

Therefore, X (HH) = 2; X (HT) = 1; X (TH) = 1 and X (TT) = 0. Hence, first we expect to perform the random experiment E, observe the outcomes (sample space $s \in S$ and then find X (s) i.e. assign a real number x to X (s).

Definition :

Let,
- E – be the experiment,
- S – be the sample space,
- s – Element of sample space (S),
- X – Random variable,
- x – Real number,
- $\mathbb{R}$ – Set of real number.

(i) A random variable (r. v.)

X on a sample space S is a function, that assigns a real number (X (s) = x) to every element $s \in S$. Mathematically, we say that

$$X(s) : S \to \mathbb{R}$$

i.e. the function X (s) = x that maps the elements of the sample space into real numbers is called random variable, associated with the experiment. For a given random variable X, the set or collection of all possible values that X can assume is called range of X and it is denoted by R_x. So, random variable X is a function whose domain is a sample space of an experiment and range set R_x is a set of real number (real number line). ($-\infty < x < \infty$).

Fig. 2.2.1

Random variable X satisfies :

(i) The range set R_X { X (s) ≤ x } is an event in S for every real number x.

(ii) **Probability Function :** The probability of an event { X ≤ x } is denoted by P (X ≤ x) is equal to the sum of the probabilities of all elementary events corresponding to { X ≤ x }.

(iii) P (X = −∞) = 0 = P (X = ∞). It is possible X (s) = −∞ or X (s) = ∞ but its probability is zero.

(iv) X (s) is a function, but we consider it is a variable, i.e. we are not interested in the functional nature of X but interested in the values of X.

(v) The random variable X is a single valued. (i.e. for every $s \in S$, there is only one value of X. Different elements of S, may have same value of X).

Ex. 3 : A fair coin is tossed twice, sample space S = { HH, HT, TH, TT }

X = Random variable denoting number of heads.

X (HH) = 2 ; X (HT) = X (TH) = 1 ;

X (TT) = 0. $R_X = \{0, 1, 2\}$

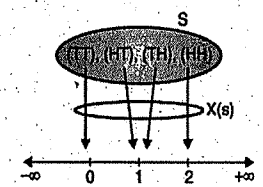


Fig. 2.2.2

Ex. 4 : Suppose the experiment E is to record a time required for project. If X = denote the time then it values from 0 to ∞.

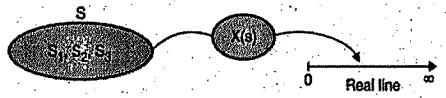


Fig. 2.2.3

In example 3, X takes discrete values and in example 4, X takes continuously all values from 0 to ∞.

2.2.1 Types of Random Variables

From the technical point of view, random variable X can be classified into two categories :

(i) Discrete random variable.
(ii) Continuous random variable.

depending upon the nature of domain.

(i) Discrete Random Variable :

- If the range of a random variable has only finitely many values or countable many values then the random variable X is called discrete random variable.
- i.e. If a random variable X takes finite or countable infinite values $x_0, x_1, x_2, \ldots x_n \ldots$ then, X is called discrete random variable or **discrete stochastic variable.**

- In above example 3, random variable X is called discrete random variable. Most of times, a discrete random variable is the result of counting process, but not always.

(ii) Continuous Random Variable

- If the range of the random variable has uncountable many values over an interval, then X is called continuous random variable.
- i.e. If random variable X can takes all values (infinite number of valves) in an intervals, then X is called a continuous random variable.
- In example 4, the random variable X is called continuous random variable. A continuous random variables usually occur in measurement process such as, height, weight, lifetime, waiting time ... etc.

Syllabus Topic : Probability Distributions

2.3 Probability Distribution of a Random Variable

Let, X – be the random variables which takes the values in the range $R_X = \{x_1, x_2, x_3, \ldots, x_n, \ldots\}$. Collection of the probabilities $\{P(X = x_1), P(X = x_2), \ldots, P(X = x_n)\}$ is called **the probability distribution function of X or cumulative distribution function** of the random variable X.

$$\therefore F_X(x) = P(X \le x); \text{ for } -\infty < x < \infty.$$

i.e. $F_X(x)$ is the probability that X takes any value in the set $(-\infty, x)$.

Ex. 5 : A fair die is rolled. Let x = the face value of the die. $R_X = \{1, 2, 3, 4, 5, 6\}$.

The probability distribution is

$P(x=1) = \frac{1}{6}$; $P(x=2) = \frac{1}{6}$, $P(x=3) = \frac{1}{6}$,

$P(x=4) = \frac{1}{6}$; $P(x=5) = \frac{1}{6}$; $P(x=6) = \frac{1}{6}$.

Ex. 6 : A fair coin is tossed twice.

X = number of times head appear.

$R_x = \{0, 1, 2\}$ is the range set

The probability distribution is

$$P(x=0) = P(TT) = \frac{1}{4};$$

$$P(X=1) = P(TH) = P(HT) = \frac{2}{4};$$

$$P(X=2) = P(HH) = \frac{1}{4}.$$

Syllabus Topic : Probability Mass Function

2.4 Probability Mass Function

- If X – be the discrete random variable which can takes the values $x_1, x_2, \ldots x_n, \ldots$ Such that $P(X = x_i) = p_i$. Then p_i – is called probability mass function or point probability function, provided,
  (i) $p_i \ge 0$ for all i and
  (ii) $\sum p_i = 1$
  $\therefore F_X(x) = P(X \le x)$ for $-\infty < x < \infty$.
- By the definition of distribution function $F_X(x)$ we have,

$$F(X = x_1) = p(x_1)$$
$$F(X = x_2) = p(x_1) + p(x_2)$$
$$F(X = x_3) = p(x_1) + p(x_2) + p(x_3)$$
$$\vdots$$
$$F(X = x_1) = p(x \le x_i)$$
$$= p(x_1) + p(x_2) + p(x_3) + \ldots + p(x_i)$$

For discrete random variable

If X is discrete random variable then the graph of distribution function $F_X(x)$ is a **staircase form**.

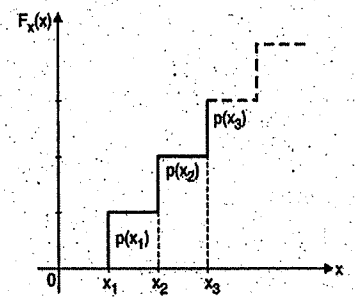


Fig. 2.4.1

$$\therefore P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 3)$$

For continuous random variable

If X is a continuous random variable then $P(X \le 0) = 0$ and for $(0 < X < x)$, probability increases linearly in (0, k) and $P(X \ge k) = 1$.

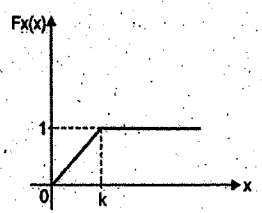


Fig. 2.4.2

It is a non-decreasing function.

2.5 Probability Density Function

- Probability density function is the derivative of the distribution function $F_X(x)$ w.r.t. x.
- i.e. If X a continuous random variable and $F_X(x)$ is a distribution function then $\frac{d}{dx} F_X(x)$ is called probability density and it is denoted by $f_X(x)$.

$$\therefore f_X(x) = \frac{d}{dx} F_X(x)$$

Ex. 2.5.1 : For random variable X, find the distribution function $F_X(x)$.

X	1	2	3	4
P(X = x)	$\frac{3}{20}$	$\frac{7}{20}$	$\frac{9}{20}$	$\frac{1}{20}$

Soln. :

Step I : Given data :

X	$x = x_1 = 1$	$x = x_2 = 2$	$x = x_3 = 3$	$x = x_4 = 4$
P(X = x)	$\frac{3}{20}$	$\frac{7}{20}$	$\frac{9}{20}$	$\frac{1}{20}$

Step II :

$$F_X(x \le 1) = P(X = 1) = \frac{3}{20}$$

$$F_X(x \le 2) = P(X = 1) + P(X = 2) = \frac{3}{20} + \frac{7}{20} = \frac{10}{20}$$

$$F_X(x \le 3) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{3}{20} + \frac{7}{20} + \frac{9}{20} = \frac{19}{20}$$

$$F_X(x \le 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = \frac{3}{20} + \frac{7}{20} + \frac{9}{20} + \frac{1}{20} = \frac{20}{20} = 1$$

2.5.1 Properties of a Distribution Function

(X – is discrete random variable).

(i) $0 \le F_X(x) \le 1$ for $-\infty < x < \infty$, it follows because $F_X(x)$ is the probability of X and it is unity.

(ii) $F_X(x)$ is a non-decreasing function of x. i.e. for $x_1 \le x_2$ then $F_X(x_1) \le F_X(x_2)$.

(iii) $\lim_{x \to \infty} F_X(x) = 1$ and $\lim_{x \to -\infty} F_X(x) = 0$

i.e. $F_X(\infty) = P(X \le \infty) = 1$ and

$F_X(-\infty) = P(X \le -\infty) = 0$

(iv) $F_X(x)$ has positive jump which is equal to probability $P(x_i)$, i = 1, 2, 3, ...

$F_X(x)$ is always constant in $[x_i, x_{i+1}]$.

(v) $F_X(x) = F_X(x_i)$ for $x_i \le x \le x_{i+1}$

and $F_X(X_{i+1}) = F_X(x_i) + p(x_{i+1})$

(vi) If X is a discrete random variable then distribution function $F_X(x)$ is a step function.

2.5.2 Properties of Probability Density Function

Probability density function $f_X(x)$ satisfies :

(i) $f_x(x) \ge 0$ for all x.

This indicates that the probability density curve $y = f_x(x)$ lies above X-axis only.

(ii) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

i.e. total area under density curve from $(-\infty < x < \infty)$ is unity.

(iii) $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt$;

Where, $-\infty < x < \infty$ is distribution function.

(iv) $P(X \in (a, b)) = P(a \le X \le b)$

$$= P(X \le b) - P(X \le a)$$
$$= \int_{-\infty}^{b} f_x(t)\,dt - \int_{-\infty}^{a} f_x(t)\,dt$$
$$= \int_{a}^{b} f_X(t)\,dt$$

i.e. the probability from $a \le X \le b$ is the area under the curve from x = a to x = b.

(v) For continuous random variable P (x = a) is always zero.

$$\because P(x = a) = \int_{a}^{a} f_X(t)\,dt = 0$$

(This is not the case for discrete random variable).

Since, it is continuous, not jumps.

(vi) Any function f (x) for real variable x, which satisfies conditions (i) and (ii), becomes probability density function.

A random variable X is called mixed type, if it's distribution function has both jumps and continuous growth.

Ex. 2.5.2 : For continuous random variable X, the probability density function

$f_X(x) = kx^2$; $0 \le x \le 2$ then find.

(i) The value of k.

(ii) Probability for $0.2 \le X \le 0.5$

(iii) Probability for $X \ge 0.75$

(iv) Distribution function $F_X(x)$

Soln. :

Step I : Given, $f_X(x) = k x^2$ for $0 \le x \le 2$

Since, we know, total probability is unity.

$$\int_a^b f_X(x)\,dx = 1$$

$$\therefore \int_{a=0}^{b=2} kx^2\,dx = 1$$

$$\Rightarrow K\left(\frac{x^3}{3}\right)_0^2 = 1 \Rightarrow K\left(\frac{2^3}{3} - 0\right) = 1$$

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

Step II : Probability for $0.2 \le X \le 0.5$

$$P(0.2 \le X \le 0.5) = \int_{a=0.2}^{b=0.5} K x^2\,dx$$

$$= \frac{3}{8}\int_{0.2}^{0.5} x^2 \cdot dx = \frac{3}{8}\left(\frac{x^3}{3}\right)_{0.2}^{0.5}$$

$$= \frac{1}{8}[(0.5)^3 - (0.2)^3]$$

$$= \frac{1}{8}[0.13 - 0.008] = \frac{1}{8}(0.12)$$

$$= 0.015$$

Step III : Probability for $X \ge 0.75$

$$P(X \ge 0.75) = \int_x^b Kx^2\,dx = \frac{3}{8}\int_{x=0.75}^{2} x^2\,dx$$

$$= \frac{3}{8}\left(\frac{x^3}{3}\right)_{0.75}^{2} = \frac{1}{8}(2^3 - (0.75)^3)$$

$$= \frac{1}{8}(8 - 0.42) = 0.95$$

Step IV : Distribution function $F_X(x)$

By definition, $F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$

$$= \int_{x=0}^{x} Kx^2\,dx = \int_{x=0}^{x} \frac{3}{8}x^2\,dx$$

$$= \frac{3}{8}\left(\frac{x^3}{3}\right)_0^x = \frac{1}{8}(x^3), 0 \le x \le 2$$

Ex. 2.5.3 : For continuous random variable X, the probability density function

$f_X(x) = K(1 - x^2)$ in $0 \le x \le 2$

Find (i) The value of K

(ii) Probability distribution function $F_X(x)$

Soln. :

Step I : Given, probability density function $f_X(x)$

$f_X(x) = K(1 - x^2)$ in $0 \le x \le 2$

Step II : To find value of K

Since, $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ ($\because$ Total area is unity)

$$K\int_{x=0}^{2}(1 - x^2)dx = 1$$

$$K\left(x - \frac{x^3}{3}\right)_0^2 = 1 \qquad \Rightarrow K\left[2 - \frac{8}{3}\right] = 1$$

$$\Rightarrow K(6 - 8) = 3 \qquad \therefore K(-2) = 3$$

$$\Rightarrow K = -\frac{3}{2}$$

Step II : The distribution function $F_X(x)$, is

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx = \int_0^x K(1 - x^2)\,dx$$

$$= -\frac{3}{2}\int_0^x (1 - x^2)\,dx = -\frac{3}{2}\left[x - \frac{x^3}{3}\right]_0^x dx$$

$$F_X(x) = -\frac{3}{2}\left(x - \frac{x^3}{3}\right)$$

$$F_X(x) = \frac{3}{2}\left(\frac{x^3}{3} - x\right)$$

Ex. 2.5.4 : Find the value of K, if the probability density function $f_X(x) = Kxe^{-4x^2}$; $0 \le x \le \infty$.

Soln. :

Step I :

Given, $f_X(x) = Kx\,e^{-4x^2}$; $0 \le x \le \infty$

Since, $\int_a^b f_X(x)\,dx = 1$

$$\Rightarrow \int_{x=0}^{\infty} Kx\,e^{-4x^2}\,dx = 1$$

$$\Rightarrow K\int_0^{\infty} x\,e^{-4x^2}\,dx = 1$$

Put $4x^2 = t$;

$8x\,dx = dt$

$x = 0$; $t = 0$

$x = \infty$; $t = \infty$

$$\therefore K\int_{t=0}^{\infty} e^{-t}\cdot\frac{dt}{8} = 1$$

$$\Rightarrow \frac{K}{8}[e^{-t}]_0^{\infty} = 1$$

$$\Rightarrow \frac{K}{8}[0 + 1] = 1$$

$$\Rightarrow \frac{K}{8} = 1 \Rightarrow K = 8$$

Ex. 2.5.5 : Find $P(1 \le X \le 3)$; $P(x \le 1.5)$, the probability density function

$$f_X(x) = \begin{cases} 2e^{-2x} & ; \quad x \ge 0 \\ 0 & ; \quad \text{for } x < 0 \end{cases}$$

Soln. :

Step I :

Given, $f_X(x) = \begin{cases} 2e^{-2x} & ; \quad x \ge 0 \\ 0 & ; \quad \text{for } x < 0 \end{cases}$

Since $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

$$\therefore P(1 \le x \le 3) = \int_{x=1}^{3} f_X(x)\,dx = \int_1^3 2e^{-2x}\,dx$$

$$= 2\left(\frac{e^{-2x}}{-2}\right)_1^3$$

$$= -[e^{-6} - e^{-4}] = e^{-4}(1 - e^{-2})$$

Step II :

$$P(x \le 1.5) = \int_0^{1.5} f_X(x)\,dx = \int_0^{1.5} 2e^{-2x}\,dx$$

$$= 2\left[\frac{e^{-2x}}{-2}\right]_0^{1.5} = -(e^{-2x})_0^{1.5}$$

$$= -[e^{-2(1.5)} - 1]$$

$$P(x \le 1.5) = [1 - e^{-3}] \qquad \ldots\left(\because 1.5 = \frac{3}{2}\right)$$

$$= 0.95021$$

Ex. 2.5.6 : In a certain region, the daily icefall (in inches) follows continuous random variable X with probability density function $f_X(x)$ given by

$$f_X(x) = \begin{cases} \frac{3}{4}(2x - x^2) & ; \quad 0 \le x \le 2 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Then find (i) $P(x < 1.5)$

(ii) Greater than 1.5 inches

(iii) Between 0.5 and 1.5 inches

Soln. :

Step I : Given probability density

Function $f_X(x) = \begin{cases} \frac{3}{4}(2x - x^2) & ; \quad 0 \le x \le 2 \\ 0 & ; \quad \text{otherwise} \end{cases}$

(1) $$P(X < 1.5) = \int_0^{1.5} \frac{3}{4}(2x - x^2)\,dx$$

$$= \frac{3}{4}\left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_0^{1.5} = \frac{3}{4}\left[x^2 - \frac{x^3}{3}\right]_0^{1.5}$$

$$= \frac{3}{4}\left[(1.5)^2 - \frac{(1.5)^3}{3}\right] = \frac{3}{4}[1.125]$$

$$P(X < 1.5) = 0.84375$$

Step II : Probability greater than 1.5

$$\therefore P(X > 1.5) = \int_{1.5}^{2} f_X(x)\,dx = \int_{1.5}^{2} \frac{3}{4}(2x - x^2)\,dx$$

$$= \frac{3}{4}\left(\frac{2x^2}{2} - \frac{x^3}{3}\right)_{1.5}^{2}$$

$$= \frac{3}{4}\left[(2)^2 - \frac{(2)^3}{3}\right] - \frac{3}{4}\left[(1.5)^2 - \frac{(1.5)^3}{3}\right]$$

$$= \frac{3}{4}\left[\left(4 - \frac{8}{3}\right) - (1.5)^2 + \frac{(1.5)^2}{3}\right]$$

$$= \frac{3}{4}\left[4 - \frac{8}{3} - (2.25) + 1.125\right]$$

$$= \frac{3}{4}(0.2083) = \frac{5}{32} = 0.15625$$

Step III : Probability between 0.5 and 1.5

i.e. $$P(0.5 < x < 1.5) = \int_{0.5}^{1.5} f_X(x)\,dx$$

$$= \int_{0.5}^{1.5} \frac{3}{4}(2x - x^2)\,dx = \frac{3}{4}\left[x^2 - \frac{x^3}{3}\right]_{0.5}^{1.5}$$

$$= \frac{3}{4}\left\{\left[(1.5)^2 - \frac{(1.5)^3}{3}\right] - \left[(0.5)^2 - \frac{(0.5)^3}{3}\right]\right\}$$

$$= \frac{3}{4}[2.25 - 1.125 - 0.25 + 0.0417]$$

$$= 0.667$$

Ex. 2.5.7 : If $P(X = x) = \frac{x}{15}$; $x = 1, 2, 3, 4, 5$ then

(i) Find the probability density function $f_X(x)$.

(ii) $P(X = 1 \text{ or } 2)$

(iii) $P\left(\frac{1}{2} < x < \frac{5}{2}\right)$

Soln. :

Step I :

Given $p(X = x) = \frac{x}{15}$ and $x = 1, 2, 3, 4, 5\ldots$

Here, the values of x are discrete, first find the probability distribution function $F_X(x)$

Step II : $Fx(x = 1) = \left(\frac{x}{15}\right)_{x=1} = \frac{1}{15}$

$$F_X(x = 2) = p(X = x = 1) + p(X = x = 2)$$

$$= \left(\frac{x}{15}\right)_{x=1} + \left(\frac{x}{15}\right)_{x=2}$$

$$= \frac{1}{15} + \frac{2}{15} = \frac{3}{15}$$

$$F_X(x=3) = p(X=2) + p(X=3)$$

$$= \frac{3}{15} + \left(\frac{x}{15}\right)_{x=3} = \frac{3}{15} + \frac{3}{15} = \frac{6}{15}$$

$$F_X(x=4) = p(X=3) + p(X=4)$$

$$= \frac{6}{15} + \left(\frac{x}{15}\right)_{x=4} = \frac{6}{15} + \frac{4}{15} = \frac{10}{15}$$

$$Fx(x=5) = p(X=4) + p(X=5)$$

$$= \frac{10}{15} + \left(\frac{x}{15}\right)_{x=5} = \frac{10}{15} + \frac{5}{15} = \frac{15}{15} = 1$$

Probability density function is

X	1	2	3	4	5
$p(X=x) = \frac{x}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{5}{15}$
$F_X(x)$	$\frac{1}{15}$	$\frac{3}{15}$	$\frac{6}{15}$	$\frac{10}{15}$	1

(ii)

Step III

Since, $p(X=1) = \left(\frac{x}{15}\right)_{x=1} = \frac{1}{15}$

And $p(X=2) = \left(\frac{x}{15}\right)_{x=2} = \frac{2}{15}$

Let $A = p(X=1)$ and $B = p(X=2)$

$$p(x=1 \text{ or } 2) = p(A \text{ or } B)$$

$$= p(A \cup B)$$

$$= p(A) + p(B) - p(A \cap B)$$

($\because$ Here, events are independent)

$$= \frac{1}{15} + \frac{2}{15} - 0$$

$$= \frac{3}{15} = \frac{1}{15}$$

(iii)

Step IV :

$$p\left(\frac{1}{2} < X < \frac{5}{2}\right) = p(X=1) + p(X=2)$$

$$= \frac{1}{15} + \frac{2}{15}$$

$$= \frac{3}{15} = \frac{1}{15}$$

($\because$ It is discrete, there is no sample point other then 1 and 2 between $\frac{1}{2} < X < \frac{5}{2}$)

Ex. 2.5.8 : If the probability density function $f_X(x) = \frac{1}{x^2}$ in $0 < x < \infty$ and $A = \{x \mid 2 < x < 3\}$; $B = \{x \mid 6 < x < 7\}$ then find $p(A \cap B)$, $p(A \cup B)$, $p(B \mid A)$

Soln. :

Step I : **The given p.d.f is**

$$f_X(x) = \frac{1}{x^2} \text{ in } 0 < x < \infty$$

$$p(A) = p(2 < x < 3) = \int_2^3 \frac{1}{x^2}\,dx = \left[-\frac{1}{x}\right]_2^3$$

$$= -\left[\frac{1}{3} - \frac{1}{2}\right] = -\left[\frac{2-3}{6}\right] = \frac{1}{6}$$

And

$$p(B) = p(6 < X < 7) = \int_6^7 \frac{1}{x^2}\,dx = \left[-\frac{1}{x}\right]_6^7$$

$$= -\left[\frac{1}{7} - \frac{1}{6}\right] = -\left[\frac{6-7}{42}\right] = \frac{1}{42}$$

Step II

(i) $p(A \cap B) = p(A \text{ and } B)$

$$= p\{2 < x < 3 \text{ and } 6 < x < 7\}$$

$$= 0$$

($\because$ The events are independent. There is no common term between these two interval)

(ii) $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

$$= \frac{1}{6} + \frac{1}{42} - 0$$

$$= \frac{42+6}{252} = \frac{48}{252}$$

(iii) $p(B \mid A) = \frac{p(A \cap B)}{p(B)} = \frac{0}{(1/42)} = 0$

Ex. 2.5.9 : Find the value of K for the probability density function. $f_X(x) = \begin{cases} kx & ; 0 \le x \le 1 \\ k & ; 1 \le x \le 2 \\ k_{(3-x)} & ; 2 \le x \le 3 \end{cases}$

Also, find $p(X \le 1.5)$

Soln. :

Step I : **Given, the probability density function is**

$$fx(x) = \begin{cases} kx & ; 0 \le x \le 1 \\ k & ; 1 \le x \le 2 \\ k(3-x) & ; 2 \le x \le 3 \end{cases}$$

since $\int_a^b f_X(x)\,dx = 1$ ($\because$ Total probability is 1)

$$\therefore \int_0^3 f_X(x)\,dx = 1$$

$$\Rightarrow \int_0^1 f_X(x)\,dx + \int_1^2 f_X(x)\,dx + \int_2^3 fx(x)\,dx = 1$$

$$\Rightarrow \int_0^1 kx\,dx + \int_1^2 k\,dx + \int_2^3 k(3-x)\,dx = 1$$

$$\Rightarrow k\left(\frac{x^2}{2}\right)_0^1 + k(x)_1^2 + k\left(3x - \frac{x^2}{2}\right)_2^3 = 1$$

$$k\left(\frac{1}{2}\right) + k(2-1) + k\left(9 - \frac{9}{2} - 6 + \frac{4}{2}\right) = 1$$

$$\Rightarrow k\left[\frac{1}{2} + 1 + 5 - \frac{9}{2}\right] = 1$$

$$\Rightarrow k(6-4) = 1$$

$$\Rightarrow 2K = 1 \Rightarrow K = \frac{1}{2}$$

Step II : The p.d.f. becomes

$$f_X(x) = \begin{cases} \frac{x}{2} & ; 0 \le x \le 1 \\ \frac{1}{2} & ; 1 \le x \le 2 \\ \frac{1}{2}(3-x) & ; 2 \le x \le 3 \end{cases}$$

$$\therefore P(x \le 1.5) = \int_0^{1.5} f_X(x)\,dx = \int_0^1 f_X(x)\,dx + \int_1^{1.5} f_X(x)\,dx$$

($\because$ $x = 1.5$ lies in the interval $1 \le x \le 2$)

$$= \int_0^1 \frac{1}{2}x\,dx + \int_1^{1.5} \frac{1}{2}\,dx = \frac{1}{2}\left(\frac{x^2}{2}\right)_0^1 + \frac{1}{2}(x)_1^{3/2}$$

$$= \frac{1}{2}\left[\frac{1}{2} - 0 + \frac{3}{2} - 1\right] = \frac{1}{2}[2-1] = \frac{1}{2}(1)$$

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

Ex. 2.5.10 : Find the probability density function $f_X(x)$ and distribution function $F_X(x)$ when a fair coin is tossed four times, where X be the random variable for number of times tail appears.

Soln. :

Step I : Fair coin is tossed four times, the sample space is

S = {HHHH,
HHHT, HHTH, HTHH, THHH,
HHTT, HTTH, HTHT, THHH,
TTHH, THHT,
HTTT, THTT, TTHT, TTTH,
TTTH}

$$p(H) = \frac{1}{2} \text{ and } p(T) = \frac{1}{2}$$

Step II : The probability density function

X	0	1	2	4	5
$P(X=x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Here, X – be the number of times tail appears.

Step III : Probability distribution function $F_X(x)$ is

$$F_X(x=0) = p(X=0) = \frac{1}{16}$$

$$F_X(x=1) = p(X=0) + p(x=1)$$

$$= \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$$

$$F_X(x=2) = p(x=1) + p(X=2)$$

$$= \frac{5}{16} + \frac{6}{16} = \frac{11}{16}$$

$$F_X(x=3) = p(X=2) + p(X=3)$$

$$= \frac{11}{16} + \frac{4}{16} = \frac{15}{16}$$

$$F_X(x=3) = p(X=3) + p(x=4)$$

$$= \frac{15}{16} + \frac{1}{16} = \frac{16}{16} = 1$$

X	0	1	2	3	4
$p(X=x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$
$Fx(x)$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{11}{16}$	$\frac{15}{16}$	1

Syllabus Topic : Mathematical Expectations

2.6 Mathematical Expectation

- As we know, the discrete random variable is completely described by its probability mass function F(X) or probability distribution P(X) and the continuous random variable described by its probability density function f(x).
- Sometimes we need very concise description, for example, a single number, or few numbers, instead of entire function. One such number or value is **Mathematical Expectation.**
- A very important concept in probability theory is the **mathematical expectation** or **expected value** or mean value or **expectation** of a random variable (X).

Definition

Let X – be random variable ;
fx – density function ;
Fx – probability distribution function ;
Pi = $P(X = x_i)$ be probability

- The mean or expected value or expectation of random variable X is denoted by E (X) or μ_x or μ and it is defined as

(i) For X as discrete random variable,

$$E(X) = \mu = \mu_x = x_1 P(x = x_1) + x_2 P(x = x_2) + x_3 P(x = x_3) + \ldots + x_n P(x = x_n)$$

$$= \sum_{i=1}^{n} x_i P(x = x_i)$$

$$\text{or} \quad E(x) = \sum_{i=1}^{n} P_i x_i = \sum_{i=1}^{n} x_i f(x) \qquad \ldots(1)$$

(ii) For X as continuous random variable.

$$E(x) = \mu = \mu_x = \int_{-\infty}^{\infty} x_i f(x)\, dx \qquad \ldots(2)$$

; provided integral is convergent.

If R.H.S. of equation (1) and (2) are not absolutely convergent then E(x) does not exists.

- The expectation of X is often called the mean of X and is denoted by μ_x.
- E(x) gives a single value of X that acts as a representative or average of the values of X.
- Let X be the continuous random variable with probability function F(x). then Y = g(x) also discrete random variable of X, then

$$E(y) = \int_{-\infty}^{\infty} Y f_y(y)\, dy$$

and $$E(Y) = \int_{-\infty}^{\infty} g(x) f_x(x)\, dx$$

If X – is discrete random variable

$$E(Y) = \Sigma g(x_i) P_i$$

Variance

The variance of a random variable X is denoted by σ_x^2 or V(x) and is defined as

$$\sigma_x^2 = V(x) = E\{(x-\mu_x)^2\}$$

(i) $Var(X) = \sigma_x^2 = \sum_i (x_i - \mu_x)^2 P_i$

...if X is discrete random variable

(ii) $Var(X) = \sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)\, dx$

... if x is continuous random variable.

The square root of variance is called standard deviation. The mean of a random variable is its average value and the variance is a measure of the spread or dispersion of the values of random variable.

Syllabus Topic : Joint Probability Distributions

2.7 Joint Probability Distribution

Let E be the experiment and S be the sample space of E. Let X = X(s) and Y = Y(s) be two functions which assigns a real number to every outcome s ∈ S. Then (X, Y) is called two dimensional sample space.

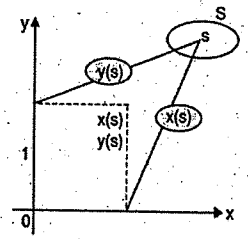


Fig. 2.7.1

s - is member of sample space S, then X(s) corresponding to one real number on X-axis. Similarly, Y(s) corresponds to one real number on Y-axis.

Let A = {X | X ≤ x} and B = {Y | Y ≤ y}

Then the event A takes all points to the left side of vertical line X = x and the event B takes all points to below of horizontal line Y = y.

A ∩ B : maps to the common area of the above two areas.

(A ∩ B) = {(X, Y) | X ≤ x and Y ≤ y}

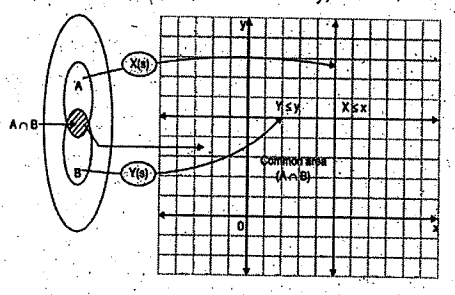


Fig. 2.7.2

As like one-dimensional random variable, in tow dimensional random variables, X(s), and Y(s) consider only values which takes X and Y but not their functional nature.

Definition

A two-dimensional random variable (X, Y) is called two dimensional **discrete random variable**, if the values of (X, Y) are finite or countably infinite.

In two-dimensional discrete random variable (X, Y) represent two-numerical characteristics X and Y.

Definition :

Joint probability distribution function :

The probability of the combine event {X ≤ x, Y ≤ y} which is a function of x and y, is called **joint (or compound) probability distribution function** and it denoted by

$$F_{X,Y}(x, y) = P(X \le x, Y \le y); -\infty \le x \le \infty, -\infty \le y \le \infty$$

Geometrically, the joint probability distribution function $F_{X,Y}(x, y)$ represents the common area between X and Y i.e.

$$F_{X,Y}(x, y) = P(X \le x, Y \le y)$$
$$= \{(X, Y) \mid X \le x, Y \le y\}$$

2.7.1 Properties of Joint Probability Distribution Function :

(i) $\lim_{\substack{x \to -\infty \\ y \to -\infty}} F_{X,Y}(x, y) = 0$

i.e. $F_{X,Y}(-\infty, -\infty) = 0$

(ii) $\lim_{x \to -\infty} F_{X,Y}(x, y) = 0$

i.e. $F_{X,Y}(-\infty, y) = 0$

(iii) $\lim_{y \to -\infty} F_{X,Y}(x, y) = 0$

i.e. $F_{X,Y}(x, -\infty) = 0$

(iv) $\lim_{\substack{x \to \infty \\ y \to \infty}} F_{X,Y}(x, y) = 1$

i.e. $F_{X,Y}(\infty, \infty) = 1$

(v) The values of $F_{X,Y}(x, y)$ lies between 0 and 1.

i.e. $0 \le F_{X,Y}(x, y) \le 1$

(vi) $F_{X,Y}(x, y)$ is increasing function of x and y.

(vii) (a) $p\{x_1 \le X \le x_2, Y \le y\} = F_{X,Y}(x_2, y) - F_{X,Y}(x_1, y)$

(b) $p\{X \le x, y_1 \le Y \le y_2\} = F_{X,Y}(x, y_2) - F_{X,Y}(x, y_1)$

(viii) $p\{x_1 \le X \le x_2; y_1 \le Y \le y_2\} \ge 0 = F_{X,Y}(x_2, y_2) + F_{X,Y}(x_1, y_1) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$

(ix) $F_{X,Y}(x, \infty) = F_x(x)$

(x) $F_{X,Y}(\infty, y) = F_y(y)$

2.7.2 Joint Probability Density Function of A Two-Dimensional Random Variables :

If X, Y are two random variables and $F_{X,Y}(x, y)$ be the joint probability distribution function of two random variables X and Y, then the **joint probability density function of two-dimensional** random variable X and Y is denoted by $f_{X,Y}(x, y)$ and is defined as

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y)$$

... For discrete random variables if partial derivative exists.

Also,

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du\, dv$$

$(\because F_{X,Y}(x, y) = P(X \le x, Y \le y))$

....For continuous random variable X and Y.

2.7.3 Properties of Joint Probability Density Function

1. $f_{X,Y}(x, y) \ge 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$

These above two properties can be used to verify whether the function is joint probability density function or not.

3. $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$
4. $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$
5. $P(x_1 < X \le x_2, y_1 < Y \le y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{X,Y}(x, y)\, dx\, dy$
6. $F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, du\, dv$
7. $F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\, du\, dv$

$$F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\, du\, dv$$

Syllabus Topic : Marginal Probability Distribution

2.8 Marginal Probability Distribution Function

If X and Y are random variables, (X, Y) - is a two-dimensional random variable.

$F_{X,Y}(x, y)$ be the distribution function of two dimensional random variable.

Then the probability that X will take a values less than or equal to x (i.e. X ≤ x) and any value of Y from – ∞ to ∞ (i.e. – ∞ ≤ Y ≤ ∞), such function is called **Marginal or (individual) distribution function** of X, which is denoted by

$F_X(x)$ and is defined as

$$F_X(x) = P(X \le x) = F_{X,Y}(x, \infty)$$

and $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$ is called **Marginal density function.**

The probability of that Y will take a values less than or equal to Y (i.e $Y \le y$) and any value of X from $-\infty$ to ∞ (i.e. $-\infty \le X \le \infty$) such function is called **Marginal distribution function of Y**, which is denoted by

$F_Y(y)$ and is defined as

$$F_Y(y) = P(Y \le y) = F_{X,Y}(\infty, y)$$

and $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$ is called **Marginal density functions**

Note that :

(i) The marginal distribution function $F_X(x)$ is obtained by putting y = upper limit of y.

(ii) The marginal distribution function $F_Y(y)$ is obtained by putting x = upper limit of x.

Ex. 2.8.1 : A two dimensional random variables (x, y) has the joint density function.

$$F_{X,Y}(x, y) = \begin{cases} \frac{4}{7}xy & ; 1 \le y \le 2; 1 \le x \le y \\ 0 & ; \text{Otherwise} \end{cases}$$

Find the marginal distributions.

Soln. :

The given joint density function is

$$f_{X,Y}(x, y) = \begin{cases} \frac{4}{7}xy & ; 1 \le y \le 2; 1 \le x \le y \\ 0 & ; \text{Otherwise} \end{cases}$$

The given curves y = 1 to y = 2 and x = 1 to x = y

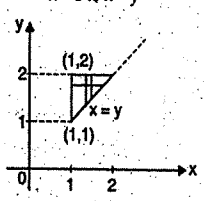


Fig. 2.8.1

To find the limits of y, draw a vertical line (parallel to y-axis) and find the limits of y within the region. The ends of vertical strip are the limits of y.

Limits may be constants or functions of x

Here, limits of y are : y = x to y = 2

Similarly, limits of x are: x = 1 to x = y

Step II :

(i) Marginal density function of X is

$$f_X(x) = \int_{y=x}^{2} f_{X,Y}(x, y)\, dy = \int_{y=x}^{2} \frac{4}{7}xy\, dy$$

Integrating w.r.t. y, keeping x as constant.

$$f_X(x) = \frac{4}{7}x\left(\frac{y^2}{2}\right)_{y=x}^{2} = \frac{2}{7}x(4 - x^2) \text{ in } 1 \le x \le 2$$

(ii) Marginal density function of y is

$$f_Y(y) = \int_{x=1}^{y} f_{X,Y}(x, y)\, dx = \int_{x=1}^{y} \frac{4}{7}xy\, dx$$

Integrating w.r.t. x, keeping y as constant.

$$f_Y(y) = \frac{4}{7}y\left(\frac{x^2}{2}\right)_{1}^{y} = \frac{2}{7}y(y^2 - 1) \text{ in } 1 \le y \le 2$$

Ex. 2.8.2 : Find the distribution function of X and Y. The two dimensional distribution function

$$F_{X,Y}(x, y) = \frac{xy}{16}(x + y); 0 \le x \le 2;\ 0 \le y \le 2$$

Also find the probability density function of (X, Y)

Soln. :

Step I : Given

$$F_{X,Y}(x, y) = \frac{xy}{16}(x + y);\ 0 \le x \le 2;\ 0 \le y \le 2 \quad \ldots(1)$$

The marginal distribution function $F_X(x)$ of the random variable X is obtained by replacing y by its upper limit y = 2.

$$F_X(x) = F_{X,Y}(x, 2) = \frac{x}{8}(x + 2) \quad \ldots(2)$$

Step II :

The marginal distribution function $F_Y(y)$ of the random variable Y is obtained by replacing x by its upper limits x = 2

$$\therefore\ F_Y(y) = F(y) = \frac{y}{8}(y + 2)$$

$$\text{Hence } F_X(x) = \begin{cases} 0 & ; x \le 0 \\ \frac{x}{8}(x + 2) & ; 0 \le x \le 2 \\ 1 & ; x \ge 2 \end{cases}$$

$$\text{and } F_Y(y) = \begin{cases} 0 & ; y \le 0 \\ \frac{y}{8}(y + 2) & ; 0 \le y \le 2 \\ 1 & ; y \ge 2 \end{cases}$$

Step III :

The joint probability density function is given by

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y}\left[\frac{1}{16}xy(x + y)\right]$$

$$= \frac{\partial^2}{\partial x \partial y}\left(\frac{1}{16}(x^2y + xy^2)\right)$$

$$= \frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}\left(\frac{1}{16}(x^2y + xy^2)\right)\right]$$

$$= \frac{\partial}{\partial x}\left[\left(\frac{1}{16}(x^2 + 2xy)\right)\right] = \left(\frac{1}{16}\right)(2x + 2y)$$

$$= \left(\frac{x + y}{8}\right)$$

$$\therefore\ f_{X,Y}(x, y) = \begin{cases} \frac{1}{8}(x + y) & ; 0 \le x \le 2; 0 \le y \le 2 \\ 0 & ; \text{Otherwise} \end{cases}$$

Ex. 2.8.3 : Consider that the lifetime (X) and the brightness (Y) of a light tube are continuous random variables. The joint probability density function

$$f_{X,Y}(x, y) = \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)}; 0 < x < \infty; 0 < y < \infty$$

Then find

(i) Marginal density functions

(ii) Joint probability distribution functions.

Soln. :

Step I : Given

$$f_{X,Y}(x, y) = \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)}$$

in $0 < x < \infty, 0 < y < \infty$

Since, marginal density functions are

(i) $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$$

$$= \int_{-\infty}^{0} f_{X,Y}(x, y)\, dy + \int_{0}^{\infty} f_{X,Y}(x, y)\, dy$$

$$= 0 + \int_{0}^{\infty} \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)}\, dy$$

$$= \lambda_1\lambda_2 e^{-\lambda_1 x}\int_{0}^{\infty} e^{-\lambda_1 y}\, dy$$

$$= \lambda_1\lambda_2 e^{-\lambda_1 x}\left(\frac{e^{-\lambda_1 y}}{-\lambda_1}\right)_{y=0}^{\infty} = -\lambda_2 e^{-\lambda_1 x}(0 - 1)$$

$$F_X(x) = \lambda_2 e^{-\lambda_1 x}$$

(ii) $$F_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

$$= \int_{-\infty}^{0} f_{X,Y}(x, y)\, dx + \int_{0}^{\infty} f_{X,Y}(x, y)\, dx$$

$$= 0 + \int_{0}^{\infty} \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)}\, dx$$

$$= \lambda_1\lambda_2 e^{-\lambda_2 y}\int_{0}^{\infty} e^{-\lambda_1 x}\, dx$$

$$= \lambda_1\lambda_2 e^{-\lambda_2 y}\left(\frac{-e^{-\lambda_1 x}}{-\lambda_1}\right)_{x=0}^{\infty} = -\lambda_2 e^{-\lambda_2 y}(0 - 1)$$

$$f_Y(y) = \lambda_2 e^{-\lambda_2 y}$$

Step II :

Since, for continuous random variables, the joint probability distribution function $F_{X,Y}(x, y)$ is

$$F_{X,Y}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} F(u, v)\, du\, dv$$

$$= \int_{0}^{x}\int_{0}^{y} \lambda_1\lambda_2 e^{-(\lambda_1 u + \lambda_2 v)}\, du\, dv$$

Integrating first w.r.t u, keeping v as constant and then integrating w.r.t. v. (Also we can integrate separately since, the X and Y are independent.)

$$= \int_{0}^{x} \lambda_1\lambda_2 e^{-\lambda_1 u} \cdot du \int_{0}^{y} e^{-\lambda_2 v}\, dv$$

$$= \lambda_1\lambda_2\left(\frac{e^{-\lambda_1 u}}{-\lambda_1}\right)_{u=0}^{x} \cdot \left(\frac{e^{-\lambda_2 v}}{-\lambda_2}\right)_{v=0}^{y}$$

$$= [1 - e^{-\lambda_1 x}] \cdot [1 - e^{-\lambda_2 y}]$$

in $0 < x < \infty; 0 < y < \infty$

Marginal probability distribution of X :

If we find the probability $p_1, p_2, p_3, \ldots p_n$ of $x_1, x_2, x_3, \ldots x_n$ only, then the probability distribution is called **Marginal Probability distribution of X.**

i.e.

X	x_1	x_2	x_3	...	x_n	Sum
P (x)	p_1	p_2	p_3	...	p_n	1

Where

$p_1 = p_{x_1y_1} + p_{x_1y_2} + p_{x_1y_3} + \ldots + p_{x_1y_n}$

$p_2 = p_{x_2y_1} + p_{x_2y_2} + p_{x_2y_3} + \ldots + p_{x_2y_n}$

$\vdots$

$P_n = p_{x_ny_1} + p_{x_ny_2} + p_{x_ny_3} + \ldots + p_{x_ny_n}$

Marginal probability distribution of Y :

If we find the probability $p_1, p_2, p_3, \ldots p_n$ of $y_1, y_2, y_3, \ldots y_n$ only, then the probability distribution is called **marginal probability distribution of Y.**

Y	y_1	y_2	y_3	...	y_n	Sum
P (Y)	p_1	p_2	p_3	...	p_n	1

$p_1 = p_{x_1y_1} + p_{x_2y_1} + p_{x_3y_1} + \ldots + p_{x_ny_1}$

$p_2 = p_{x_1y_2} + p_{x_2y_2} + p_{x_3y_2} + \ldots + p_{x_ny_2}$

$\vdots$

$P_n = p_{x_1y_n} + p_{x_2y_n} + p_{x_3y_n} + \ldots + p_{x_ny_n}$

Note :

The marginal distributions of two dimensional discrete random variable are obtained just adding the probabilities horizontally and vertically.

Ex. 2.8.4 : The joint probability density function of two random variables is given by

$$f_{X,Y}(x,y) = \begin{cases} 12\,e^{-3x-4y} & ; \ x>0\,;x<0 \\ 0 & ; \ \text{Otherwise} \end{cases}$$

Find

(i) The probability that $1<X<2\,;2<Y<3\,;X<2\,;Y>2$

(ii) Find marginal distributions of X and Y.

Soln. :

Step I : Given function is

$$f_{X,Y}(x,y) = \begin{cases} 12\,e^{-3x-4y} & ; \ x>0\,;y>0 \\ 0 & ; \ \text{Otherwise} \end{cases}$$

Since, $P(x_1<X<x_2, y_1<Y<y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)\,dx\,dy$

$$\therefore \ P(1<X<2, 2<Y<3) = \int_{x=1}^{2}\int_{y=2}^{3} 12\,e^{-3x-4y}\,dy\,dx$$

Here, limits of integrations are independent to each other, so we integrate separately,

$$= 12\int_{x=1}^{2} e^{-3x}\,dx \cdot \int_{y=2}^{3} e^{-4y}dy$$

$$= 12\left(\frac{e^{-3x}}{-3}\right)_1^2\left(\frac{e^{-4y}}{-4}\right)_{y=2}^{3}$$

$$= (e^{-6}-e^{-3})(e^{-12}-e^{-8})$$

$$= e^{-6}.e^{-12}-e^{-6}.e^{-8}-e^{-3}.e^{-12}+e^{-3}.e^{-8}$$

$$= e^{-18}-e^{-14}-e^{-15}+e^{-11}$$

Step II :

(i) $P(x<2, y>2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)\,dx\,dy$

$$= \int_{x=0}^{2}\int_{y=2}^{\infty} 12\,e^{-3x-4y}.dy\,dx$$

$$= 12\int_{x=0}^{2} e^{-3x}\,dx \cdot \int_{y=2}^{\infty} e^{-4y}dy$$

$$= 12\left(\frac{e^{-3x}}{-3}\right)_{x=0}^{2}\cdot\left(\frac{e^{-4y}}{-4}\right)_{y=2}^{\infty}$$

$$= (e^{-6}-1)(0-e^{-8})$$

$$= e^{-8}(1-e^{-6})$$

Step III : Marginal probability distribution functions

(i) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$

$$= \int_0^{\infty} 12\,e^{-3x-4y}\,dy$$

($\because$ Here, interval is $x>0$)

$$= 12\left(\frac{e^{-3x-4y}}{-4}\right)_{y=0}^{\infty}$$

$$= -3\,[0-e^{-3x}]$$

$$= \begin{cases} 3\,e^{-3x} & ; \ x>0 \\ 0 & ; \ \text{otherwise} \end{cases}$$

Step IV

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{x=0}^{\infty} 12\,e^{-3x-4y}\,dx$$

$$= 12\left(\frac{e^{-3x-4y}}{-3}\right)_{x=0}^{\infty} = -4\,[0-e^{-4y}]$$

$$= \begin{cases} 4\,e^{-4y} & ; \ y>0 \\ 0 & ; \ \text{otherwise} \end{cases}$$

Ex. 2.8.5 : The joint probability distribution function of X and Y is given by

$$P(X=x, Y=y) = \frac{x+3y}{24}\ ;\ x=1,2\,;y=1,2$$

Find the marginal probability distribution functions.

Soln. :

Step I : Given functions

$$P(X,Y) = \frac{x+3y}{24}\ ;\ x=1,2\,;y=1,2$$

Substitute the values of x = 1, 2 and y = 1, 2 in probability function P (X, Y)

X \ Y	$y_1=1$	$y_2=2$	Total
$x_1=1$	$P_{11}=\frac{1}{6}$	$P_{12}=\frac{7}{24}$	$\frac{11}{24}$
$x_2=2$	$P_{21}=\frac{5}{24}$	$P_{22}=\frac{1}{3}$	$\frac{13}{24}$
Total	$\frac{3}{8}$	$\frac{5}{8}$	1

Step II :

Now, The marginal probability distributions

(i) of X

X	1	2	Total
P (X)	$P_1=\frac{11}{24}$	$P_2=\frac{13}{24}$	1

$$P_1 = P_{11}+P_{12} = \frac{1}{6}+\frac{7}{24} = \frac{11}{24}$$

$$P_2 = P_{21}+P_{22} = \frac{5}{24}+\frac{1}{3} = \frac{13}{24}$$

(ii) of Y

Y	1	2	Total
P (Y)	$P_1=\frac{3}{8}$	$P_2=\frac{5}{8}$	1

$$P_1 = P_{11}+P_{21} = \frac{1}{6}+\frac{5}{24} = \frac{3}{8}$$

and $P_2 = P_{12}+P_{22} = \frac{7}{24}+\frac{1}{3} = \frac{5}{8}$

Ex. 2.8.6 : Find the marginal probability distributions if the joint probability distribution function

$$P(X=x, Y=y) = \frac{1}{27}(x+2y)\,;x=0,1,2$$

and y = 0, 1, 2

Soln. :

Step I : Given joint probability distribution function is

$$P(X,Y) = \frac{1}{27}(x+2y) \qquad ...(1)$$

For x = 0, 1, 2 and y = 0, 1, 2

The joint probability distribution in tabular form is obtained by putting the values of x and y in Equation (1)

X \ Y	$y_1=0$	$y_2=1$	$y_3=2$	Total
$x_1=0$	$P_{11}=0$	$P_{12}=\frac{2}{27}$	$P_{13}=\frac{4}{27}$	$\frac{2}{9}$
$x_2=1$	$P_{21}=\frac{1}{27}$	$P_{22}=\frac{3}{27}$	$P_{23}=\frac{5}{27}$	$\frac{1}{3}$
$x_3=2$	$P_{31}=\frac{2}{27}$	$P_{32}=\frac{4}{27}$	$P_{33}=\frac{6}{27}$	$\frac{4}{9}$
Total	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{5}{9}$	1

Step II : The marginal probability distribution

	X	P (X)
(i) of X	0	$\frac{2}{9}$
	1	$\frac{1}{3}$
	2	$\frac{4}{9}$
	Total	1

(i) of Y

Y	P (Y)
0	$\frac{1}{9}$
1	$\frac{1}{3}$
2	$\frac{5}{9}$
Total	1

Ex. 2.8.7 : Suppose X and Y represent the operating lives of bulbs (X) and Tubes (Y) in a months. Their joint probability density function is given by

$$f_{X,Y}(x,y) = \begin{cases} e^{-x-y} & ; \ x\ge 0, y\ge 0 \\ 0 & ; \ \text{otherwise} \end{cases}$$

Find :

(i) The probability of bulbs has life less than or equal to 1 month.

(ii) The probability of tubes has life more than 2 months.

Soln. :

Step I : Given joint p.d.f is

$$f_{X,Y}(x,y) = \begin{cases} e^{-x-y} & ; \ x\ge 0, y\ge 0 \\ 0 & ; \ \text{otherwise} \end{cases}$$

Step II : Marginal probability density function

(i) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$

$$= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy + \int_0^{\infty} f_{X,Y}(x,y)\,dy$$

$$= \int_0^{\infty} e^{-x-y}\,dy = e^{-x}\int_0^{\infty} e^{-y}\,dy$$

$$= e^{-x}\left(-e^{-y}\right)_0^{\infty} = -e^{-x}(0-1) = e^{-x}$$

$\therefore$ **Probability of bulbs life less than 1 months**

$$= \int_0^1 f_X(x)\,dx = \int_0^1 e^{-x}\,dx = \left(-e^{-x}\right)_0^1$$

$$= -(e^{-1}-1)$$

$$= 1-e^{-1} = 1-0.37 = 0.63$$

Step II :

(ii) $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$

$$= \int_0^{\infty} e^{-x-y}\,dx = e^{-y}\int_0^{\infty} e^{-x}\,dx$$

$$= e^{-y}\left(\frac{-e^{-x}}{-1}\right)_0^{\infty} = e^{-y}(0-1) = e^{-y}$$

$\therefore$ Probability of life greater than 2 months

$$= \int_2^{\infty} f_Y(y)\,dy = \int_2^{\infty} e^{-y}\,dy = \left[-e^{-y}\right]_2^{\infty}$$

$$= \left[0+e^{-2}\right] = e^{-2} = 0.13$$

Ex. 2.8.8 : The joint probability distribution of a bivariate random variate (X,Y) is

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.3	0.1

Obtain marginal distribution of X and Y

Soln. :

Step I : Given

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.3	0.1

The marginal probability distributions

(i) of X

X	P (X)	
1	0.3	$\Rightarrow$ $0.1 + 0.2 = 0.3$
2	0.4	$\Rightarrow$ $0.1 + 0.3 = 0.4$
3	0.3	$\Rightarrow$ $0.2 + 0.1 = 0.3$
Total	1	

(ii) of Y

X	P (X)	
1	0.4	$\Rightarrow$ $0.1 + 0.1 + 0.2 = 0.4$
2	0.6	$\Rightarrow$ $0.2 + 0.3 + 0.1 = 0.6$
Total	1	

Ex. 2.8.9 : For the following probability distribution

Y \ X	−1	0	1
0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
1	$\frac{1}{5}$	$\frac{2}{15}$	$\frac{1}{15}$
2	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{2}{15}$

Obtain marginal distribution of X and Y

Soln. :

Step I : From the given joint probability distribution, we can find the marginal probability density distributions of

(i) of X

X	P (X)
−1	$P_1 = \frac{6}{15}$
0	$P_2 = \frac{5}{15}$
1	$P_3 = \frac{4}{15}$
Total	1

$$P_1 = \frac{1}{15} + \frac{1}{5} + \frac{2}{15} = \frac{6}{15}$$

$$P_2 = \frac{2}{15} + \frac{2}{15} + \frac{1}{15} = \frac{5}{15}$$

$$P_3 = \frac{1}{15} + \frac{1}{15} + \frac{2}{15} = \frac{4}{15}$$

(i) of Y

Y	P (X)
0	$P_1 = \frac{4}{15}$
1	$P_2 = \frac{6}{15}$
2	$P_3 = \frac{5}{15}$
Total	1

$$P_1 = \frac{1}{15} + \frac{2}{15} + \frac{1}{15} = \frac{4}{15}$$

$$P_2 = \frac{1}{5} + \frac{2}{15} + \frac{1}{15} = \frac{6}{15}$$

$$P_3 = \frac{2}{15} + \frac{1}{15} + \frac{2}{15} = \frac{5}{15}$$

Ex. 2.8.10 : The joint distribution of (X, Y) is given by

$$P(X, Y) = \frac{x + y}{21} \; ; x = 1, 2, 3 \; ; y = 1, 2$$

Find the marginal probability functions of X and Y

Soln. :

Step I : Given

$$P(X, Y) = \frac{x + y}{21} \; ; x = 1, 2, 3 \; ; y = 1, 2$$

The joint probability distribution in tabular form is obtained by putting the values of x and y in P (X, Y)

Y \ X	1	2	3	Total
1	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{9}{21}$
2	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{12}{21}$
Total	$\frac{5}{21}$	$\frac{7}{21}$	$\frac{9}{21}$	1

Step II : The marginal probability distribution

(i) of X

X	P (X)	
1	$\frac{5}{21}$	$\Rightarrow \frac{2}{21} + \frac{3}{21} = \frac{5}{21}$
2	$\frac{7}{21}$	$\Rightarrow \frac{3}{21} + \frac{4}{21} = \frac{7}{21}$
3	$\frac{9}{21}$	$\Rightarrow \frac{4}{21} + \frac{5}{21} = \frac{9}{21}$
Total	1	

(i) of Y

Y	P (X)	
1	$\frac{9}{21}$	$\Rightarrow \frac{2}{21} + \frac{3}{21} + \frac{4}{21} = \frac{9}{21}$
2	$\frac{12}{21}$	$\Rightarrow \frac{3}{21} + \frac{4}{21} + \frac{5}{21} = \frac{12}{21}$
Total	1	

Ex. 2.8.11 : Two dimensional random variable (X, Y) has the joint density function

$$f_{X,Y}(x, y) = \begin{cases} \frac{x^3 y^3}{16} & ; \; 0 \le x \le 2 \; ; 0 \le y \le 2 \\ 0 & ; \; \text{otherwise} \end{cases}$$

Find marginal density functions.

Soln. :

Step I : The given joint p.d.f. is

$$f_{X,Y}(x, y) = \begin{cases} \frac{x^3 y^3}{16} & ; \; 0 \le x \le 2 \; ; 0 \le y \le 2 \\ 0 & ; \; \text{otherwise} \end{cases}$$

The marginal density distribution of X and Y is

(i) $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_{y=0}^{2} \frac{x^3 y^3}{16}\, dy$$

$$= \frac{x^3}{16} \left(\int_{y=0}^{2} y^3\, dy \right)$$

$$= \frac{x^3}{16} \left(\frac{y^4}{4} \right)_{y=0}^{2} = \frac{x^3}{16} \left(\frac{2^4}{4} \right)$$

$$f_X(x) = \frac{x^3}{4} \; ; \; 0 < x < 2$$

Step II :

(ii) $$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx = \int_{0}^{2} \frac{x^3 y^3}{16}\, dx$$

$$= \frac{y^3}{16} \left(\frac{x^4}{4} \right)_0^2 = \frac{y^3}{16} \left(\frac{2^4}{4} \right)$$

$$= \frac{y^3}{4} \; ; \; 0 < y < 2$$

Exercise

Ex. 1 : The joint probability distribution of (X, Y) is given by

$$P(X = x, Y = y) = \frac{2x + 3y}{72} \; ; \; x = 0, 1, 2 \; ; \; y = 1, 2, 3$$

Find the marginal probability distribution of X and Y.

Ex. 2 : The joint probability distribution of discrete random variables X and Y is given below :

Y \ X	1	2	3
1	$\frac{1}{12}$	$\frac{1}{6}$	0
2	0	$\frac{1}{6}$	$\frac{1}{4}$
3	$\frac{1}{8}$	0	$\frac{5}{24}$

Find the marginal distributions of X and Y.

Ans. :

X	1	2	3
P (X)	$\frac{5}{24}$	$\frac{1}{3}$	$\frac{11}{24}$

;

Y	1	2	3
P (Y)	$\frac{1}{4}$	$\frac{5}{12}$	$\frac{1}{3}$

Ex. 3 : A two dimensional random variable (X, Y) has the joint density function

$$f_{X,Y}(x, y) = x + y : 0 \le (x, y) \le 1$$

Find the marginal distribution probability distributions

Ans. :

$$f_X(x) = x + \frac{1}{2} \; ; \; 0 \le x \le 1$$

$$f_Y(y) = \frac{1}{2} + y \; ; \; 0 \le y \le 1$$

Ex. 4 : A two dimensional random variable (X, Y) has the joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & ; \; 0 \le x \le 1, \; 0 \le y \le 1 \\ 0 & ; \; \text{otherwise} \end{cases}$$

Find :

(i) $P(0.2 \le X \le 0.5, 0.4 \le Y \le 0.6)$

Ans. : 0.1868

(ii) Marginal probability density function

Ans. :

$$f_X(x) = \begin{cases} \frac{6}{5}\left(x + \frac{1}{3}\right) & ; \; 0 \le x \le 1 \\ 0 & ; \; \text{otherwise} \end{cases}$$

and $$f_Y(y) = \begin{cases} \frac{6}{5}\left(\frac{1}{2} + y^2\right) & ; \; 0 < y < 1 \\ 0 & ; \; \text{otherwise} \end{cases}$$

Syllabus Topic : Properties of Expectations and Variance with Proof

2.9 Properties of Expectation and Variance

(i) If a is any constant, then

$$E(ax) = a\,E(x)$$

(ii) If X and Y are any random variables then

$$E(X+Y) = E(x) + E(y)$$

(iii) If X and Y are independent variables then

$$E(XY) = E(X)\cdot E(Y)$$

(iv) $E(ax+b) = a\,E(x) + b$; a, b are any constants

Proofs

(i) since, $E(x) = \sum_i x_i P_i$

$$\therefore\ E(ax) = \sum_i a\,x_i P_i = a\,\Sigma x_i P_i$$

$$= a\,E(x)$$

(ii) By definition

$$E(X+Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}(x_i + y_j)$$

$$= \sum_i\sum_j P_{ij}x_i + \sum_i\sum_j P_{ij}y_j$$

$$= \sum_i x_i\left(\sum_j P_{ij}\right) + \sum_j y_j\left(\sum_i P_{ij}\right)$$

$$= \sum_i x_i P_i + \sum_j y_j P_i$$

$$E(X+Y) = E(X) + E(Y)$$

(iii) Since

$$E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} P_i P_j' x_i x_j$$

$$= \left(\sum_i P_i x_i\right)\left(\sum_{j=1}^{n} P_j' y_j\right)$$

$$= E(x)\cdot E(Y)$$

(iv) $E(ax+b) = \Sigma P_i(a\,x_i + b)$

$$= a\,\Sigma P_i x_i + b$$

$$= a\,\Sigma x + b$$

Properties of variance

(v) $Var(x) = E(x^2) - [E(x)]^2$

Proof :

Since $Var(x) = E[(X-\mu_x)^2]$

$$= E\left[X^2 - 2\mu_x X + \mu_x^2\right]$$

$$= E(X^2) - 2\mu_x E(X) + \mu_x^2$$

($\because$ μ_x is mean = constant)

$$= E(X^2) - 2\mu_x\cdot\mu_x + \mu_x^2$$

($\because$ $E(x) = \mu_x$)

$$= E(X^2) - \mu_x^2$$

$$= E(X^2) - [E(X)]^2$$

(vi) $Var(ax) = a^2\,Var(x)$

Proof :

since, $Var(ax) = E(X^2) - [E(X)]^2$

$$\therefore\ V_{ar}(ax) = E(a^2X^2) - [E(aX)]^2$$

$$= a^2E(X^2) - [a\,E(X)]^2 = a^2[E(X^2) - [E(X)]^2]$$

$$= a^2\,Var(X)$$

(vii) Variance of a constant is zero ; $Var(c) = 0$

Proof :

$$Var(c) = E(c^2) - [E(c)]^2$$

$$= c^2 - [c]^2$$

$$= 0$$

(viii) $Var(ax+b) = a^2V(x)$

Also note that :

(i) $E[(x-a)^2]$ is a maximum ; when $a = \mu = E(X)$

(ii) $Var(X+Y) = Var(X) + Var(Y)$

(iii) $Var(X-Y) = Var(X) + Var(Y)$

(iv) Since, σ_x^2 is always positive

$$\therefore\ E(x^2) \geq [E(x)]^2$$

Ex. 2.9.1 : Find the expected value E(x) for the probability function for X.

X	0	20	0	40	0	−30
$P(x_i = x)$	1/6	1/6	1/6	1/6	1/6	1/6

Soln. :

$$E(x) = \Sigma x_i P_i$$

$$= (0)\left(\frac{1}{6}\right) + (20)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + 40\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (-30)\left(\frac{1}{6}\right)$$

$$= 5$$

Ex. 2.9.2 : If the density function of X is given

$f_x(x) =$

Then find expected value.

Soln. :

Step 1 : Given $f_x(x) = \begin{cases} \frac{1}{2}x; & 0 < x \leq 2 \\ 0, & \text{otherwise} \end{cases}$

$$E(x) = \int_{-\infty}^{\infty} x_i f(x)\,dx = \int_0^2 x\left(\frac{1}{2}x\right)dx$$

$$= \frac{1}{2}\int_{x=0}^{2} x^2\,dx$$

$$= \frac{1}{2}\left(\frac{x^3}{3}\right)_0^2 = \frac{4}{3}$$

Ex. 2.9.3 : The probability density function

$$f_x(x) = \frac{x}{a^2}e^{-\frac{x^2}{2a^2}};\ x > 0$$

Find (i) $E(x^n)$; (ii) Mean;

(iii) Variance of x.

Soln. :

Step I : Given $f_x(x) = \frac{x}{a^2}e^{-\frac{x^2}{2a^2}}$

Since, $E(x) = \int_{-\infty}^{\infty} x_i f_x(x)\,dx$

$$\Rightarrow E(x^n) = \int_{-\infty}^{\infty} x^n\cdot f_x(x)\,dx$$

$$= \int_0^{\infty} x^n\frac{x}{a^2}\cdot e^{-\frac{x^2}{2a^2}}\,dx$$

put $\frac{x^2}{2a^2} = t \Rightarrow \frac{2x\,dx}{2a^2} = dt$

$$\Rightarrow \frac{x}{a^2}\,dx = dt$$

$$\therefore\ E(x^n) = \int_0^{\infty}(2a^2t)^{n/2}\,e^{-t}\,dt$$

$$= 2^{\frac{n}{2}}\,a^n\int_0^{\infty} e^{-t}\,t^{\frac{n}{2}}\,dt$$

$$= 2^{\frac{n}{2}}\,a^n\int_0^{\infty} e^{-t}\,t^{\left(\frac{n}{2}+1\right)-1}\,dt$$

By gamma function

$$E(x^n) = 2^{\frac{n}{2}}\cdot a^n\,\overline{\left|\frac{n}{2}+1\right.} \quad \ldots(1)$$

Step II :

Put $n = 1$ we get mean, from (1)

$$E(x) = 2^{\frac{1}{2}}\cdot a\,\overline{\left|\frac{1}{2}+1\right.} = \sqrt{2}\,a\cdot\overline{\left|\frac{1}{2}\right.}\cdot\frac{1}{2}$$

$$= \frac{\sqrt{2}\,a}{2}\left(\sqrt{\pi}\right) = \frac{a}{\sqrt{2}}\left(\sqrt{\pi}\right)$$

Step III :

Put n = 2 in Equation (1)

$$E(x^2) = 2a^2\,\overline{|2} = 2a^2$$

($\because \overline{|2} = 1!$)

$$\therefore\ V(x) = E(x^2) - [E(x)]^2$$

$$= 2a^2 - \left(\frac{a}{\sqrt{2}}\sqrt{\pi}\right)^2 = 2a^2 - \frac{a^2\pi}{2}$$

$$= \left(2 - \frac{\pi}{2}\right)a^2$$

Ex. 2.9.4 : Find mean and variance of

(i) $f(x) = 6x^2(1-x^3)$; $0 \leq x \leq 1$

Soln. :

Step I : Given density function

$$f_x(x) = 6x^2(1-x^3);\ 0 \leq x \leq 1$$

Since, mean $E(x) = \int_{-\infty}^{\infty} x\,f_x(x)\,dx$

$$\therefore\ E(x) = \int_{-\infty}^{0} x\,f_x(x)\,dx + \int_0^1 x\,f_x(x)\,dx + \int_1^{\infty} x\,f_x(x)\,dx$$

$$= 0 + \int_0^1 x\cdot 6x^2(1-x^3)\,dx + 0$$

$$= \int_0^1 (6x^3 - 6x^6)\,dx = 6\left[\frac{x^4}{4} - \frac{x^7}{7}\right]_0^1$$

$$= 6\left[\frac{1}{4} - \frac{1}{7}\right] = 6\left(\frac{7-4}{28}\right) = \frac{18}{28} = \frac{9}{14}$$

Step II :

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f_x(x)\,dx = \int_0^1 x^2 f_x(x)\,dx$$

$$= \int_0^1 x^2\cdot 6x^2(1-x^3)\,dx = 6\int_0^1 (x^4 - x^7)\,dx$$

$$E(x^2) = 6\left[\frac{x^5}{5} - \frac{x^8}{8}\right]_0^1 = 6\left[\frac{1}{5} - \frac{1}{8}\right] - [0]$$

$$= 6\left[\frac{8-5}{40}\right] = \frac{18}{40} = \frac{9}{20}$$

Step III :

$$\text{Variance } (X) = E(x^2) - [E(x)]^2$$
$$= \left(\frac{9}{20}\right) - \left(\frac{9}{14}\right)^2 = \left(\frac{9}{20}\right) - \left(\frac{81}{196}\right)$$
$$= \frac{1764 - 1620}{3920} = \frac{144}{3920}$$
$$= \frac{9}{245}$$

Ex. 2.9.5 : Find the mean afid variance of the function

x:	1	3	4	5
$f_x(X = x)$:	0.4	0.1	0.2	0.3

Soln. :

Step I : Given distribution

x:	$x_1 = 1$	$x_2 = 3$	$x_3 = 4$	$x_4 = 5$
$f_x(x) = P(x)$	$P_1 = 0.4$	$P_2 = 0.1$	$P_3 = 0.2$	$P_4 = 0.3$

(i) $$\text{Mean } E(x) = \sum P_i x_i$$
$$= x_1 P_1 + x_2 P_2 + x_3 P_3 + x_4 P_4$$
$$= 1(0.4) + 3(0.1) + 4(0.2) + 5(0.3)$$
$$= 0.4 + 0.3 + 0.8 + 1.5 = 3.0$$

(ii) $$E(x^2) = \sum x_i^2 P_i$$
$$= x_1^2 P_1 + x_2^2 P_2 + x_3^2 P_3 + x_4^2 P_4$$
$$= (1)^2 (0.4) + (3)^2 (0.1) + (4)^2 (0.2) + (5)^2 (0.3)$$
$$= 0.4 + 0.9 + 3.2 + 7.5$$
$$= 12.0$$

Step II :

$$\text{Variance Var } (x) = E(x^2) - \{E(x)\}^2$$
$$= 12 - (3)^2 = 12 - 9$$
$$= 3$$

Ex. 2.9.6 : The probability distribution of a discrete random variable is

X:	0	1	2	3
P (X = x) :	0.1	0.3	0.5	0.1

If $Y = x^2 + 2x$, then find mean and variance of Y.

Soln. :

Step I : Given

X:	0	1	2	3
P (x = x) :	0.1	0.3	0.5	0.1

Given $Y = x^2 + 2x$

Put $x = 0, 1, 2, 3$ in above equation

Y:	0	3	8	15
P (Y = y)	0.1	0.3	0.5	0.1

Step II :

$$\text{Mean } E(Y) = \sum y_i P_i$$
$$= y_1 P_1 + y_2 P_2 + y_3 P_3 + y_4 P_4$$
$$= (0)(0.1) + (3)(0.3) + (8)(0.5) + (15)(0.1)$$
$$= 0 + 0.9 + 4 + 1.5 = 6.4$$

(ii) $$E(Y^2) = \sum y_i^2 P_i$$
$$= y_1^2 P_1 + y_2^2 P_2 + y_3^2 P_3 + y_4^2 P_4$$
$$= (0)^2(0.1) + (3)^2(0.3) + (8)^2(0.5) + (15)^2(0.1)$$
$$= 0 + 2.7 + 32 + 22.5 = 57.2$$

Step III :

$$\text{Var } (Y) = E(y^2) - \{E(y)\}^2$$
$$= (57.2) - (6.4)^2 = 16.24$$

Ex. 2..9.7 : If the random variable X follows N (0,2) and $Y = 3x^2$, find the mean and variance of y.

Soln. :

Step I : Given X follows $N(0, 2) = N(\bar{x}, \sigma)$

$$\therefore E(x) = \bar{x} = 0 \text{ and } \sigma = 2$$
$$\text{Var } (x) = \sigma^2 = 4$$

Since $$\text{Var } (x) = E(x^2) - \{E(x)\}^2$$
$$\therefore E(x^2) = \text{Var } (x) + \{E(x)\}^2$$
$$= 4 + (0)^2 = 4$$

Step II :

$$E(Y) = E(3x^2)$$
$$= 3(E(x^2)) = 3 \times 4 = 12$$
$$E(Y^2) = E(3x^2)^2$$
$$= E(9x^4) = 9E(x^4)$$
$$= 9 \times 3 \times 2^4$$
$$= 27 \times 2^4$$

($\because$ Normal distribution $N \le \sigma$)

$$E(x^{2r}) = \frac{(2r)!\,\sigma^{2r}}{2^r\, r!}$$

Hence, $$\text{Var } (Y) = E(Y^2) - \{E(Y)\}^2$$
$$= 27 \times 2^4 - (12)^2$$
$$= 288$$

Exercise

Ex. 1 : A random variable X has the following probability mass function.

$$P(X = x) = \frac{x}{15}, \quad x = 1, 2, 3, 4, 5, 6$$

Find $E(x)^2$; $E(2x^2 + 1)$; $E(3x + 8)$

Ex. 2 : If a random variable X has a probability density function

$f_x(x) = 0.5 \cos x$, find the mean and variance.

Ex. 3 : If a continuous random variable X has probability density function.

$$f_x(x) = \begin{cases} x2 & ; 0 \le x \le 1 \\ 3 - x & ; 1 \le x \le 2 \\ 0 & ; \text{0theswise} \end{cases}$$

Find E (Y), where, $Y = (x - 1)2$

$$\left(\because E(Y) = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx\right)$$

Ex. 4 : Find mean and variance of

(i) $f_x(x) = \frac{3}{4} x (2 - x) ; 0 \le x \le 2$

(ii) $f_x(x) = \frac{1}{\pi} \cdot \frac{2}{1 + x2} ; -\infty < x < \infty$

(iii) $f_x(x) = \begin{cases} 2x2 & ; 0 \le x \le 1 \\ (2 - x2) & ; 1 \le x \le 2 \end{cases}$

Multiple Choice Questions

Q. 1 If f(x) is probability density function and G(x) is called probability distribution function then, G(X) is defined as

(a) $\int_{-\infty}^{\infty} f(x)\, dx$ (b) $\int_{-\infty}^{x} f(x)\, dx$

(c) $\int_{0}^{\infty} f(x)\, dx$ (d) $\int_{x}^{\infty} f(x)\, dx$

Ans. : (b)

Explanation : If f(x) is the probability density function then F(x) is called probability distribution function and is defined as

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\, dx$$

Q. 2 If $F(x) = \int_{-\infty}^{x} f(x)\, dx$ be the continuous probability distribution function then $F(\infty)$ is equal to

(a) 0 (b) −1 (c) 1 (d) 10

Ans. : (c)

Explanation : Total area under density curve from $(-\infty < x < \infty)$ is unity.

i.e. $$\int_{-\infty}^{\infty} f_x(x)\, dx = 1$$

Q. 3 If $G(x) = \begin{cases} e^{-x} & ; x \ge 0 \\ 0 & ; x < 0 \end{cases}$ is the probability density function of variate x in (1, 2), then the probability function G(2) is

(a) $1 - e^{-2}$ (b) $1 + e^{-2}$

(c) $1 - e^2$ (d) $1 + e^2$

Ans. : (a)

Explanation : Probability distribution function

Since, $$F(x) = \int_{-\infty}^{x} f(x)\, dx$$

$$\therefore F(2) = \int_{-\infty}^{2} f(x)\, dx = \int_{-\infty}^{0} f(x)\, dx + \int_{0}^{2} f(x)\, dx$$
$$= \int_{-\infty}^{0} 0\, dx + \int_{0}^{2} e^{-x}\, dx = (-e^{-x})_0^2$$
$$= -(e^{-2} - 1) = 1 - e^{-2}$$

Q. 4 If the probability density function of a variate X is :

X	0	1	2	3	4	5	6
P(X)	k	3k	5k	7k	9k	11k	13k

then the value of k is

(a) 1 (b) $\frac{1}{49}$ (c) 49 (d) $\frac{1}{2}$

Ans. : (b)

Explanation : Let X-be the random variable.

Since, $\sum p(x_i) = 1$

$$\therefore k + 3k + 5k + 7k + 9k + 11k + 13k = 1$$
$$\Rightarrow 49k = 1$$
$$\therefore k = \frac{1}{49}$$

Q. 5 If the probability density function of a variate X is :

X	0	1	2	3	4	5	6
P(X)	k	3k	5k	7k	9k	11k	13k

Q. 6 with $k = \frac{1}{49}$ then $P(X \le 4)$ is

(a) 1 (b) $\frac{4}{49}$ (c) $\frac{25}{49}$ (d) −1

Ans. ; (c)

Explanation : Let X-be the random variable.

Since, $\sum p(x_i) = 1$

$$p(X \le 4) = k + 3k + 5k + 7k + 9k = 25k$$
$$= 25 \times \left(\frac{1}{49}\right) = \frac{25}{49}$$

Q. 7 If the probability density function of a variate X is :

X	0	1	2	3	4	5	6
P(X)	k	3k	5k	7k	9k	11k	13k

and $P(x \le 3) > 4$ then the value of k is

(a) $k = 1$ (b) $k = 4$ (c) $k > \frac{1}{4}$ (d) $k < \frac{1}{4}$

Ans. : (c)

Explanation : $p(X \le 3) = k + 3k + 5k + 7k = 16k$

But given $p(X \le 3) > 4$

$\therefore 16k > 4 \Rightarrow k > \frac{4}{16}$

Q. 8 If μ be the mean value of the probability distribution of a variate X then its Expectation E(X) for continuous distribution is defined as

(a) $\int_{-\infty}^{x} f(x)\,dx$ (b) $\int_{-\infty}^{\infty} x\,f(x)\,dx$

(c) $\int_{-\infty}^{x} x\,f(x)\,dx$ (d) $\int_{0}^{\infty} x\,f(x)\,dx$

Ans. : (b)

Explanation : Since, μ is the variance of a random variable X.

The mean value (μ) of the probability distribution of a variate X is known as its expectation and it is denoted by E(X) and is defined as,

$$E(X) = \sum x_i f(x_i) \text{ – for discrete distribution}$$

$$= \int_{-\infty}^{\infty} x\,f(x_i)dx \text{ – for continuous distribution}$$

Where $f(x_i)$ - be the probability density function of X.

Q. 9 The continuous random variable X has the probability distribution

X	−3	6	9
P(x = X)	1/6	½	1/3

the Expectation E(X) for discrete distribution is

(a) 12 (b) −162 (c) $\frac{11}{2}$ (d) $\frac{-1}{2}$

Ans. : (c)

Explanation :

Since, $E(x) = \sum x\,f(x) = -3 \times \frac{1}{6} + 6 \times \frac{1}{2} + 9 \times \frac{1}{3}$

$= -\frac{1}{2} + 3 + 3 = \frac{11}{2}$

Q. 10 The continuous random variable X has the probability distribution

X	−3	6	9
P(x = X)	1/6	½	1/3

then the expectation $E(x^2)$ is equal to

(a) $\frac{11}{2}$ (b) 126 (c) $\frac{93}{2}$ (d) 0

Ans. : (c)

Explanation :

$$E(x^2) = \sum x^2 f(x) = 9 \times \frac{1}{6} + 36 \times \frac{1}{2} + 81 \times \frac{1}{3}$$

$$= \frac{2}{3} + 18 + 27 = \frac{93}{2}$$

Q. 11 For continuous random variable X, the probability density function $f_X(x) = kx^2;\ 0 \le x \le 2$ then the value of k is .

(a) $\frac{3}{8}$ (b) $\frac{8}{3}$ (c) 0 (d) 1

Ans. : (a)

Explanation : Given, $f_X(x) = k\,x^2$ for $0 \le x \le 2$

Since, we know, total probability is unity.

$$\int_{a}^{b} f_X(x)\,dx = 1$$

$$\therefore \int_{a=0}^{b=2} kx^2\,dx = 1$$

$$\Rightarrow K\left(\frac{x^3}{3}\right)_0^2 = 1 \Rightarrow K\left(\frac{2^3}{3} - 0\right) = 1$$

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

Q. 12 For the probability density function

$$f_X(x) = \begin{cases} 2e^{-2x} & ;\ x \ge 0 \\ 0 & ;\ \text{for } x < 0 \end{cases}$$

the value of $P(x \le 1.5)$ is ,

(a) 0.65021 (b) 0.75020

(c) 0.85027 (d) 0.95021

Explanation :

$$P(x \le 1.5) = \int_0^{1.5} f_X(x)\,dx = \int_0^{1.5} 2e^{-2x}\,dx$$

$$= 2\left[\frac{e^{-2x}}{-2}\right]_0^{1.5} = -(e^{-2x})_0^{1.5}$$

$$= -[e^{-2(1.5)} - 1]$$

$$P(x \le 1.5) = [1 - e^{-3}]$$

$$= 0.95021$$

Q. 13 If X, Y are two continuous random variables and $F_{X,Y}(x, y)$ be the joint probability distribution function of two random variables X and Y, then the joint probability density function of two-dimensional random variable X and Y is denoted by $f_{X,Y}(x, y)$ and is defined as

(a) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$

(b) $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x, y)$

(c) $F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\,du\,dv$

(d) $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$

Ans. : (c)

Explanation : If X, Y are two random variables and $F_{X,Y}(x, y)$ be the joint probability distribution function of two random variables X and Y, then the joint probability density function of two-dimensional random variable X and Y is denoted by $f_{X,Y}(x, y)$ and is defined as

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\,du\,dv$$

$$(\because F_{X,Y}(x, y) = P(X \le x,\ Y \le y)$$

....For continuous random variable X and Y.

Q. 14 If X, Y are two discrete random variables and $F_{X,Y}(x, y)$ be the joint probability distribution function of two random variables X and Y, then the joint probability density function of two-dimensional random variable X and Y is denoted by $f_{X,Y}(x, y)$ and is defined as

(a) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$

(b) $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x, y)$

(c) $F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\,du\,dv$

(d) $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$

Ans. : (b)

Explanation : If X, Y are two random variables and $F_{X,Y}(x, y)$ be the joint probability distribution function of two random variables X and Y, then the joint probability density function of two-dimensional random variable X and Y is denoted by $f_{X,Y}(x, y)$ and is defined as

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x, y)$$

... For discrete random variables if partial derivative exists.

Q. 15 Marginal or (individual) density function of continuous variable X is defined as

(a) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$

(b) $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x, y)$

(c) $F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\,du\,dv$

(d) $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$

Ans. : (a)

Explanation : $F_{X,Y}(x, y)$ be the distribution function of two dimensional random variable.

The probability that X will take a values less than or equal to x (i.e. $X \le x$) and any value of Y from $-\infty$ to ∞

(i.e. $-\infty \le Y \le \infty$), such function is called Marginal density function of X, which is defined as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$$

Q. 16 Marginal or (individual) density function of continuous variable Y is defined as

(a) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$

(b) $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x, y)$

(c) $F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\,du\,dv$

(d) $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$

Ans. : (d)

Explanation :

The probability of Y will take a values less than or equal to y (i.e $Y \le y$) and any value of X from $-\infty$ to ∞ i.e. $-\infty \le X \le \infty$) such function is called marginal density functions of Y, which is defined as

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$$

Q. 17 The joint probability distribution of a bivariate random variate (X,Y) is

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.3	0.1

Then the marginal distribution of X is

(a)

X	P (X)
1	0.3
2	0.4
3	0.3
Total	1

(b)

X	P (Y)
1	0.4
2	0.6
Total	1

(c)

X	P (X)
1	0.0
2	0.6
3	0.4
Total	1

(d)

X	P (Y)
1	0.5
2	0.5
Total	1

Ans. : (a)

Explanation : Given

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.3	0.1

The marginal probability distributions of X

X	P (X)	
1	0.3	$\Rightarrow$ $0.1 + 0.2 = 0.3$
2	0.4	$\Rightarrow$ $0.1 + 0.3 = 0.4$
3	0.3	$\Rightarrow$ $0.2 + 0.1 = 0.3$
Total	1	

Q. 18 The joint probability distribution of a bivariate random variate (X,Y) is

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.3	0.1

Then the marginal distribution of Y is

(a)

X	P (X)
1	0.3
2	0.4
3	0.3
Total	1

(b)

X	xP (Y)
1	0.4
2	0.6
Total	1

(c)

X	P (X)
1	0.0
2	0.6
3	0.4
Total	1

(d)

X	P (Y)
1	0.5
2	0.5
Total	1

Ans. : (b)

Explanation : Given

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.3	0.1

The marginal probability distributions of Y

X	P (Y)	
1	0.4	$\Rightarrow$ $0.1 + 0.1 + 0.2 = 0.4$
2	0.6	$\Rightarrow$ $0.2 + 0.3 + 0.1 = 0.6$
Total	1	

Q. 19 For expectation, if a, b are any arbitrary constants then the value of E (ax + b) is

(a) a E(x) + b E(x) (b) a E(x) – b E(x)

(c) a E(x) . b E(x) (d) a E(x) + b

Ans. : (d)

Explanation : $E(ax + b) = \Sigma P_i (a x_i + b)$

$= a \Sigma P_i x_i + b$

$= a \Sigma x + b$

Q. 20 For expectation, if X and Y are independent variables then the value of E (XY) is

(a) E(X) · E(Y) (b) E(x) + E(x)

(c) E(X + Y) (d) E(x) – E(x)

Ans. : (a)

Explanation : Since

$$E(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} P_i P_j' x_i x_j$$

$$= \left(\sum_i P_i x_i\right)\left(\sum_{j=1}^{n} P_j' y_j\right)$$

$$= E(x) \cdot E(Y)$$

Q. 21 For expectation, if X is independent variable then the value of Var (X) is

(a) $E(X^2)$

(b) $\{E(X)\}^2$

(c) $E(X^2) - \{E(X)\}^2$

(d) $E(X^2) + \{E(X)\}^2$

Ans. : (c)

Explanation :

Since $Var(x) = E\{(X - \mu_x)^2\}$

$= E\{X^2 - 2\mu_x X + \mu_x^2\}$

$= E(X^2) - 2\mu_x E(X) + \mu_x^2$

($\because$ μ_x is mean = constant)

$= E(X^2) - 2\mu_x \cdot \mu_x + \mu_x^2$

($\because$ $E(x) = \mu_x$)

$= E(X^2) - \mu_x^2$

$= E(X^2) - \{E(X)\}^2$

Chapter Ends...

□□□

CHAPTER 3

UNIT III

Probability Distribution

Syllabus

Binomial distribution, Poisson distribution, Normal distribution, Fitting of binomial distributions. Properties of binomial, Poisson and normal distributions, Relation between binomial and normal distributions, Relation between Poisson and normal distributions, Importance of normal distribution, Examples.

Syllabus Topic : Binomial Distribution

3.1 Binomial Distribution

Q. Explain Binomial Distribution in detail and give its formula.

If the trials of a repetitive nature of an experiment in which only occurrence or non-occurrence, **success or failure**, acceptance or rejection, yes or no of a particular event then we can use Binomial distribution.

In a single trial,

p = is the probability success

q = is the probability of failure

Then $p + q = 1$

If trials are independent and P is constant for every trial then,

r successes can be obtained in n trials in ${}^nC_r$ ways

$$\therefore P(X = r) = {}^nC_r\, P(\underbrace{S\,SS \ldots S}_{r\ times}\ \underbrace{FFF \ldots F}_{n-r\ times})$$

$$= {}^nC_r\, \underbrace{P(S)\,P(S)P(S) \ldots P(S)}_{r\ factors}\ \underbrace{P(F)P(F)P(F) \ldots P(F)}_{(n-r)\ factors}$$

$$= {}^nC_r\, \underbrace{P\,PP \ldots P}_{r\ factors}\ \underbrace{q\,qq \ldots q}_{(n-r)\ factors}$$

$$P(X = r) = {}^nC_r\, p^r q^{n-r}$$

Hence $P(X = r) = {}^nC_r\, p^r q^{n-r}$

Where $p + q = 1$ and $r = 0, 1, 2, 3, \ldots, n$

The probability of number of successes so obtained is called the Binomial distribution.

The sum of the probabilities.

$$= {}^nC_0\, p^0 q^n + {}^nC_1\, p\, q^{n-1} + {}^nC_2\, p^2 q^{n-2} + \ldots + {}^nC_n\, p^n$$

$$= (q + p)^n$$

$$= (1)^n (\because p + q = 1)$$

$$= 1$$

$\therefore$ Sum of the probabilities of 0, 1, 2, ……,r,…., n successes is equal to 1.

X - is called Binomial variate.

Hence,

Probability of r success in a series of n trial is

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

Note :

(i) $P(X = r)$ is usually as $P(r)$

(ii) In Binomial distribution n, number of trial are finite.

(iii) Each trial has only two possible outcomes usually called success and failure.

(iv) All the trials are independent.

(v) p (hence q) is constant for all the trials.

Formulae :

1. Sum of probabilities of all possible outcomes is equal to 1.

   $P(x = 0) + P(x = 1) + P(x = 2) + \ldots + P(x = m - 1) + P(x = m) + \ldots + P(x = n) = 1$

2. Probability of getting r successes in n trials,

   $$P(x = r) = {}^nC_r\, p^r q^{n-r}$$

3. The probability of getting at least r successes in n trials

   $P(x \geq r) = P(x = r) + P(x = r + 1) + P(x = r + 2 + \ldots + P(x = 4)$

   $= 1 - [P(x = 0) + P(x = 1) + \ldots + P(x = r - 1)]$

4. The probability of getting at most r successes in n trials

   $P(x \leq r) = P(x = 0) + P(x = 1) + P(x = 2) + \ldots + P(x = r)$

   $= 1 - [(x = r + 1) + P(x = r + 2) + P(x = r + 3) + \ldots + P(x = n)]$

3.1.1 Examples on Binomial Distribution

Ex. 3.1.1 : The probability that a bomb dropped from a plane will strike the target is $\frac{1}{5}$. If six bombs are dropped find the probability that exactly two will strike the target.

Soln. :

Given : Probability of a bomb dropped from a plane will strike the target i.e. $p = \frac{1}{5} = 0.2$

$\therefore$ will not strike $= q = 1 - p = 1 - 0.2 = 0.8$ the target

$\left\{\begin{array}{l} \because p + q = 1 \\ \Rightarrow q = 1 - p \end{array}\right.$

We know, by Binomial distributions, Probability of r success is

$$P(x = r) = {}^nC_r\, p^r q^{n-r} \quad \ldots(1)$$

Here $n = 6$

$\therefore$ Probability of exactly two will strike the target (i.e. $r = 2$) is

$\therefore$ Equation (1) $\Rightarrow P(x = 2) = {}^6C_2\, p^2 q^4$

$$= {}^6C_2\, \underbrace{(0.2)^2}_{p}\, \underbrace{(0.8)^2}_{q} \quad \because P = 0.2 \text{ and } q = 0.8$$

$$\therefore P(x = 2) = \frac{6!}{2!\,(6-2)!}(0.2)^2 (0.8)^2$$

➔ **Using formula of combination :** $\left[{}^nC_r = \frac{n!}{r!\,(n-r)!}\right]$

Here $n = 6$ $r = 2$

$$\therefore P(x = 2) = \frac{\cancel{6}^{3} \times 5 \times \cancel{4!}}{(\cancel{2} \times 1)\,\cancel{4!}}(0.2)^2 (0.8)^2$$

$\therefore P(x = 2) = 15\,(0.2)^2 (0.8)^2 = 0.2458$ on Simplification

$\therefore$ Probability that exactly two will strike the target

$= 0.2458$ ✓ **...Ans.**

Ex. 3.1.2 : If 30% of the bulbs produced are defective. Find probability that out 4 bulbs selected.

(a) One is defective

(b) At the most two are defective.

Soln. : Given : 30% of the bulbs produced are defective

$\therefore$ Probability of defective bulbs = 30%

$$\therefore p = 30\% = \frac{30}{100} = 0.3$$

$P = 0.3$ {Probability of defective}

Since, $p + q = 1$,

$\therefore q = 1 - p = 1 - 0.3 = 0.7$

Here $n = 4$ (Total number of bulbs)

We know, Binomial distribution for, probability of r success is,

$$P(X = r) = {}^nC_r\, p^r q^{n-r} \quad \ldots(1)$$

(i) One is defective :

Probability that one is defective (i.e. $r = 1$) is

$P(X = 1) = {}^4C_1\, p\, q^3$

$\therefore C_1^n = n$ ...[Here $n = 4$, $r = 1$]

Since $p = 0.3$ and $q = 0.7$

$= {}^4C_1\, (0.3)\, (0.7)^3$

$$P(X = 1) = 4\,(0.3)\,\underbrace{(0.7)^3}_{0.343}$$

$$= \underbrace{4\,(0.3)\,(0.343)}_{0.4116}$$

$= 0.4116$

$P(X = 1) = 0.4116$ on simplification

$\therefore$ Probability that one blub is defective = 0.4116 ✓ **...Ans.**

(ii) At most two are defective :

Probability that at most two (i.e $r = 0, 1, 2 \ldots$) are defective is,

Almost two means 0 or 1 or 2 i.e. upto 2

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

Using formula (1) with $n = 4$,

$$= \underbrace{{}^4C_0\, p^0}_{1} q^4 + \underbrace{{}^4C_1}_{4}\, p\, q^3 + {}^4C_2\, p^2 q^2$$

➔ **Using formula of combination :** $\left[{}^nC_r = \frac{n!}{r!\,(n-r)!}\right]$

$${}^nC_0 = 1, {}^nC_1 = n, {}^4C_2 = \frac{4 \times 3}{1 \times 2} = \frac{12}{2} = 6$$

$$= (1)\,(1)\, q^4 + (4)\, pq^3 + (6)\, p^2 q^2$$

$$= q^4 + 4\,p\,q^3 + 6\,p^2 q^2$$

Since, $p = 0.3$ and $q = 0.7$

$$= \underbrace{(0.7)^4}_{0.2401} + 4\,\underbrace{(0.3)}_{0.343}\,(0.7)^3 + 6\,\underbrace{(0.3)^2}_{0.09}\,\underbrace{(0.7)^2}_{0.49}$$

$$= \underbrace{(0.2401) + (4)\,(0.3)\,(0.343)}_{0.4116} + \underbrace{6\,(0.09)\,(0.49)}_{0.2646}$$

$= 0.2401 + 0.4116 + 0.2646$

$= 0.9163$

Probability that at most two bulbs are defective = 0.9163. ✓ **...Ans.**

Ex. 3.1.3 : The probability that a pen manufactured by a company will be defective is 1 /10. If 12 such

pens are manufactured, find the probability that :

(i) Exactly two will be defective.

(ii) At least two will be defective.

(iii) None will be defective.

Soln. :

Given : Probability of defective pen $= p = \frac{1}{10} = 0.1$

Since $p + q = 1$

$\therefore\ q = 1 - p = 1 - 0.1$

$q = 0.9$ and There are 12 pens. $n = 12$

We know, Binomial distribution,

$P(X = r) = {}^nC_r\, p^r q^r$ ...(1)

(i) Probability of exactly two pens will be defective

$P(X = r = 3) = {}^{12}C_3\, p^3 q^9 = {}^{12}C_3\, (0.1)^3 (0.9)^9 = 0.0852$

Probability of exactly two pens will be defective = 0.0852. ✓ **...Ans.**

(ii) Probability of at least two pens will be defective means 2 or 3 or 4 .....

$P(X = r \geq 2) = P(2) + P(3) + \ldots + P(12)$

Since total Probability is 1

$P(0) + P(1) + P(2) + P(3) + P(4) + \ldots + P(12) = 1$

$\therefore\ P(2) + P(3) + P(4) + \ldots + P(12) = 1 - [P(0) + P(1)]$

$= 1 - [P(0) + P(1)]$

Put r = 0 and r = 1 in Equation (1)

$= 1 - [{}^{12}C_0\, p^0 q^{12} + {}^{12}C_1\, p^0 q^{11}]$

$= 1 - [(0.9)^{12} + 12 \times (0.1) \times (0.9)^{11}] = 1 - 0.6590$

$P(X = r \geq 2) = 0.3409$

$\therefore$ **Probability of at least two pens will be defective = 0.3409** ✓ **...Ans.**

(iii) Probability of none will be defective. i.e. r = 0

$P(X = r = 0) = {}^{12}C_0\, p^0 q^{12}$

[put r = 0 in Equation (1)] $= (0.9)^{12}$ ...$\left[\because {}^{12}C_0 = 1 \text{ and } p^0 = 1\right]$

$P(X = r = 0) = 0.2824$

Probability of no pen will be defective = 0.0284. ✓ **...Ans.**

Ex. 3.1.4 : If the chance that out of 10 telephone lines one of the line is busy at any instant is 0.2. What is the chance that 5 of the lines are busy ?

Soln. :

Given : The chance out of the 10 lines are busy at any instant is 0.2

$\therefore\ n = 10$

And $P = 0.2$ (Probability of busy lines)

$\therefore\ q = 1 - p = 1 - 0.2$ (not busy)

$\therefore\ p + q = 1 \Rightarrow q = 1 - p$

$\therefore\ q = 1 - 0.2 = 0.8$

We know,

By Binomial distributions Probability of r success is

$P(x = r) = {}^nC_r\, p^r q^{n-r}$ ...(1)

The chance that 5 out of the lines are busy

i.e. probability that 5 lines are busy (i.e. r = 5) is

$\therefore$ Equation (1) $\Rightarrow P(x = 5) = {}^{10}C_5\, p^5 q^5$ $\quad\therefore\ n = 10$ and $r = 5$

➔ **Using formula of combination :** ...$\left[{}^nC_r = \frac{n!}{r!\,(n-r)!}\right]$

here $p = 0.2$, $q = 0.8$

$\therefore\ P(x = 5) = \frac{10!}{5!\,(10-5)!}\ \underbrace{(0.2)^5}_{p}\ \underbrace{(0.8)^5}_{q}$

$\therefore\ P(x = 5) = \underbrace{\frac{\overset{2}{\cancel{10}} \times 9 \times \overset{2}{\cancel{8}} \times 7 \times \overset{2}{\cancel{6}} \times \cancel{5}!}{\cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times 1 \times \cancel{5}!}} \times (0.2)^5 \times (0.8)^5$

$\therefore\ P(x = 5) = 9 \times 7 \times 4 \times (0.2)^5 \times (0.8)^5$

$= 252 \times (0.00032) \times (0.32768)$ on Simplification

$\therefore\ P(x = 5) = 0.0264$ on Simplification

$\therefore$ **The chances that 5 lines are busy at any instant = 0.0264** ✓ **...Ans.**

Ex. 3.1.5 : The provability that a managed 65 will live to 75 is 0.65. What is the probability that out of 10 men which are now 65, 7 will live to 75 ?

Soln. :

Given : Probability that a man aged 65 will live 75 is 0.65

$\therefore\ P = 0.65$

$\therefore$ Will not Live $\Rightarrow\ q = 1 - P = \underbrace{1 - 0.65}_{0.35}$ $\because\ P + q = 1$ $\Rightarrow q = 1 - P$

$\therefore\ q = 0.35$

Here $n = 10$

We know, by Binomial distribution

Probability of r success is

$P(x = r) = {}^nC_r\, p^r q^{n-r}$ ...(1)

$\therefore$ Probability that 7 men live to 75 (i.e. r = 7) is

$\therefore\ P(x = 7) = {}^{10}C_7\, p^7 q^3$ putting values in Equation (1)

➔ **Using formula of combination :** ...$\left[{}^nC_r = \frac{n!}{r!\,(n-r)!}\right]$

$\therefore\ P(x = 7) = \frac{10!}{7!\,(10-7)!}\, p^7 q^3$ $\quad\because n = 10$ and $r = 7$

$\therefore\ P(x = 7) = \frac{10 \times 9 \times 8 \times \cancel{7!}}{\cancel{7!}\,(3)!}\,(0.65)^7 (0.35)^3$ $\left\{\begin{array}{l}\therefore p = 0.65 \\ \text{and } q = 0.35\end{array}\right.$

$\therefore\ P(x = 7) = \frac{10 \times \overset{3}{\cancel{9}} \times \overset{4}{\cancel{8}}}{\cancel{3} \times \cancel{2} \times 1}\,(0.65)^7 (0.35)^3$

$= 10 \times 3 \times 4 \times (0.65)^7 (0.35)^3$

$= 120\,(0.049)\,(0.42)$ ...on Simplification

$= 0.2522$

$\therefore$ **Probability that out of 10 man 7 will live to 75 = 0.2522** ✓ **...Ans.**

Ex. 3.1.6 : On an average a box containing 10 articles is likely to have 2 defective. If we consider a consignment of too boxes how many of them are expected to have three or less defective.

Soln. :

Given. : On an average a box containing to articles is likely to have 2 defective i.e. out of 10 articles 2 are defective.

$\therefore$ Probability of defective articles $= P = \frac{2}{10}$

$\therefore\ P = 0.2$

$\therefore\ q = 1 - 0.2 = 0.8$ $\quad\because p + q = 1 \Rightarrow q = 1 - p$

Here $n = 10$

We know, By Binomial distribution Probability of success is

$P(x = r) = {}^nC_r\, p^r q^{n-r}$ ...(1)

$P(x \leq 3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$

$= {}^{10}C_0\, p^0 q^{10} + {}^{10}C_1\, p^1 q^9 + {}^{10}C_2\, p^2 q^8 + {}^{10}C_3\, p^3 q^7$

[By Equation (1) with n = 10 and r = 0, 1, 2, 3]

$P(x \leq 3) = {}^{10}C_0\,(0.2)^0 (0.8)^{10} + {}^{10}C_1\,(0.2)^1 (0.8)^9$

$+ {}^{10}C_2\,(0.2)^2 (0.8)^8 + {}^{10}C_3\,(0.2)^3 (0.8)^7$

$\because P = 0.2,\ q = 0.8$

➔ **Using formula of combination :** ...$\left[{}^nC_r = \frac{n!}{r!\,(n-r)!}\right]$

$\therefore\ P(x \leq 3) = \frac{10!}{0!\,(10-0)!}\underbrace{(0.2)^0}_{1}(0.8)^{10}$

$+ \frac{10!}{1!\,(10-1)!}\underbrace{(0.2)^1}_{0.2}(0.8)^9 + \frac{10!}{2!\,(10-2)!}(0.2)^2 (0.8)^8$

$+ \frac{10!}{3!\,(10-3)!}(0.2)^3 (0.8)^7 + \ldots\ldots$

$\therefore\ P(x \leq 3) = \frac{\cancel{10!}}{\cancel{10!}}(1)(0.8)^{10} + \frac{10!}{1\,(9)!}(0.2)(0.8)^9$

As ${}^nC_1 = n$ and ${}^nC_0 = 1$ $\Big\}$ $+ \frac{10!}{2!\,(8)!}(0.2)^2 (0.8)^8 + \frac{10!}{3!\,7!}(0.2)^3 (0.8)^7$

$\therefore\ P(x \leq 3) = (0.8)^{10} + \frac{10 \times \cancel{9!}}{\cancel{9!}}(0.2)(0.8)^9 +$

$\frac{10! \times 9 \times \cancel{8!}}{(2 \times 1)\,\cancel{8!}}(0.2)^2 (0.8)^8 + \frac{10 \times 9 \times 8 \times \cancel{7!}}{3! \times \cancel{7!}}(0.2)^3 (0.8)^7$

$\therefore\ P(x \leq 3) = (0.8)^{10} + 10\,(0.2)\,(0.8)^9 + 90\,(0.2)^2 (0.8)^5$

$+ \frac{10 \times 9 \times 8}{3 \times 2 \times 1}(0.2)^3 (0.8)^7$

$= 0.1074 + 0.2684 + 0.3020 + 0.2013$ ...on Simplification

$\therefore\ P(x \leq 3) = 0.8791$

$\therefore$ Probability of three or less defective articles = 0.8791

$\therefore$ Expected boxes haring three or less defective

$= 0.8791 \times 100$

$= 87.91 \approx 88$ **boxes.** ✓ **...Ans.**

Ex. 3.1.7 : The overall percentage of failure in a certain examination is 20. It six candidates appear in an examinations, what is the probability that at least five pass the examinations.

Soln. : 20% failures in a certain examination

$\therefore\ P = 20\% = \frac{20}{100}$

$P = 0.2$ probability of failures

$\therefore\ q = 1 - p = \underbrace{1 - 0.2}_{0.8}$ $\quad\because p + q = 1$ i.e. $q = 1 - p$

$\therefore\ q = 0.8$ probability of pass

Here six candidates appear in an examination,

$\therefore\ n = 6$

We know, By binomial distributions,

Probability of r success is,

$P(x = r) = {}^nC_r\, p^r q^{n-r}$ ...(1)

$\therefore$ Probability that out of 6 at least five pass the examination means one or none will be fail. (i.e. r = 0, r = 1).

$P(x \leq 1) = P(x = 0) + P(x = 1)$

$= \underbrace{{}^6C_0\, p^0 q^6}_{r=0} + \underbrace{{}^6C_1\, p^1 q^5}_{r=1}$ by Equation (1) [with n = 6, r = 0, 1]

$\therefore\ P(x \leq 1) = \underbrace{{}^6C_0}_{1}\,\underbrace{p^0}_{1}\, q^6 + \underbrace{{}^6C_1}_{6}\,\underbrace{p^1}_{p}\, q^5$

$\therefore\ P(x \leq 1) = \underbrace{(1)\,(1)\, q^6}_{q^6} + \underbrace{6\,(p)\, q^5}_{6\,p\,q^5}$ $\left\{\begin{array}{l}\because {}^nC_0 = 1 \\ p^0 = 1\end{array}\right.$

$\therefore\ P(x \leq 1) = q^6 + 6\,p\,q^5$

$= \underbrace{(0.8)^6}_{0.2621} + \underbrace{6\,(0.2)\,(0.8)^5}_{0.3276}$

$= 0.2621 + 6\,(0.2)\,(0.3276)$ on Simplification

$= 0.6554$

$\therefore\ P(x \leq 1) = 0.6554$

$\therefore$ **Probability that at least five candidates are pass = 0.6554** ✓ **...Ans.**

Ex. 3.1.8 : If 20% of the bolts produce by a machine are defective. Find the probability that out of 4 bolts drawn.

(i) One is defective

(ii) At most two are defective

Soln. :

Given : 20% of bolts produce by a machine are defective.

$\therefore P = 20\% = \frac{20}{100}$

$\therefore P = 0.2$ (Probability defective bolts)

$\therefore q = 1 - p = \underbrace{1 - 0.2}_{0.8}$

$\therefore q = 0.8$ (probability of non-defective bolts)

Her $n = 4$ ($\because$ 4 bolts drawn)

We know, By Binomial distribution Probability of r success is

$P(x = r) = {}^nC_r p^r q^{n-r}$ ...(1)

(i) One is defective

Probability that one defective i.e. $(r = 1)$ is

$P(x = 1) = {}^4C_1 p^1 q^3$

$= {}^4C_1 (0.2)^1 (0.8)^3$

By Equation (1) and here $p = 0.2$ and $q = 0.8$

➔ **Using formula of combination :** $\left[{}^nC_r = \frac{n!}{r!(n-r)!}\right]$

$\therefore P(x = 1) = \frac{4!}{\underbrace{1!(4-1)!}_{4}}(0.2)(0.8)^3$ Also, $n = 4, r = 1$

$\therefore P(x = 1) = 4(0.2)(0.8)^3$

$= 0.4096$ on Simplification

${}^4C_1 = \frac{4!}{1!(4-1)!} = \frac{4 \times 3!}{1!\,3!} = 4$

$\therefore$ Probability that one bolt is defectine = 0.4096 ✓ ...Ans.

(ii) At most two are defective

Probability that at most two (i.e. $r = 0, 1, 2$) are defective is,

$\therefore P(x \le 2) = P(x = 0) + P(x = 1) + P(x = 2)$

$= \underbrace{{}^4C_0}_{1} \underbrace{p^0}_{1} q^4 + \underbrace{{}^4C_1}_{4} p^1 q^3 + {}^4C_2 p^2 q^2$

by Equation (1) with $n = 4$ and $r = 0, 1, 2$

$\therefore P(x \le 2) = (1)(1) q^4 + (4) p^1 q^3 + \frac{4!}{2!(4-2)!} p^2 q^2$

On Simplification

$\left\{\therefore {}^nC_r = \frac{n!}{r!(n-r)!} \text{ and } {}^nC_0 = 1, {}^nC_1 = h, p^0 = 1\right.$

$\therefore P(x \le 2) = q^4 + 4p^1 q^3 + \frac{\cancel{4}^{2} \times 3 \times \cancel{2!}}{\cancel{2} \times 1 \times (\cancel{2!})} p^2 q^2$

$\therefore P(x \le 2) = q^4 + 4p^1 q^3 + 6p^2 q^2$

$\therefore P(x \le 2) = (0.8)^4 + 4(0.2)(0.8)^3 + 6(0.2)^2(0.8)^2$

$= 0.4096 + 0.4096 + 0.1536$ on Simplification

$\therefore P(x \le 2) = 0.9728$

$\therefore$ Probability that at most two bolts are defective

$= 0.9728$ ✓ ...Ans.

Ex. 3.1.9 : The probability that a student who is freshman will graduate is 0.4. Determine the probability that out of 5 students.

(i) None will be graduate

(ii) At least one will be graduate

Soln. :

Given : Probability that a student who is freshman will graduate i.e.

$P = 0.4$

$q = 1 - p = 0.6$ $\because (p + q = 1)$

Here $n = 5$ $\because$ 5 students

We know, By Binomial distribution Probability of r success is $P(x = r) = {}^nC_r p^r q^{n-r}$ ...(1)

(i) None will be graduate

Probability that none (i.e. $r = 0$) will be graduate is

$P(x = 0) = {}^5C_0 p^0 q^5$ {by Equation (1) with $n = 5$ $r = 0$

$\therefore P(x = 0) = \underbrace{{}^5C_0}_{1} (0.4)(0.6)^5$ $\therefore {}^nC_0 = 1$

$= (1)(0.4)(0.6)^5$

$\therefore$ **$P(x = 0) = 0.0778$ ✓** ...Ans. on calculation (2)

(ii) At least one will be graduate

Probability that at least one will (i.e. $r = 1, 2, 3, 4, 5$) be graduate is,

$P(x \ge 1) = P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4) + P(x = 5)$

i.e. $P(x \ge 1) = 1 - [P(x = 0)]$

($\because$ sum of probability is equal to 1)

$\therefore P(x \ge 1) = 1 - 0.0778$ by Equation (2)

$\therefore$ **$P(x \ge 1) = 0.9222$ ✓** ...Ans.

Ex. 3.1.10 : If 10% of riverts produced by a machine are defective find the probability that out of 5 riverts chosen at random.

(i) None will be defective

(ii) One will be defective

(iii) At least two will be defective

Soln. :

Given : 10% of riverts are defective,

$\therefore p = 10\% = \frac{10}{100} = 0.1$

$\therefore q = 1 - p = 1 - 0.1 = 0.9$

$\therefore (p + q = 1)$

Here $n = 5$ $\therefore$ Total 5 riverts

We know, By Binomial distribution Probability of r success is $P(x = r) = {}^nC_r p^r q^{n-r}$ ...(1)

(i) None will be defective :

Probability of none (i.e. $r = 0$) will be defective is,

$P(x = 0) = \underbrace{{}^5C_0}_{1} \underbrace{p^0}_{1} q^5$ [By Equation (1),

with $n = 5, r = 0$]

$\therefore P(x = 0) = (1)(1) q^5$ $\because {}^nC_0 = 1, p^0 = 1$

$= (1)(0.9)^5$ $\because q = 0.9$

$\therefore$ **$P(x = 0) = (0.9)^5 = 0.5905$ ✓** ...Ans.

(ii) One will be defective

Probability of one (i.e. $r = 1$) will be defective is

$P(x = 1) = {}^5C_1 p^1 q^4$ by (1)

$= \underbrace{{}^5C_1}_{5} (0.1)(0.9)^4$ $\therefore {}^nC_1 = n$

$\therefore P(x = 1) = 5(0.1)(0.9)^4$

$= 0.3281$ on Simplification

$\therefore$ **$P(x = 1) = 0.3281$ ✓** ...Ans.

(iii) At least two will be defective :

Probability of least two (i.e. $r = 2, 3, 4, 5$) will be defective is

$P(x \ge 2) = P(x = 2) + P(x = 3) + P(x = 4) + P(x = 5)$

$= 1 - [P(x = 0) + P(x = 1)]$

$\because$ Total sum of probability = 1

$\therefore P(x \ge 2) = 1 - [0.5905 + 0.3281]$ by (i) and (ii)

$\therefore P(x \ge 2) = 1 - 0.9186$ on Simplification

$= 0.0814$

$\therefore$ **$P(x \ge 2) = 0.0814$ ✓** ...Ans.

Ex. 3.1.11 : 10% of the components manufactured by a company are defective. If twelve components are selected at random, find the probability that at least two will be defective.

Soln. :

Given : 10% of the components manufactured by a company are defective.

$\therefore P = 10\% = \frac{10}{100} = 0.1$ probability of defective

$\therefore q = 1 - p = 1 - 0.1$ $\therefore P + q = 1$

$\Rightarrow q = 1 - p$

$\therefore q = 0.9$

Here $n = 12$

$\therefore$ [Twelve components are selected]

We know, By Binomial distribution probability of r success is

$P(x = r) = {}^nC_r p^r q^{n-r}$ ...(1)

Probability of at least two will be defective

(i.e. $r = 2, 3, 4 \ldots\ldots 12$) is

$P(x \ge 2) = P(x = 2) + P(x = 3) + \ldots\ldots + P(x = 12)$

$= 1 - [P(x = 0) + P(x = 1)]$

$\because$ Total probability = 1

$\therefore P(x \ge 2) = 1 - [\underbrace{{}^{12}C_0}_{1} \underbrace{p^0}_{1} q^{12} + \underbrace{{}^{12}C_1}_{12} p^1 q^{11}]$

by Equation (1) with $n = 12$ and $r = 0, 1$

$\therefore P(x \ge 2) = 1 - [(1)(1) q^{12} + (12) p^1 q^{11}]$

$\because {}^nC_0 = 1, {}^nC_1 = n$

$\therefore P(x \ge 2) = 1 - [q^{12} + 12 p^1 q^{11}]$

$= 1 - [(0.9)^{12} + 12 \times (0.1) \times (0.9)^{11}]$

$\because p = 0.1, q = 0.9$

$\therefore P(x \ge 2) = 1 - [0.2824 + 0.3766]$ by Simplification

$\therefore P(x \ge 2) = 1 - 0.659 = 0.341$

$\therefore$ Probability that at least two components are defective

$= 0.341$ ✓ ...Ans.

Ex. 3.1.12 : An unbiased coin is tossed 5 times. Find the probability of getting (a) three heads (b) at least 4 heads.

Soln. :

Given, an unbiased coin is tossed 5 times $\therefore n = 5$

probability of getting head is $P = \frac{1}{2}$

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

$\left\{\begin{array}{l}\because \text{ for one coin } S = \{H, T\} \\ \text{probability of head} = \frac{n(A)}{n(S)} = \frac{1}{2}\end{array}\right.$

We know, By Binomial distribution, probability of r success is

$P(x = r) = {}^nC_r p^r q^{n-r}$ ...(1)

(a) Three heads :

Probability of getting these heads (i.e. $r = 3$) is,

$P(x = 3) = \underbrace{{}^5C_3}_{q^2} p^3 q^{5-3}$

by Equation(1) with $n = 5, r = 3$

$\therefore P(x = 3) = {}^5C_3 p^3 q^2$

➔ **Using formula of combination :** $\left[{}^nC_r = \frac{n!}{r!(n-r)!}\right]$

Also, here $P = q = \frac{1}{2}$

$\therefore P(x = 3) = \frac{5!}{3!(5-3)!} \times \overbrace{\left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2}^{\left(\frac{1}{2}\right)^{3-2}}$ $\{\because x^{m+n} = x^m x^n$

$\therefore P(x = 3) = \frac{5 \times 4 \times 3!}{3!(2)!} \times \left(\frac{1}{2}\right)^{3+2}$

$\therefore P(x=3) = \frac{5\times \cancel{4}^{2}}{\cancel{2}\times 1}\times\left(\frac{1}{2}\right)^5$ $\therefore$ on Simplification

$= \cancel{10}^{5}\times\frac{1}{\cancel{32}_{16}} = \frac{5}{16}$ Div N^r and D^r by 2

$\therefore P(x=3) = \frac{5}{16}$

$\therefore$ **Probability of getting three heads** $= \frac{5}{16}$ ✓ **...Ans.**

(b) At least 4 heads :

Probability of getting at least 4 (i.e. r = 4, 5) is

$\therefore P(x\geq 4) = P(x=4)+P(x=5)$

$= {}^5C_4\, p^4 q^1 + \underbrace{{}^5C_5}_{1}\, p^5 \underbrace{q^0}_{1}$

by Equation (1) with n = 5 and r = 4, 5

$\therefore P(x\geq 4) = {}^5C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 + (1)\left(\frac{1}{2}\right)^5(1)$

$\because P = q = \frac{1}{2}$

${}^nC_n = 1$

➔ **Using formula of combination :** $\ldots\left[{}^nC_r = \frac{n!}{r!(n-r)!}\right]$

and $x^{m+n} = x^m x^n$

$\therefore P(x\geq 4) = \frac{5!}{4!(5-4)!}\times\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^5$

$\therefore P(x\geq 4) = \frac{5\times\cancel{4!}}{\cancel{4!}1!}\left(\frac{1}{2}\right)^{4+1} + \left(\frac{1}{2}\right)^5$

$= \underbrace{\frac{5}{1}\left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5}_{\left(\frac{1}{2}\right)^5(5+1)}$

Taking $\left(\frac{1}{2}\right)^5$ common.

$\therefore P(x\geq 4) = \left(\frac{1}{2}\right)^5(6)$

$= \frac{1}{32}(6)$ on Simplification

Div N^r and D^r by 2 we get

$\therefore P(x\geq 4) = \frac{\frac{6}{2}}{\frac{32}{2}} = \frac{3}{16}$

Probability of getting at least 4 heads $= \frac{3}{16}$ ✓ **...Ans.**

Ex. 3.1.13 : In 256 sets of 12 tosses of a coin, in how many cases one can expect 8 heads and 4 tails.

Soln. :

Total 12 tosses of a coin, n = 12 probability of getting head is $P = \frac{1}{2}$

$\therefore$ for one coin s = {H, T}

Probability of head $= \frac{n(A)}{n(S)} = \frac{1}{2}$

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

We know, by binomial distribution, probability of r success is,

$P(x=r) = {}^nC_r\, p^r q^{n-r}$ ...(1)

Probability of getting 8 heads (i.e. r = 8) is

$P(x=8) = {}^{12}C_8\, \underbrace{p^8}_{q^4} q^{12-8}$

by (1) with n = 12 and r = 8

➔ **Using formula of combination :** $\ldots\left[{}^nC_r = \frac{n!}{r!(n-r)!}\right]$

$\therefore P(x=8) = {}^{12}C_8\times\underbrace{\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^4}_{\left(\frac{1}{2}\right)^{8+4}}$ $\because p = q = \frac{1}{2}$

$\therefore P(x=8) = \frac{12!}{8!\underbrace{(12-8)!}_{4}}\left(\frac{1}{2}\right)^{8+4}$ $\because x^{m+n} = x^m x^n$

$\therefore P(x=8) = \frac{12\times 11\times 10\times 9\times \cancel{8!}}{\cancel{8!}\ 4!}\left(\frac{1}{2}\right)^{12}$

$\therefore P(x=8) = \frac{\cancel{12}^{\cancel{3}}\times 11\times\cancel{10}^{5}\times 9}{\cancel{4}\times\cancel{3}\times\cancel{2}\times 1}\left(\frac{1}{2}\right)^{12}$

$\therefore P(x=8) = (11\times 5\times 9)\left(\frac{1}{2}\right)^{12}$

$= 495\times\left(\frac{1}{4096}\right) = 0.1208$

$\therefore P(x=8) = 0.1208$

Total number of sets = 256

$\therefore$ In total 256 sets, total no. of cases Getting 8 heads and 4 tails

$= 0.1208\times 256 = 30.92 \approx 31$

$\therefore$ **Expected cases = 31** ✓ **...Ans.**

Ex. 3.1.14 : Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) at least one boy (iii) no girl (iv) at most two girls.

Soln.:

Step I : Given : n = 4 ; N = 800, $P = \frac{1}{2}$; $q = \frac{1}{2}$.

By Binomial distribution, $NP(r) = N\,{}^nC_r\, p^r q^{n-r}$

Step II :

(i) The expected number of families having 2 boys and 2 girls

$= N\,P(r) = 800\times\left[{}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2\right] = 800\times 6\times\frac{1}{16}$

$= 300$

Step III :

(ii) The expected number of families having at least one boy = P(1) + P(2) + P(3) + P(4)

$= 800\left[{}^4C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 + {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + {}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + {}^4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0\right]$

$= 750$ or $N[1 - P(x=r=0)] = N\left[1 - {}^4C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{4}\right)^4\right]$

$= 800\left[1 - \frac{1}{2^4}\right]$

Step IV :

(iii) The expected number of families having no girls i.e. having 4 boys

$P(r=4) = 800\times{}^4C_4\left(\frac{1}{2}\right)^4 = 50$

Step V :

(iv) The expected number of families having at most two girls i.e. having at least two boys

$= N[P(r=2) + P(r=3) + P(r=4)]$

$= 800\left[{}^4C_2\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + {}^4C_3\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^1 + {}^4C_4\left(\frac{1}{2}\right)^4\right]$

$= 550$

Ex. 3.1.15 : 10 coins thrown simultaneously. Find the probability of getting.

(i) 8 Heads (ii) At least 6 heads

Soln. :

Step I : Here, n = 10

$p = \frac{1}{2}$, $q = \frac{1}{2}$

Step II : By Binomial distribution

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

(i) Probability of getting exactly 8 heads,

$P(r=8) = {}^{10}C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 = \frac{10!}{8!\,2!}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2$

$= \frac{10!}{8!\times 2!\times 2^{10}}$

$= \frac{9\times 10}{2\times 2^{10}} = \frac{45}{1024} = 0.0439$

Probability of getting exactly 8 heads = 0.0439

Step III :

(ii) At least 6 heads

r = 6, 7, 8, 9, 10

$P(r=7,8,9,10) = \left[{}^{10}C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3 + {}^{10}C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + {}^{10}C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^1 + {}^{10}C_{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0\right]$

$= \left[\frac{120}{2^{10}} + \frac{45}{2^{10}} + \frac{10}{2^{10}} + \frac{1}{2^{10}}\right] = \frac{176}{2^{10}} = 0.3769$

Probability getting at least 6 heads = 0.3769

Ex. 3.1.16 : The mean and variance of Binomial distribution are 4 and 2 respectively. Find $P(r\geq 2)$.

Soln. :

Step I : Given : mean = μ = 4, variance = $\sigma^2 = 2$

Since, $\sigma^2 = npq$ and $\mu = np$

$\therefore 2 = (np)q$

$= 4q \Rightarrow q = \frac{1}{2}$

$\therefore p = 1 - q = \frac{1}{2}$ and $n = \frac{\mu}{p} = \frac{4}{\frac{1}{2}} = 8$

$\therefore P(r\geq 2) = 1 - P(r=0,1)$

Step II : By Binomial distribution,

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

$P(r\geq 2) = 1 - [P(r=0) + P(r=1)]$

$= 1 - \left[{}^8C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^8 + {}^8C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7\right]$

$= 1 - \left[1\times\left(\frac{1}{2}\right)^8 + 8\times\frac{1}{2}\times\left(\frac{1}{2}\right)^7\right]$

$= 1 - .03516 = 0.96484$

Syllabus Topic : Poisson Distribution

3.2 Poisson's Distribution

Poisson's distribution related to the probabilities of event which are extremely rare, but which have a large number of independent opportunities for occurrence.

This distribution can be used if n very large and p very small.

3.2.1 Mean of Poisson's Distribution

Mean or average value of the Poisson's distribution is given by,

$M = np$

3.2.2 Variance of Poisson's Distribution

Variance of the Poisson's distribution is given by,

$V(X) = \sigma^2 = M$

3.2.3 Standard Deviation of Poisson's Distribution

Standard deviation (S.D) of Poisson's distribution is given by,

$$\text{S.D.} = \sigma = \sqrt{M}$$

3.2.4 Examples on Poisson's Distribution

Ex. 3.2.1 : The probability of getting an item defective is 0.005. What is the probability that exactly 3 items in a sample of 200 are defective (Given $e^{-1} = 0.3679$)

Soln. :

Given probability of getting an item defective is P = 0.005,

Also, n = 200

Note that, here P is very small and n is large.

∴ Use Poisson's Distribution we know, by Poisson distribution Probability of r success is

$$P(x=r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

Where m = mean = np = $\underbrace{200 \times 0.005}_{1}$

∴ m = 1

Probability that exactly 3 items (i.e. r = 3) are defective is

$$P(x=3) = \frac{m^3 e^{-m}}{3!} \quad \because r = 3 \text{ by Equation (1)}$$

$$\therefore P(x=3) = \frac{(1)^3 e^{-1}}{3!} \quad \because m = 1$$

$$= \frac{(1)\, e^{-1}}{3 \times 2 \times 1} \quad \because n! = n(n-1)$$

$$= \frac{1 e^{-1}}{3 \times 2} = \frac{e^{-1}}{6}$$

$$\therefore P(x=3) = \frac{1}{6}(0.3679) \quad \because \text{Given } e^{-1} = 0.3679$$

$$= 0.0613 \quad \text{on calculation}$$

Probability that exactly 3 are defective = 0.0613 ✓ ...Ans.

Ex. 3.2.2 : If 2% of the electric bulbs manufactured by a company are defective. Find the probability that in a sample of 100 bulbs

(i) 3 are defective

(ii) At least two are defective.

Soln. :

Given : 2% of the electric bulbs one defective

$$\therefore P = 2\% = \underbrace{\frac{2}{100}}_{0.02}$$

$$\therefore P = 0.02 \quad \{\text{Probability of defective}\}$$

Here n = 100

Note that here p is small and n is large there for use Poisson's distribution we know, by Poisson's distribution.

Probability of r success is

$$P(x=r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

Where m = mean = np

$$\therefore \text{mean} = m = 100 \times 0.02 \quad \because n = 100, p = 0.02$$

$$\therefore m = 2 \quad ...\text{on calculation}$$

(i) Three bulbs are defective :

Probability that 3 are defective i.e. (r = 3) is

$$P(x=3) = \frac{m^r e^{-m}}{r!}$$

$$= \frac{2^3 e^{-2}}{3!} \quad \because m = 2, r = 3 \text{ by Equation (1)}$$

$$\therefore P(x=3) = \frac{8 e^{-2}}{3 \times 2 \times 1} \quad \left\{\begin{array}{l} \because 2^3 = 8 \\ \because e^{-2} = 0.1353 \\ \text{and } 3! = 3 \times 2 \times 1 \end{array}\right.$$

$$= \frac{\cancel{8}^{4} \times (0.1353)}{\cancel{6}_{3}} \quad ...\text{On simplification}$$

$$= \frac{0.5412}{3} = 0.1804$$

$$\therefore P(x=3) = 0.1804$$

Probability that 3 bulbs are defective = 0.1804 ✓ ...Ans.

(ii) At least two bulbs are defective :

Probability that at least two are defective (i.e. r = 2, 3, 4, 5 ...100) is

∴ Total probability = 1

Also, $P(x \geq 2) = [P(x=2) + P(x=3) + \ldots\ldots P(x=100)]$

$$\therefore P(x \geq 2) = 1 - [\underbrace{P(x=0)}_{\frac{m^0 e^{-m}}{0!}} + \underbrace{P(x=1)}_{\frac{m^1 e^{-m}}{1!}}]$$

$$\therefore P(X \geq 2) = 1 - \left[\frac{m^0 e^{-m}}{0!} + \frac{m^1 e^{-m}}{1!}\right] \quad ...\text{by(1)}$$

$$\therefore P(x \geq 2) = 1 - \underbrace{\left[\frac{1 e^{-m}}{1} + \frac{m e^{-m}}{1}\right]}_{(1+m)\, e^{-m}}$$

With r = 0, 1 ∴ 0! = 1! = 1 $m^0 = 1$ $m^1 = m$

$$\therefore P(x \geq 2) = 1 - (1+m)\, e^{-m} \quad ...\text{on simplification}$$

$$= 1 - (1+2)\, e^{-2} \quad \because m = 2$$

$$\therefore P(x \geq 2) = 1 - (3)\, e^{-2}$$

$$= 1 - (3 \times 0.1353) \quad \because e^{-2} = 0.1353$$

$$= 1 - 0.4059 \quad ..\text{on simplification}$$

$$= 0.5941$$

∴ Probability that at least two bulbs are defective = 0.5941 ✓ ...Ans.

Ex. 3.2.3 : If probability that an electric motor is defective is 0.01. What is probability that sample of 300 electric motor will contain exactly 5 defective motor. ($e^{-3} = 0.0498$).

Soln. :

Given :

Probability of an electric motor is defective = p = 0.01

Sample of 300 electric motor : n = 300

Note that here p is very small and n is large

∴ Use Poisson's Distribution,

We know, Probability of r success is,

$$P(X=r) = \frac{m^r e^{-m}}{r!}$$

Where m = mean = np = 300 × 0.01

∴ m = 3

∴Probability of exactly 5 (i.e. r = 5) defective motors is,

$$P(X=5) = \frac{m^5 e^{-m}}{5!} = \frac{(m)^5 e^{-m}}{5!}$$

$$= \frac{(3)^5 (0.0498)}{5!} = 12.1014 \quad (\because e^{-3} = 0.0498)$$

$$P(X=5) = \frac{12.1014}{5!} = 0.1008$$

∴ Probability of exactly 5 defective motors = 0.1008 ✓

...Ans.

Ex. 3.2.4 : One percent of articles from a certain machine are defective. What is the probability of (i) no defective, (ii) one defective and (iii) two or more defective in a sample of 100

Soln. : Given : one percent of articles are defective P = 1% $= \frac{1}{100}$

∴ P = 0.01 Here n = 100

∴ p is very small and n is large, so, use Poisson's distribution we know by Poisson's distribution

$$P(x=r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

Where m = mean = np

$$\text{mean} = m = \underbrace{100 \times 0.01}_{1}$$

∴ m = 1

(i) No articles is defective

Probability of number defective articles (i.e. r = 0) is

$$P(x=0) = \underbrace{\frac{m^0 e^{-m}}{0!}}_{e^{-m}} \quad ...(1)$$

$\because r = 0$, $\because m^0 = 1$, $0! = 1$

$$P(x=0) = e^{-m}$$

$$= e^{-1} \quad \because m = 1$$

$$P(x=0) = 0.3679 \quad \because e^{-1} = 0.3679 \quad ...(2)$$

Probability of no defective articles = 0.3679 ✓ ...Ans.

(ii) One article is defective :

Probability of one defective articles (i.e. r = 1) is

$$P(x=1) = \frac{m e^{-m}}{1!} \quad \because n = 1 \text{ in (1) by Equation (1)}$$

$$P(x=1) = \frac{1 e^{-1}}{1} \quad \because m = 1 \text{ and } 1! = 1$$

...by Equation (1)

$$P(x=1) = e^{-1}$$

$$= 0.3679 \quad \because e^{-1} = 0.3679 \quad ...(3)$$

Probability of one defective articles = 0.3679 ✓ ...Ans.

(iii) Two or more articles are defective

Probability of two or more defective (i.e. r ≥ 2) (r = 2, 3, ...) is

$$P(x \geq 2) = P(x=2) + P(x=3) + \ldots$$

Total Probability = 1

$$P(x \geq 2) = 1 - [\underbrace{p(x=0)}_{0.3679} + \underbrace{p(x=1)}_{0.3679}]$$

$$= 1 - [0.3679 + 0.3679]$$

By Equations (2) and (3)

$$P(x \geq 2) = 1 - 0.7358$$

$$P(x \geq 2) = 0.2642 \quad \text{on calculation}$$

Probability of two or more defective articles = 0.2642 ✓

...Ans.

Ex. 3.2.5 : A firm produces articles of which 0.1 % are defective, out of 500 articles. If wholesaler purchases 100 such cases, how many can be expected to have one defective :

Given $e^{-0.5} = 0.605$.

Soln. : A firm produces articles of which 0.1 % are defective, out of 500 articles.

$$\therefore P = 0.1\% = \frac{0.1}{100} = 0.001$$

n = 500 (Total articles)

We know, by Poisson's distribution,

Probability of r success is,

$$\therefore \text{mean} = m = np = 500 \times 0.001 = 0.5$$

$$P(X=r) = \frac{m^r e^{-m}}{r} \quad ...(1)$$

∴ probability of one defective (i.e. r = 1)

Substitute r = 1 in Equation (1)

$$P(r=1) = \frac{m e^{-m}}{1!} \quad \text{Since } m = 0.5$$

$$= \frac{(0.5)\, e^{-0.5}}{1}$$

$$= \frac{(0.5)(0.6065)}{1} \quad (\because e^{-0.5} = 0.6065)$$

$$P(r=1) = 0.3031$$

$\therefore$ Probability of one defective = 0.3031

If wholesaler purchases 100 such cases then

$$= 100 \times 0.3031 = 30.31 \approx 30$$

$\therefore$ **30 cases can be expected to have one defective out of 100 cases.** ✓ ...Ans.

Ex. 3.2.6 : 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 at least 3 defective parts.

Soln. :

Given :

The mean number of defectives in a sample of 20 is 2

$\therefore n = 20$ and mean $= m = 2$

We know, by Possion's Distribution,

Probability of r success is,

$$P(r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

Probability of at least 3 defective parts means 3, 4, 5, ... defective parts

$\therefore$ Probability that 3 defective i.e. $(x \geq 3)$ are,

$$P(r \geq 3) = P(3) + P(4) + P(5) + ... + P(20)$$

We know, probability of total defective parts is 1

i.e. $\underbrace{P(0) + P(1) + P(2)} + P(3) + ... = 1$

$$\therefore P(3) + P(4) + P(5) + ... + P(20) = 1 - [P(0) + P(1) + P(2)]$$

$$P(r \geq 3) = 1 - [\underbrace{P(0)}_{\frac{m^0 e^{-m}}{0!}} + \underbrace{P(1)}_{\frac{m^1 e^{-m}}{1!}} + \underbrace{P(2)}_{\frac{m^2 e^{-m}}{2!}}]$$

...Using formula (1)

With r = 0, 1, 2

Since, m = 2,

$$P(r \geq 3) = 1 - \left[\frac{2^0 e^{-2}}{0!} + \frac{2 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!}\right]$$

$$0! = 1, 1! = 1 \text{ and } 2! = 1 \times 2 = 2$$

$$= 1 - \left[\frac{e^{-2}}{1} + \frac{2 e^{-2}}{1} + \frac{2^2 e^{-2}}{2}\right]$$

$$= 1 - e^{-2}[1 + 2 + 2]$$

$$= 1 - e^{-2}(5) = 1 - \underbrace{(0.1353)}_{e^{-2}}(5)$$

$$= 1 - 0.6765 \text{ ... on simplification}$$

$$P(r \geq 3) = 0.3235$$

$\therefore$ Out of 1000 sample contain at least 3 defective parts

$$= 0.3235 \times 1000 = 323.5 \approx 324 \checkmark$$ ...Ans.

Ex. 3.2.7 : **If 5% of the electric bulbs manufacturing by a company are defective, use Poisson's Distribution to find the probability that in a sample of 100 bulbs.**

(i) None is defective

(ii) Five bulbs are defective

(Given $e^{-5} = 0.007$).

Soln. :

Given 5 % of the electric bulbs are defective

$$\therefore p = 5\% = \frac{5}{100} = 0.05$$

In a sample of 100 bulbs, $\therefore n = 100$

Note that here p is small and n is large therefore use Poisson's Distribution

We know, by Poisson's distribution,

Probability of r success is,

$$P(X = r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

Where mean $= m = np$

mean $= m = 100 \times 0.05 = 5$ $(\because n = 100, p = 0.05)$

$m = 5$

(i) None bulb is defective out of 100 bulbs,

Probability of zero bulb defective (i.e. r = 0) is,

$$P(X = r = 0) = \underbrace{\frac{m^0 e^{-m}}{0!}}_{e^{-m}} \quad [\because \text{Put } r = 0 \text{ in Equation (1)}]$$

$$\because m^0 = 1 \text{ and } 0! = 1 = \frac{(1)e^{-m}}{(1)} = e^{-m}$$

since mean $= m = 5 = e^{-5} = 0.007$

$\therefore$ **Probability that none bulb is defective = 0.007** ✓

...Ans.

(ii) Five bulbs are defective :

Probability of five bulbs defective (i.e. r = 5) are,

$$P(X = r = 5) = \frac{m^5 e^{-m}}{5!} \quad [\because r = 5 \text{ in Equation (1)}]$$

$$\text{Since } m = 5 = \frac{(5)^5 e^{-5}}{5!} = \frac{3125}{120} \times 0.007 = 0.1823$$

$\therefore$ **Probability that five bulbs are defective = 0.1823** ✓

...Ans.

Ex. 3.2.8 : Assuming that the probability of a fatal accident during the year is $\frac{1}{1200}$, Calculate the probability that in a factory employing 300 workers there will be at least two fatal accidents in a year. [$e^{-0.25} = 0.7788$].

Soln. :

Given :

Probability of a fatal accident during the year $= P = \frac{1}{1200}$

Total number of workers $= n = 300$

We know, Mean $= m = np = 300 \times \frac{1}{1200} = \frac{1}{4} = 0.25$

We know, Poisson's Distribution of r success,

$$P(X = r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

$\therefore$ Probability of at least two fatal accidents in a year means 2 or 3 or 4 or 5.....

$$P(X = r \geq 2) = P(2) + P(3) + P(4) + ...$$

We know total probability is 1

$$\therefore P(0) + P(1) + P(2) + P(3) + ... = 1$$

$$\therefore P(2) + P(2) + P(3) + ... = 1 - [P(0) + P(1)]$$

$$= 1 - \left[\frac{m^0 e^{-m}}{0!} + \frac{m e^{-m}}{1!}\right]$$

Using Equation (1) with r = 0 and r = 1

$$= 1 - \left[\frac{(1)e^{-m}}{(1)} + \frac{m e^{-m}}{1}\right] \quad ...[\because 0! = 1 \text{ and } 1! = 1]$$

$$= 1 - [e^{-m} + m e^{-m}] = 1 - e^{-m}(1 + m)$$

$$P(X = r \geq 2) = 1 - [P(0) + P(1)]$$

Since m = 0.25

$$= 1 - e^{-0.25}[1 + 0.25] = 1 - (0.7788)[1.25]$$

$$= 1 - 0.9735$$

$\therefore$ **P (X = r ≥ 2) = 0.0265** ✓ ...Ans.

Ex. 3.2.9 : Using Poisson distribution, find the probability that the ace of spades will be drawn from a pack of well shuffled cards getting at least one is 104 consecutive trials.

Soln. : In a pack 52 cards and lace of spades

Probability of drawn card is ace of spade is

$$p = \frac{1}{52} = 0.0192 \quad \text{(p is very small)}$$

Here 104 consecutive trials.

$$n = 104$$

We know, by poisson's distribution, Probability of r success is

$$P(x = r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

Where m = Mean = np

$$m = 104 \times \frac{1}{52} \quad \therefore n = 104 \text{ and } P = \frac{1}{52}$$

$$\therefore m = 2$$

Proof of at least one drawn card is ace of spade (i.e. r = 1, 2, 3, ...)

$$P(x \geq 1) = P(x = 1) + P(x = 2) + P(x = 3) ...$$

Total Probability = 1

$$P(x \geq 1) = 1 - \underbrace{p(x = 0)}_{\frac{m^0 e^{-m}}{0!}}$$

...by Equation (1) Here r = 0

$$P(x \geq 1) = 1 - \underbrace{\frac{m^0 e^{-m}}{0!}}_{e^{-m}} \quad \because m^0 = 0! = 1$$

$$p(x \geq 1) = 1 - e^{-m}$$

$$= 1 - e^{-2} \quad \because m = 2$$

$$= 1 - 0.1353 \quad \because e^{-2} = 0.1353$$

$$P(x \geq 1) = 0.8647$$

Probability of at least one drawn card is ace of spade = 0.8647 ✓ ...Ans.

Ex. 3.2.10 : If the probability of a bad reaction from a certain injection is 0.001, determine the chance that out of 2,000 individuals more than two will get a bad reactions . (Given $e^2 = 7.4$)

Soln. : Given Probability of bad reaction

$$p = 0.001 \quad \text{and} \quad n = 2000$$

we know by Poisson's distribution Probability by r success is

$$P(x = r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

where m = mean = np $\quad \because n = 2000$ and $p = 0.001$

$$\therefore m = 2000 \times 0.001;$$

$$\therefore m = 2 \quad \text{on calculation,}$$

Probability that more than two will get a bad reaction i.e. $(r \geq 2)$ (r = 3, 4, 5,...) is

$$P(x > 2) = P(x = 3) + P(x = 4) + ...$$

Total Probability = 1

$$P(x > 2) = 1 - [\underbrace{P(x = 0)}_{\frac{m^0 e^{-m}}{0!}} + \underbrace{P(x = 1)}_{\frac{m^1 e^{-m}}{1!}} + \underbrace{P(x = 2)}_{\frac{m^2 e^{-m}}{2!}}]$$

$$P(x > 2) = 1 - \left[\underbrace{\frac{m^0 e^{-m}}{0!}}_{e^{-m}} + \underbrace{\frac{m^1 e^{-m}}{1!}}_{m e^{-m}} + \underbrace{\frac{m^2 e^{-m}}{2!}}_{\frac{m^2 e^{-m}}{2!}}\right]$$

...by Equation (1)

with r = 0, 1, 2

$$P(x > 2) = 1 - \left[e^{-m} + m e^{-m} + \frac{1}{2} m^2 e^{-m}\right]$$

...on simplification

$$\because m^0 = 0! = 1; \; m^1 = m; \; 2! = 2 \times 1 = 2$$

$$P(x>2) = 1-e^{-m}\left[1+m+\frac{1}{2}m^2\right]$$

...Taking e^{-m} common

$$P(x>2) = 1-e^{-2}\left[1+2+\underbrace{\frac{1}{2}(2)^2}_{4}\right]$$

$$P(x>2) = \underbrace{1-e^{-2}}_{\frac{1}{e^2}}\left[1+2+\underbrace{\frac{\cancel{4}^{2}}{\cancel{2}}}_{2}\right]$$

$$P(x>2) = 1-\frac{1}{e^2}\underbrace{(1+2+2)}_{5} \quad \because e^{-m}=\frac{1}{m}$$

$$P(x>2) = 1-\frac{1}{7.4}(5) \quad \because e^2=7.4$$

$$P(x>2) = 1-(0.1351\times 5) \quad \because \frac{1}{7.4}=0.1301$$

$$= 1-0.6755$$

$$= 0.3245 \quad \text{on simplification}$$

Chance that more than two will get bad reaction = 0.3245 ✓

...Ans.

Ex. 3.2.11 : Number of road accidents follows a Poisson's distribution with mean 5. Find the probability that in a certain month number of accidents on the high way will be (i) less than 3 (ii) more than 3.

Soln. :

Given : Mean = m = 5

We know, by Poisson's distribution, probability of r success is

$$P(x=r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

(i) Less than 3 accidents

Probability that accidents less than 3 (r < 3) is

$$P(x<3) = \underbrace{P(x=0)}_{\frac{m^0e^{-m}}{0!}}+\underbrace{P(x=1)}_{\frac{m^1e^{-m}}{1!}}+\underbrace{P(x=2)}_{\frac{m^2e^{-m}}{2!}}$$

$$\therefore P(x<3) = \underbrace{\frac{m^0e^{-m}}{0!}}_{e^{-m}}+\underbrace{\frac{m^1e^{-m}}{1!}}_{m^1e^{-m}}+\underbrace{\frac{m^2e^{-m}}{2!}}_{\frac{m^2e^{-m}}{2}}$$

...by Equation (1) with r = 0, 1, 2

$$P(x<3) = e^{-m}+m^1e^{-m}+m^2e^{-m} \quad \because m^0=0!=1,\ m^1=m,\ 2!=2\times1=2$$

$$P(x<3) = e^{-m}\left(1+m+\frac{m^2}{2}\right) \quad ...\text{Taking } e^{-m} \text{ common}$$

$$P(x<3) = e^{-5}\left(1+5+\frac{(5)^2}{2}\right)$$

$$= (0.0067)\left(1+5+\underbrace{\frac{25}{2}}_{12.5}\right)$$

$$\therefore P(x<3) = (0.0067)\underbrace{(1+5+12.5)}_{18.5}$$

$$P(x<3) = 0.0067\times 18.5 = 0.12395$$

$$P(x<3) = 0.1240$$

Probability that less than 3 accidents = 0.1240✓ ...Ans.

(ii) More than 3 accidents

Probability that accidents more than 3 (r > 3) is,

$$P(x>3) = P(x=4)+P(x=5)+.....$$

$\therefore$ total probability = 1

$$\therefore P(x>3) = 1-[P(x=0)+P(x=1)+P(x=2)+P(x=3)]$$

$$P(x>3) = 1-\left[\underbrace{\frac{m^0e^{-m}}{0!}}_{e^{-m}}+\underbrace{\frac{m^1e^{-m}}{1!}}_{me^{-m}}+\underbrace{\frac{m^2e^{-m}}{2!}}_{\frac{m^2e^{-m}}{2}}+\underbrace{\frac{m^3e^{-m}}{3!}}_{\frac{m^3e^{-m}}{3\times2\times1}}\right]$$

...by using Equation (1) with r = 0, 1, 2, 3....

$$P(x>3) = 1-\left[e^{-m}+me^{-m}+\frac{1}{2}m^2e^{-m}+\underbrace{\frac{1}{3\times2\times1}}_{\frac{1}{6}}m^3e^{-m}\right]$$

$$P(x>3) = 1-\left[e^{-m}+me^{-m}+\frac{1}{2}m^2e^{-m}+\frac{1}{6}m^3e^{-m}\right]$$

$$= 1-e^{-m}\left[1+m+\frac{1}{2}m^2e^{-m}+\frac{1}{6}m^3\right]$$

Taking e^{-m} common

$$P(x>3) = 1-e^{-5}\left[1+5+\underbrace{\frac{1}{2}(5)^2}_{25}+\underbrace{\frac{1}{6}(5)^3}_{125}\right]$$

$$\therefore P(x>3) = 1-e^{-5}[1+5+\underbrace{25/2}_{12.5}+\underbrace{125/6}_{20.83}]$$

$$= 1-(0.006)[1+5+12.5+20.83]$$

$$= 1-\underbrace{(0.006)(0.00076)}_{0.2650}$$

$$= 1-0.2650$$

$$P(x>3) = 0.7350$$

Probability that more than 3 accidents = 0.7350 ✓ ...Ans.

Ex. 3.2.12 : The number of road accidents meet with by taxi drivers follows Poisson's distribution with mean 2 Out of 5000 taxi in the city find the number of drivers

(i) who does not meet an accident

(ii) who meet with an accident more than 3 times

(Given e^{-2} = 0.1.353)

Soln. : Given : Mean = 2 and

Total number of taxi = total no of drivers

$$n = 5000$$

We know, by Poisson's distribution, probability of r success is

$$p(x=r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

(i) who does not meet an accident

Probability of no, accident (i.e. r = 0) is

$$P(x=0) = \frac{m^0e^{-m}}{0!} \quad \because m^0=0!=1$$

$$P(x=0) = \frac{1e^{-m}}{1} = e^{-m}$$

$$P(x=0) = e^{-2} \quad [\because m=2$$

$$P(x=0) = 0.1353$$

No. of drivers who does not meet an accident

$$= 0.1353\times 5000 \quad \left\{\begin{array}{l}\text{Total no of}\\ \text{taxi} = 5000\end{array}\right.$$

$$= 676.5$$

$$\approx 677$$

No. of drivers who does not meet an accidnet = 677✓

...Ans.

(ii) Who meet with an accident more than 3 times.

Probability of more than 3 times accident (i.e. r > 3) (r = 4, 5, 6,...) is

$$P(x>3) = P(x=4)+P(x=5)+....$$

Total Probability = 1

$$P(x>3) = 1-[P(x=0)+P(x=1)+P(x=2)+P(x=3)]$$

$$= 1-\left[\frac{m^0e^{-m}}{0!}+\frac{m^1e^{-m}}{1!}+\frac{m^2e^{-m}}{2!}+\frac{m^3e^{-m}}{3!}\right]$$

By Equation (1) with r = 0, 1, 2, 3....

$$P(x>3) = 1-\left[e^{-m}+me^{-m}+\frac{m^2e^{-m}}{2}+\frac{m^3e^{-m}}{3\times2}\right]$$

...On simplification

$$P(x>3) = 1-e^{-m}\left[1+m+\frac{m^2}{2}+\frac{m^3}{6}\right]$$

Now, put m = 2, we get

$$P(x>3) = \underbrace{1-e^{-2}}_{0.1353}\left[1+2+\underbrace{\frac{(2)^2}{2}}_{\frac{4}{2}}+\underbrace{\frac{(2)^3}{6}}_{\frac{8}{6}}\right]$$

$$P(x>3) = 1-0.1353\left[1+2+\frac{\cancel{4}^{2}}{\cancel{2}}+\frac{\cancel{8}^{4}}{\cancel{6}_{3}}\right]$$

$$P(x>3) = 1-0.1353\left[1+2+2+\frac{4}{3}\right]^{-3}$$

$$P(x>3) = 1-(0.1353)\left(\frac{19}{3}\right) \quad ..\text{on simplification}$$

$$= 0.1431 \quad \text{on classification}$$

No. of drivers who meet with an accident more than 3.

$$= 0.1431\times 5000 \quad \because \text{total no. of taxi} = 5000$$

$$= 715.5$$

$$\approx 716$$

No. of drivers = 716 ✓ ...Ans.

Ex. 3.2.13 : In a factory of rax or blades, there is a chance of $\frac{1}{500}$ for any blade to be defective blades are supplied in packets of 10. By using Poisson's distribution calculate approximate number of packets containing no defective and two defective blades in 10,000 packets.

Soln. :

Given, chance of defective blades is $\frac{1}{500}$

$$p = \frac{1}{500} \quad \text{Probability of defective blades}$$

$$\text{Here } n = 10 \quad \because \text{In packets of 10}$$

$$\text{mean} = m = np$$

$$= 10\left(\frac{\cancel{1}}{\cancel{500}_{50}}\right) = \frac{1}{50}$$

$$m = 0.02$$

We know, by poisons distribution probability of r success is

$$p(x=r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

(i) No blade is defective

Probability of no defective (i.e. r = 0) is

$$P(x=0) = \underbrace{\frac{m^0e^{-m}}{0!}}_{e^{-m}} \quad \text{In Equation (1), put } r=0$$

$$\because m^0=0!=1$$

$$p(x=0) = e^{-m}=e^{-0.02} \quad \because m=0.02$$

$$= 0.0196 \quad \because e^{-0.02}=0.0196$$

There are 10,000 packets

$$= 0.0196\times 10{,}000$$

$$= 196.0397 \approx 196$$

(ii) Probability of two defective

Probability two defective blades (i.e. r = 2) is

$$P(x=2) = \frac{m^0e^{-m}}{2!} \quad \text{In Equation (1), with } r=2$$

$$= \frac{(0.02)^2 e^{-0.02}}{2} \qquad \because m = 0.02$$

$$P(x=2) = 0.002 \qquad \text{...on calculation}$$

There are 10,000 packets.

$$= 0.0002 \times 10{,}000$$
$$= 1.9604$$
$$\approx 2 \checkmark \qquad \text{...Ans.}$$

Ex. 3.2.14 : An insurance company found that only 0.01 % of the population is involved in certain type of accident each year. If its 1000 policy holders were randomly selected from the population what is the probability that not more than two of its clients are involved in such an accident next year ? [Given $e^{-0.1} = 0.9048$]

Soln. :

Given, 0.01 % of the population is involved in certain type of accident each year.

$$P = \frac{0.01}{100} = 10^{-4}$$

Here, $n = 1000$ Total policy holder

We know, by Poisson's distribution

Probability of r success is

$$P(r) = \frac{m^r e^{-m}}{r!} \qquad \text{...(1)}$$

Where mean $= m = np$

$$= 1000 \times \underbrace{\frac{1}{10{,}000}}_{10^{-4}}$$

$$\therefore m = 0.1$$

Probability that not more than two (i.e. $r \le 2$) ($r = 0, 1, 2$) of its certain clients are involved in accident is

$$P(x \le 2) = \underbrace{P(x=0)}_{\frac{e^m m^0}{0!}} + \underbrace{P(x=1)}_{\frac{e^m m^1}{1!}} + \underbrace{P(x=2)}_{\frac{e^m m^2}{2!}}$$

$$P(x \le 2) = \underbrace{\frac{e^{-m} m^0}{0!}}_{\frac{e^{-m}(1)}{(1)}} + \underbrace{\frac{e^{-m} m^1}{1!}}_{\frac{e^{-m} m}{1}} + \underbrace{\frac{e^{-m} m^2}{2!}}_{\frac{e^{-m} m^2}{2}} \quad \text{By Equation (1)} \quad \because r = 0, 1, 2$$

$$P(x \le 2) = e^{-m} + e^{-m} m + \frac{e^{-m} m^2}{2}$$

$$= e^{-m}\left[1 + m + \frac{m^2}{2}\right] \quad \text{Taking } e^{-m} \text{ common}$$

$$P(x \le 2) = e^{-0.1}\left[1 + 0.1 + \frac{(0.1)^2}{2}\right]$$

$$= \underbrace{e^{-0.1}}_{0.9048} \underbrace{[1 + 0.1 + 0.05]}_{1.105}$$

$$P(x \le 2) = 0.9048 \times 1.105$$

$$= 0.9998$$

$$P(x \le 2) = 0.9998 \checkmark \qquad \text{...Ans.}$$

Ex. 3.2.15 : A manufacturer of cotter pins know that 3% of his product is defective. If he sells cotter pins in box of 100 pins and guarantees than not more than 6 pins will be defective in a box. Find the probability that a box will fail to meet the guaranteed quality.

Soln. :

Given 3% of cotter pins are defective

Probability of defective cotter pine $= p = \frac{3}{100} = 0.03$

Total no. of pins in a box $= n = 100$

We know, by Poisson's distribution, probability of r success is

$$P(x=r) = \frac{e^{-m} m^n}{r!} \qquad \text{...(1)}$$

Where mean $= m = np = 100 \times 0.03 = 3$

$$m = 3$$

It is guarantees that not more than 6 pins are defective.

To fail guarantees means, the probability that a box will contains more than 6 pins (i.e. $r > 6$) are defective is

$$P(x > 6) = P(x=6) + P(x=7) + P(x=8)$$
$$= 1 - [P(x=0) + P(x=1) + P(x=2)\, P(x=3) + P(x=4) + P(x=5) + P(x=6)]$$

$\because$ Total Probability = 1

$$P(x>6) = 1 - \left[\frac{e^m m^0}{0!} + \frac{e^{-m} m^1}{1!} + \frac{e^{-m} m^2}{2!} + \frac{e^{-m} m^3}{3!} + \frac{e^{-m} m^4}{4!} + \frac{e^{-m} m^5}{5!} + \frac{e^{-m} m^6}{6!}\right]$$

By Equation (1) and $r = 0, 1, 2, 3, 4, 5, 6$

Taking e^{-m} common

We get,

$$p(x>6) = 1 - e^{-m}\left[\underbrace{\frac{m^0}{0!}}_{1} + \underbrace{\frac{m^1}{1!}}_{m} + \frac{m^2}{2!} + \frac{m^3}{3!} + \frac{m^4}{4!} + \frac{m^5}{5!} + \frac{m^6}{6!}\right]$$

$$p(x>6) = 1 - e^{-m}\left[1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \frac{m^4}{4!} + \frac{m^5}{5!} + \frac{m^6}{6!}\right]$$

$$p(x>6) = 1 - e^{-3}\left[1 + 3 + \frac{(3)^2}{2!} + \frac{(3)^3}{3!} + \frac{(3)^4}{4!} + \frac{(3)^5}{5!} + \frac{(3)^6}{6!}\right]$$

$$p(x>6) = 1 - \underbrace{e^{-3}}_{0.0497}\left[1 + 3 + \frac{9}{2} + \frac{27}{6} + \frac{81}{24} + \frac{243}{120} + \frac{729}{720}\right] \qquad \text{on simplification}$$

$$= \underbrace{1 - (0.0497)(19.4125)}_{0.9965} \qquad \text{on calculation}$$

$$p(x>6) = 1 - 0.9965 = 0.0035$$

$$\therefore p(x>6) = 0.0035 \checkmark \qquad \text{...Ans.}$$

Ex. 3.2.16 : In a random variable has a Poisson's distribution such that p (1) = p (2) Find (i) Mean of the distribution (ii) p (4)

Soln. : We know By Poisson's distribution probability of r success is

$$P(x=r) = \frac{m^r e^{-m}}{r!} \qquad \text{...(1)}$$

Where mean $= m = np$

(i) Given p (1) = P (2)

i.e. $\underbrace{P(x=1)}_{\frac{m e^{-m}}{1!}} = \underbrace{P(x=2)}_{\frac{m^2 e^{-m}}{2!}}$ by Equation (1)

$$\frac{\cancel{m}\cancel{e^{-m}}}{1!} = \frac{\cancel{m^2}\cancel{e^{-m}}}{2!} \qquad \because r = 1, 2$$

$$\Rightarrow \underbrace{\frac{1}{1!}}_{1} = \underbrace{\frac{m}{2!}}_{\frac{m}{2 \times 1}} \qquad \text{...on simplification}$$

$$1 = \frac{m}{2} \Rightarrow m = 2$$

$$\text{Mean} = m = 2 \checkmark \qquad \text{...Ans.}$$

(ii) p (4)

Now,

$$P(4) = P(x=4) = \frac{m^4 e^{-m}}{4!} \qquad \text{by Equation (1) and } r = 4$$

$$= \frac{(2)^4 e^{-2}}{4 \times 3 \times 2 \times 1}$$

$$= \frac{\overset{4}{\cancel{32}}\ (0.1353)}{\cancel{4} \times 3 \times \cancel{2} \times 1}$$

$$P(x=4) = \frac{4 \times (0.1353)}{3} = 0.09 \quad \text{on Calculation}$$

$$\therefore P(4) = 0.09 \checkmark \qquad \text{...Ans.}$$

Ex. 3.2.17 : In a poisson distribution, if p (r =1) = 2p (r = 2), find p (r = 3)

Soln. : We know, by poisson distribution,

Probability of r success is,

$$P(x=r) = \frac{m^r e^{-m}}{r!} \qquad \text{...(1)}$$

Where m = mean = np

Given : $\underbrace{p(x=1)}_{\frac{m^1 e^{-m}}{1!}} = \underbrace{2p(x=2)}_{\frac{2m^2 e^{-m}}{2!}}$ by Equation (1) and r = 1, 2

$$\frac{\cancel{m}^1 \cancel{e^{-m}}}{1!} = \frac{2m^{\cancel{2}} \cancel{e^{-m}}}{2!} \qquad \text{on simplification}$$

$$\underbrace{\frac{1}{1!}}_{1} = \underbrace{\frac{\cancel{2}m}{\cancel{2} \times 1}}_{m} \qquad \because 2! = 2$$

$$1 = m \Rightarrow m = 1$$

$$\text{mean} = m = 1$$

$$P(r=3) = p(x=3) = \frac{m^3 e^{-m}}{3!} \qquad \text{by Equation(1) } r = 3$$

$$P(x=3) = \frac{(1)^3 e^{-1}}{3!}$$

$$= \frac{1 \times (0.3678)}{3 \times 2 \times 1}$$

$$= 0.0613$$

$$\therefore P(x=3) = 0.0613 \checkmark \qquad \text{...Ans.}$$

Ex. 3.2.18 : If a random variable has a Poisson's distribution such that P (2) = p (30 , find p (5)

Soln. : We know, by Poisson's distribution,

$$P(x=r) = \frac{m^r e^{-m}}{r!} \qquad \text{...(1)}$$

where m = mean = np

Given $P(2) = p(3)$

$$\underbrace{P(x=2)}_{\frac{m^2 e^{-m}}{2!}} = \underbrace{p(x=3)}_{\frac{m^3 e^{-m}}{3!}} \qquad \text{by Equation (1)}$$

$$\frac{\cancel{m^2}\cancel{e^{-m}}}{2!} = \frac{m^{\cancel{3}}\cancel{e^{-m}}}{3!}$$

$$\underbrace{\frac{1}{2!}}_{\frac{1}{2 \times 1}} = \underbrace{\frac{m}{3!}}_{\frac{m}{3 \times 2 \times 1}} \qquad \text{on simplification} \qquad \because n! = n(n-1)(n-2)\ldots$$

$$\frac{1}{2} = \frac{m}{6} \qquad \text{m simplification}$$

$$2m = 6 \qquad \text{on cross multiplication}$$

$$m = \frac{6}{2} = 3 \qquad \text{i.e. } m = 3$$

$$\text{Now, } p(5) = p(x=5) \frac{m^5 e^{-m}}{5!}$$

by Equation (1) and r = 5

$$P(x=5) = \underbrace{\frac{(3)^5 e^{-3}}{5!}}_{\frac{243 \times 0.0498}{120}} \qquad \left\{\begin{array}{l} \because 5! = 120 \\ e^{-3} = 0.0498 \end{array}\right.$$

$$p(x=5) = 0.1008$$

$$p(5) = 0.1008 \checkmark \qquad \text{...Ans.}$$

Fitting of Poisson's Distribution :

Ex. 3.2.19 : Fit a Poisson distribution for following observation :

x_1	20	30	40	50	60	70
f_1	8	12	30	10	6	4

Soln.: We know, mean $= \dfrac{\sum f_i X_i}{\sum f_i}$ ...(1)

X_i	f_i	$f_i X_i$
20	8	160
30	12	360
40	30	1200
50	10	500
60	6	360
70	4	280
$\sum$	70	2860
	$\sum f_i$	$\sum f_i X_i$

$f_i x_i$ is multiplication of f_i and r_i values

Substitute values $\sum f_i X_i = 2860$ and $\sum f_i = 70$ in Equation (1), we get,

$$\text{mean} = m = \frac{2860}{70} = 40.8571$$

$$\text{mean} = m = 40.8571$$

We know, by Poisson's Distribution

Probability of r success is,

$$P(X = r) = \frac{m^r e^{-m}}{r!}$$

$$P(X = r) = \frac{(40.8571)^r e^{-40.8571}}{r!}$$

Where r = 20, 30, 40, 50, 60, 70 and m = 40.8751

This is required Poisson's Distribution. ✓ ...Ans.

Ex. 3.2.20: Fit a Poisson's distribution to set of observations

x_i	0	1	2	3	4
f_i	122	60	15	2	1

Soln.: We know,

We know, Mean $= \dfrac{\sum f_i x_i}{\sum f_i}$ ...(1)

x_i	f_i	$f_i x_i$
0	122	0
1	60	60
2	15	30
3	2	6
4	1	4
Σ	200	100

$\because f_i x_i$ is multiplication of f_i and x_i values

$$\therefore \text{mean} = m = \frac{100}{200} \quad \text{By Equation (1)}$$

$$\therefore m = \frac{1}{2} = 0.5 = 0.5$$

We know by poisson's distribution probability of r success is,

$$P(x = r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

$$p(x = r) = \frac{(0.5)^r e^{-0.5}}{r!} \checkmark \quad ...\text{Ans.}$$

Where r = 0, 1, 2, 3, 4 and m = 0.5

This is required Poisson's distribution

Ex. 3.2.21: A skilled typist, on routine work, kept a record of mistakes per day during 300 working days. Fit a Poisson's distribution to the set of observations.

x	0	1	2	3	4	5	6
y	143	90	42	12	9	3	1

Soln.: We know, Poisson's distribution

$$P(x = r) = \frac{m^r e^{-m}}{r!} \quad ...(1)$$

Where m = mean = np

Given

x (mistake) (xi)	0	1	2	3	4	5	6	M
y (days) (fi)	143	90	42	12	9	3	1	350
xy = fi xi	0	90	84	36	36	15	6	253

$$\therefore \text{Mean} = \frac{\sum f_i x_i}{\sum f_i}$$

$$\text{Mean} = \underbrace{\frac{253}{300}}_{} \quad (\because \text{mean} = np)$$

$m = 0.8433$ On simplification

Mean of distribution = m = 0.8433

We know, by Poisson's distribution probability of r success is

$$P(x = r) = \frac{m^r e^{-m}}{r!}$$

$$P(x = r) = \frac{(0.8433)^r e^{-0.8433}}{r!} \quad \text{where } r = 0,1,2,3,4,5,6$$

This is required Poisson's distribution. ✓ ...Ans.

Ex. 3.2.22: The accidents per shift in a factory are given by the table :

Accident (x) per shift	0	1	2	3	4	5
Frequency (f)	142	158	67	27	5	1

Fit a Poisson distribution to the above table and calculate the frequencies as per Poisson fit.

Soln.:

Step I: According to Poisson's distribution,

$$P(r) = \frac{e^{-M} \cdot M^r}{r!}$$

Step II: Here, M = np = mean and $\begin{cases} N = \sum f_i = 400 \\ n = 6 \end{cases}$

Step III:

x_i	f_i	$\sum f_i x_i$
0	142	0
1	158	158
2	67	134
3	27	81
4	5	20
5	1	5
	$\sum f_i = 400$	$\sum f_i x_i = 398$

$$\text{Mean} = M = \frac{\sum f_i x_i}{N} = \frac{398}{400} = 0.995$$

Step IV: $P(r) = \dfrac{e^{-0.995}(0.995)^r}{r!}$; r = 0, 1, 2, 3, 4/5

$$\therefore P(r = 0) = \frac{e^{-0.995} \times (0.995)^0}{0!} = 0.3698$$

$$P(1) = \frac{e^{-0.995} \times (0.995)^1}{1!} = 0.3679$$

$$P(2) = \frac{e^{-0.995} \times (0.995)^2}{2!} = 0.1830$$

$$P(3) = \frac{e^{-0.995} \times (0.995)^3}{3!} = 0.0607$$

$$P(4) = \frac{e^{-0.995} \times (0.995)^4}{4!} = 0.0151$$

$$P(5) = \frac{e^{-0.995} \times (0.995)^5}{5!} = 0.003$$

Step V: Hence, the Theoretical frequency distribution is,

$$P(0) \times \sum f_i = (0.3698) \times 400 = 147.92$$

$$P(1) \times \sum f_i = (0.3679) \times 400 = 147.16$$

$$P(2) \times \sum f_i = (0.183) \times 400 = 73.2$$

$$P(3) \times \sum f_i = (0.0607) \times 400 = 24.28$$

$$P(4) \times \sum f_i = (0.0151) \times 400 = 6.04$$

$$P(5) \times \sum f_i = (0.003) \times 400 = 1.2$$

Syllabus Topic : Normal Distribution

3.3 Normal Distribution

Q. Explain normal distribution in detail.

- The normal distribution is a continuous distribution.
- The general equation of the normal distribution is given by,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Where the variable x can assume all values from $-\infty$ to $+\infty$.

- μ and σ are called the parameters of the distributions, are respectively the mean and the standard deviation of the distribution and $-\infty < \mu < \infty$, $\sigma > 0$. X is called normal variate.
- f (x) is called probability density function of the normal distribution. The graph of the normal distribution is called the normal curve.
- It is bell-shaped and symmetrical about the mean μ as shown in Fig. 3.3.1

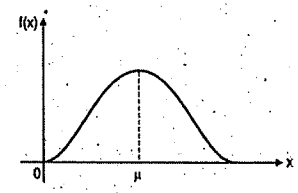


Fig. 3.3.1

- The end points of the curve extent to $+\infty$ and $-\infty$ towards the positive and negative directions of the X-axis respectively.
- The line $X = \mu$ divides the area under the normal curve above X-axis into two equal parts. The median of the distribution also coincides with mean and mode.
- The area under the normal curve between any two given ordinates $x = x_1$ and $x = x_2$ represent the probability of values falling into the given interval.

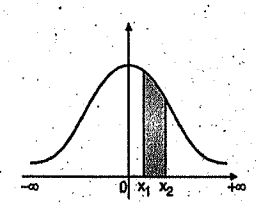


Fig. 3.3.2

- The total area under the normal curve above the X-axis divided into two parts as shown in Fig. 3.3.3.
- Left hand side area and right hand side area to z = 0 is 0.5.

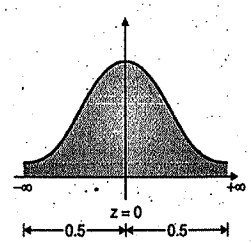


Fig. 3.3.3

3.3.1 Standard Form of the Normal Distribution

Q. Explain and give Standard Form of the Normal Distribution.

- If X is a normal random variable with mean μ and standard deviation σ, then the random variable.

$$z = \frac{X-\mu}{\sigma}$$

has the normal distribution with mean 0 and standard deviation 1.

- The random variable z is called the standardized (or standard) normal random variable.
- The probability density function for the normal distribution in standard form is given by,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

- This help us to find area under the normal probability curve by making use of standard table.
- If f(z) is the probability density function for the normal distribution then,

$$P(z_1 \le Z \le z_2) = \int_{z_1}^{z_2} f(z)\,dx = F(z_2) - F(z_1)$$

$$\text{Where } F(z) = \int_{-\infty}^{\infty} f(z)\,dz = P(Z \le 2)$$

- The function F (z) defined above is called the distribution function for the normal distribution.

3.3.2 Steps to find Probability with different z

Q. Explain Steps to find Probability with different z.

(I) To find,

$$P(z \ge z_1) = P(0 \le z \le \infty) - P(0 \le z \le z_1)$$
$$= 0.5 - P(0 \le z \le z_1)$$

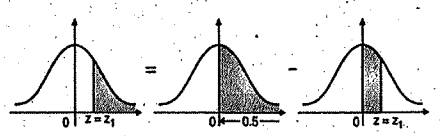


Fig. 3.3.4

(II) $P(-z_1 \le z \le z_2) = P(-z_1 \le z \le 0) + P(0 \le z \le z_2)$

$= P(0 \le z \le z_1) + P(0 \le z \le z_2)$

$P(z \le -z_1) = P(z \ge z_1)$

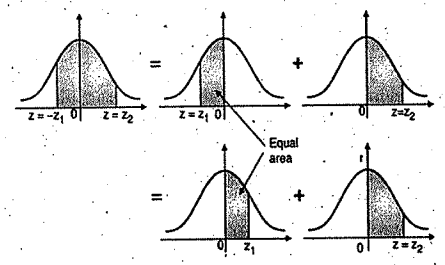


Fig. 3.3.5

(III) $P(z \le -z_1) = P(z \ge z_1) = 0.5 - P(0 \le z \le z_1)$

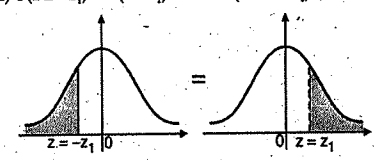


Fig. 3.3.6

(IV) $P(z \le z_1) = P(-\infty \le z \le 0) + P(0 \le z \le z_1)$

$= 0.5 + P(0 \le z \le z_1)$

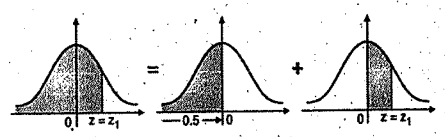


Fig. 3.3.7

Note : Use standard table for Area under the normal curve.

Table 3.3.1 : Area under the normal curve

z	0	1	2	3	4	5	6	7	8	9
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	0.0754
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	0.11591	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	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	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	0.2996	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	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
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
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993

3.3.1 Examples on Normal Distribution

Ex 3.3.1 : Students of a class were given an aptitude test. Their marks were found to be normally distributed with mean 60 and standard deviation 5. What is the percentage of student scored more than 60 marks ?

Soln. :

Given : Mean $= \mu = 60$; Standard deviation $= \sigma = 5$

Percentage of student scored more than 60 :

For $X = 60$,

$$Z = \frac{60-60}{5}$$

$$Z = 0$$

$$\therefore P(X \ge 60) = P(z \ge 0)$$
$$= P(0 \le z \le \infty)$$
$$= 0.5$$

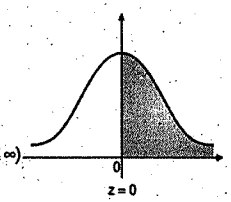


Fig. P. 3.3.1

$\therefore$ **Percentage of student scored more than 60 = 50% ✓**

... Ans.

Ex. 3.3.2 : In a test on 2000 electric bulbs, it was found that the life of particular make was normally distributed with average life of 2040 hours and standard deviation of 60 hours. Estimate the number of bulbs likely to burn for :

(a) between 1920 hours and 2160 hours

(b) more than 2150 hours

[Given that A(2) = 0.4772, A(1.83) = 0.4664]

Soln. : Given, average life is 2040 hours

$$\text{Mean} = \mu = 2040$$

and standard deviation $= \sigma = 60$ hours

For the standard normal curve,

$$z = \frac{X-\mu}{\sigma}$$

$$z = \frac{X-2040}{60} \quad \ldots(1)$$

Total electric bulbs $= n = 2000$

(a) Between 1920 hours and 2160 hours

For $X = 1920$

$$z = \frac{1920-2040}{60} \quad \text{(by Equation (1))}$$

$$= \frac{-120}{60} = -2$$

For $X = 2160$, $\quad z = \frac{2160-2040}{60}$

$$= \frac{120}{60} = 2$$

$\therefore P(1920 \le X \le 2160) = P(-2 \le z \le 2)$

$= P(-2 \le z \le 0) + P(0 \le z \le 2)$

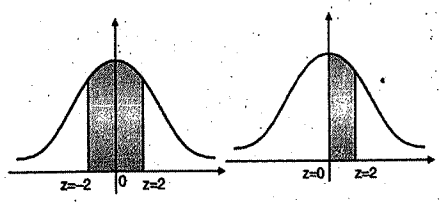
Note : Area from $z = -2$ to $z = 2$ can divide into area from $z = -2$ to $z = 0$ and $z = 0$ to $z = 2$. Also note area $z = -2$ to $z = 0$ is equal to $z = 0$ to $z = 2$ as shown in Figs. 9.12.2(a), (b), (c) and (d)

$= P(0 \le z \le 2) + P(0 \le z \le 2)$

$= 2P(0 \le z \le 2)$

$= 2 \times 0.4772$ [Given A(2) = 0.4772]

$p(1920 \le X \le 2160) = 0.9544$



(a) (b)

Fig. P. 3.3.2

$\therefore$ Number of bulbs likely to burn between 1920 hours to 2160 hours.

$= 0.9544 \times 2000 = 1908.8$

≈ 1909

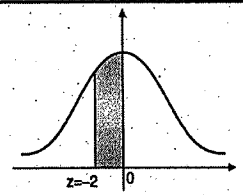


Fig. P. 3.3.3.2(c)

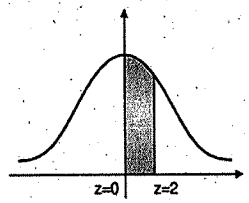


Fig. P. 3.3.2(d)

$\therefore$ **Required number of bulbs = 1909 ✓** ...Ans.

(b) More than 2150 hours

For $X = 2150$,

$$z = \frac{2150-2040}{60} = \frac{110}{60} = 1.8333$$

$\therefore P(X \ge 2150) = P(z \ge 1.83)$

$= 0.5 - P(0 \le z \le 1.83)$

Note : Since area between $z = 0$ to $z = \infty$ is 0.5

$\therefore$ Area for $z \ge 1.83$ is equal to 0.5 – area $0 \le z \le 1.83$

$= 0.5 - 0.4664$

$= 0.0336$ ($\because$ Given A (1.83) = 0.4664)

$\therefore$ Number of bulbs likely to burn more than 2150 hours

$= 0.0336 \times 2000$ ($\because$ Total bulbs = 2000)

$= 67.2 \approx 67$

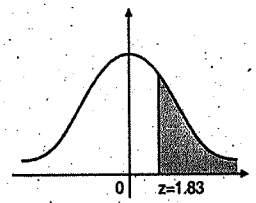


Fig. P. 3.3.2(e)

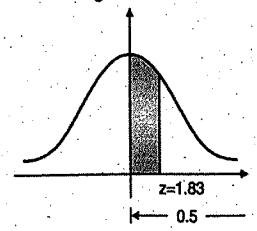


Fig. P. 3.3.2(f)

$\therefore$ **Required number of bulbs = 67 ✓** ...Ans.

Ex. 3.3.3 : In a sample of 1000 cases the mean of a certain test is 14 and standard deviation is 2.5.

Assuming the distribution to be normal.

Find.

1. How many students score between 12 and 15 ?

2. How many students score above 18 ?

[Given : A (0.8) = 0.2881, A (0.4) = 0.1554, A (1.6) = 0.4452]

Soln. :

Given : $n = 1000$ [Total sample cases of a test]

Mena $= \mu = 14$, Standard deviation $= \sigma = 2.5$

For the standard normal curve.

$$z = \frac{X-\mu}{\sigma} = \frac{X-14}{2.5} \quad \ldots(1)$$

(a) Student score between 12 and 15

For $X = 12$, $\quad z = \frac{12-14}{2.5} = \frac{-2}{2.5} = -0.8$

For $X = 15$,

$$z = \frac{15-14}{2.5} \quad \text{(by Equation (1))}$$

$$= \frac{1}{2.5} = 0.4$$

$\therefore P(12 \le X \le 15) = P(-0.8 \le z \le 0.4)$

Note : Area $-0.8 \le z \le 0.4$ is equal to area of $-0.8 \le z \le 0$ and $0 \le z \le 0.4$. Since area of $-0.8 \le z \le 0$ is equal to area of $0 \le z \le 0.8$

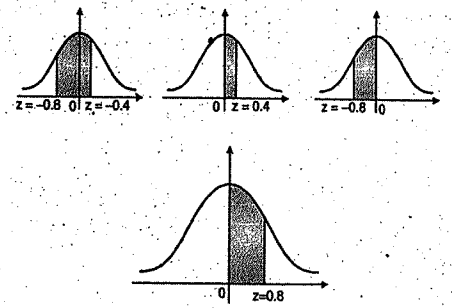


Fig. P. 3.3.3(a)

$P(12 \le X \le 15) = P(-0.8 \le z \le 0.4)$

$= P(-0.8 \le z \le 0) + P(0 \le z \le 0.4)$

$= P(0 \le z \le 0.8) + P(0 \le z \le 0.4)$

$= 0.2881 + 0.1554$

[$\because$ Given A(0.8) = 0.2881 and A (0.4) = 0.1554]

$P(12 \le X \le 15) = 0.4435$

$\therefore$ Total number of students scored between 12 and 15.

$= 0.4435 \times 1000 = 443.5$

≈ 444

$\therefore$ **Total number of student scored between 12 and 15 = 444 ✓** ...Ans.

(b) Student score above 18

For $X = 18$, $\quad z = \frac{18-14}{2.5} = \frac{4}{2.5} = 1.6$

Note : Since area for $z = 0$ to $z = \infty$ is 0.5

$\therefore$ Area for $z \ge 1.6$ is equal to 0.5 – area for $0 \le z \le 1.6$

$\therefore P(X \ge 18) = P(z \ge 1.6) = 0.5 - P(0 \le z \le 1.6)$

$= 0.5 - 0.4452$ (Given A (1.6) = 0.4452)

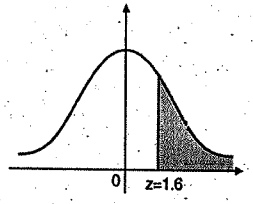
$P(X \ge 18) = 0.0548$

$\therefore$ Total number of students scored above 18

$= 0.548 \times 1000 = 54.8$

≈ 55

$\therefore$ **Total number of student scored above 18 = 55 ✓** ...Ans.



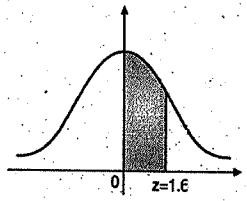


Fig. 3.3.3(b)

Ex. 3.3.4 : I. Q.'s are normally distributed with mean 100 and standard deviation 15. Find the probability that a randomly selected person has :

(i) An I.Q. more than 130

(ii) An I.Q. between 5 and 115.

[z = 2, Area = 0.4772, z = 1, Area = 0.3413]

Soln. :

Given : I · Q's are normally distributed with mean 100.

$\therefore$ mean $= \mu = 100$ and

Standard deviation $= \sigma = 15$

For the standard normal curve,

$$z = \frac{X-\mu}{\sigma} \quad \text{where P(z) is the area under the curve}$$

$$z = \frac{X-100}{15}$$

Probability that a randomly selected person has

(i) An I.Q. more than 130 means $X > 130$

$\therefore$ Probability of I.Q is more than 130 is, $P(X > 130)$

$\therefore$ $P(X > 130)$

$$= P\left(z > \underbrace{\frac{130-\mu}{\sigma}}_{\frac{X-\mu}{\sigma}}\right) = P\left(z > \underbrace{\frac{130-100}{15}}_{\frac{X-\mu}{\sigma}}\right) = p\left(z > \frac{30}{15}\right)$$

$$= P(z > 2)$$

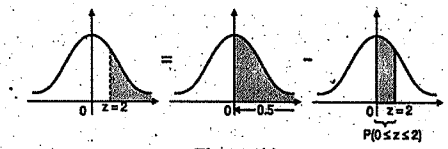


Fig. 3.3.4(a)

$$= 0.5 - P(0 \le z \le 2)$$

$$= 0.5 - 0.4772 \quad \text{(Given } z = 2, \text{ Area} = 0.4772)$$

$$P(X > 130) = 0.0228$$

i.e. Probability of randomly selected person has I.Q. more than 130 is 0.0228. ✓ ...Ans.

(ii) An I.Q. between 85 and 115

Probability of I.Q's between 85 and 115 is $P(85 \le X \le 115)$

$$\therefore P(85 \le X \le 115) = P\left(\underbrace{\frac{85-\mu}{\sigma}}_{\frac{X-\mu}{\sigma}} \le z \le \underbrace{\frac{115-\mu}{\sigma}}_{\frac{X-\mu}{\sigma}}\right)$$

$$= P\left(\frac{85-100}{15} \le z \le \frac{115-100}{15}\right)$$

$$= P\left(\frac{-15}{15} \le z \le \frac{15}{15}\right) = P(-1 \le z \le 1)$$

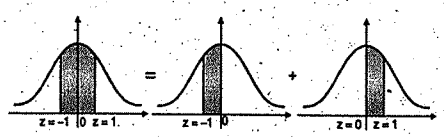


Fig. 3.3.4(b)

$$= P(-1 \le z \le 0) + P(0 \le z \le 1)$$

$$= P(0 \le z \le 1) + P(0 \le z \le 1)$$

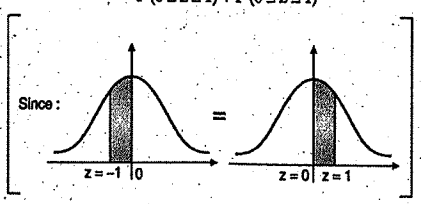


Fig. 3.3.4(c)

$$= 2 \times P(0 \le z \le 1) = 2 \times (0.3413)$$

[$\because$ Give : $z = 1$, Area = 0.3413]

$P(85 \le X \le 115) = 0.6826$

i.e. Probability of randomly selected person has an I.Q. between 5 and 115 is 0.6826. ✓ ...Ans.

Ex. 3.3.5 : In a certain examination 500 students appeared average score is 68 with S.D. 8. Find the number of students scoring

(a) less than 50 (b) more than 60.

[Given A (2.25) = 0.4878, A (1) = 0.3413]

Soln. :

Given, Total students = n = 500, Mean score = $\mu = 68$, standard deviation = $\sigma = 8$

For the standard normal curve,

$$z = \frac{X-\mu}{\sigma}, \qquad z = \frac{X-68}{8} \qquad ...(1)$$

(a) Less than 50 : [$X \le 50$]

$$\text{For } X = 50, \quad z = \frac{50-68}{8} \text{ (by Equation (1))}$$

$$= \frac{-18}{8} = -2.25$$

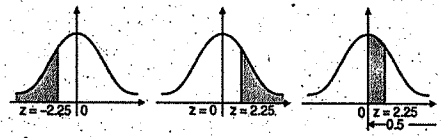


Fig. P.3.3.5(a)

Area for $z \le -2.25$ is equal to area for $z \ge 2.25$

Area for $z \ge 2.25$ is equal to 0.5 – area for $0 \le z \le 2.25$ as shown in Fig. P. 3.3.5(a).

$$\therefore P(X \le 50) = P(z \le -2.25) = P(z \ge 2.25)$$

$$= 0.5 - P(0 \le z \le 2.25)$$

$$= 0.5 - 0.4878 \quad \text{[Given A (2.25) = 0.4878]}$$

$$= 0.0122$$

$\therefore$ Total number of students scoring less than 50

$$= 0.0122 \times 500 = 6.1 \approx 6$$

$\therefore$ **Total number of students scoring less than 50 = 6**

(b) More than 60 : [$X \ge 60$]

$$\text{For } X = 60, \quad z = \frac{60-68}{8} = \frac{-8}{8} = -1$$

$$\therefore P(X \ge 60) = P(z \ge -1)$$

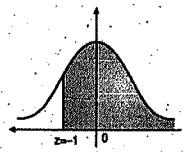


Fig. P. 3.3.5(b)

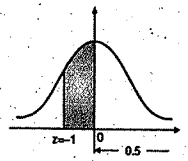


Fig. P. 3.3.5(c)

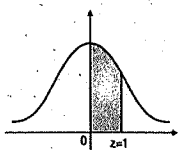


Fig. P. 3.3.5(d)

Area for $z \ge -1$ is equal to area for $-1 \le z \le 0$ and $0 \le z \le \infty$

Since Area for $-1 \le z \le 0$ is equal to $0 \le z \le 1$

$\therefore$ Area for $z \ge -1$ is equal to 0.5 + area for $0 \le z \le 1$ as shown in Fig. P. 3.3.5

$$P(X \ge 60) = P(z \ge -1) = P(-1 \le z \ge 0) + P(0 \le z \le \infty)$$

$$= P(-1 \le z \le 0) + 0.5 = P(0 \le z \le 1) + 0.5$$

$$= 0.3413 + 0.5 \qquad \text{[Given A(1) = 0.3413]}$$

$$P(X \ge 60) = 0.8413$$

$\therefore$ Total number of students scoring more than 60

$$= 0.8413 \times 500 = 420.65 \approx 421$$

$\therefore$ **Total number of students scoring less than 60 = 421 ✓**

...Ans.

Ex. 3.3.6 : Sacks of sugar packed by an atomic loader having an average weight of 100 kg with standard deviation 0.250 kg. Assuming normal distribution find chance of sack get weighing less than 99.5 kg. [Given : A(2) = 0.4772]

Soln. : Given, Average weight of sacks = μ = 100 kg

Standard deviation = σ = 0.250 kg

For the standard normal curve,

$$Z = \frac{X-\mu}{\sigma} = \frac{X-100}{0.250} \qquad ...(1)$$

$$\text{For } X = 99.5, \quad z = \frac{99.5-100}{0.250} \qquad \text{(by Equation (1))}$$

$$= \frac{-0.5}{0.250} = -2$$

The chance of sacks weighing less than 99.5 kg is,

Area for $z \le -2$ is equal to area for $z \ge 2$

Since area for $0 \le z \le \infty$ is 0.5

$\therefore$ Area for $z \ge 2$ is equal to 0.5 – area for $0 \le z \le 2$ as shown in Fig. P. 3.3.6.

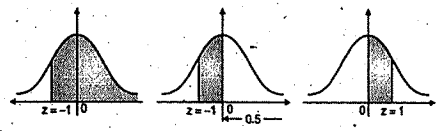


Fig. 3.3.6(a)

$$P(X \le 99.5) = P(z \le -2) = P(z \ge 2) = 0.5 - P(0 \le z \le 2)$$

$$= 0.5 - 0.4772 \quad \text{[Given A (2) = 0.4772]}$$

$$P(X \le 99.5) = 0.0228$$

$\therefore$ **The chance of sacks weighing less than 99.5 kg = 0.0228✓ ...Ans.**

Ex. 3.3.7 : Heights of students follow normal distribution with mean 190 cm and variance 70 cm^2. In a school of 800 students how many are expected to have heights above 200 cm.

($z = 1.195$, A = 0.3840)

Soln. :

Step I : Given, Mean = μ = 190

$$\sigma^2 = 70 \Rightarrow \sigma = \sqrt{70}$$

$$x = 200$$

Step II : Let, $z = \frac{x-\mu}{\sigma} = \frac{200-190}{\sqrt{70}} = \frac{10}{\sqrt{70}} = 1.195$

Also given A (z = 1.195) = 0.3840

Step III : The probability of expected to have height above 200 cm is,

$$P(x > 200) = P(z > 1.195)$$

$$= 0.5 - A(1.195) = 0.5 - 0.3840$$

$$= 0.116$$

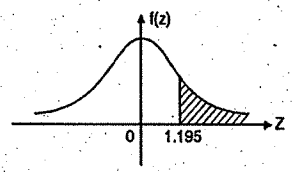


Fig. P. 3.3.7

Step IV : Expected number of students,

$$= 800 \times 0.116 = 92.8$$

$$\cong 93 \text{ students}$$

Ex. 3.3.8 : In a certain examination test, 200 students appeared in a subject of statistics. Average marks obtained were 50% with standard derivation 5%. How many students do you expect to obtain more than 60% of marks, Supposing that marks are distributed normally ?

(Given : Area corresponding to x = 0.6) = 0.4772.

Soln. :

Step I : Given : $\mu = 50\% = \frac{50}{100} = 0.5$

$$\sigma = 5\% = \frac{5}{100} = 0.05$$

$$x = 60\% = \frac{60}{100} = 0.6$$

Step II : Let $z = \frac{x-\mu}{\sigma} = \frac{0.6-0.5}{0.05} = 2$

$$A(z_1) = A(0 < z < z_1)$$

Here $z_1 = 2$

Step III : $A(2) = A(0 < z < 2) = 0.4772$ (Given)

$\therefore A(0 < z < 2) = 0.5 - A(z = 2) = 0.5 - 0.4772 = 0.0228$

Step IV : The expected number of student, getting more than 60% marks

$= 0.0228 \times 200 \approx 46$

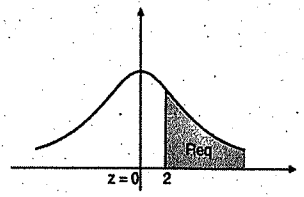


Fig. P. 3.3.8

Ex. 3.3.9 : In a distribution, exactly normal 7% of the items are under 35 and 89% are under 63. Find the mean and standard deviation of the distribution. $A_1 = 0.43$, $z_1 = 1.48$; $A_2 = 0.39$, $z_2 = 1.23$.

Soln. :

Given : 7% of items are under 35 and 89% are under 63

7% of items are under 35 means area under 35 = 0.07

Similarly Area under 63 = 89% = 0.89

Hence, $A(x > 63) = (100 - 89) = 11\% = 0.11$

$\therefore P(x < 35) = 0.07$

$P(x > 63) = 0.11$

$\therefore x_1 = 35$ and $x_2 = 63$

$$z = \frac{x - \mu}{\sigma}$$

$$\therefore z_1 = \frac{x_1 - \mu}{\sigma} \Rightarrow -z_1 = \frac{35 - \mu}{\sigma}$$

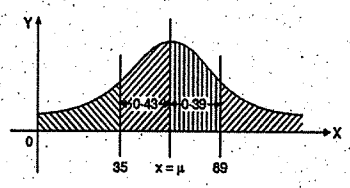


Fig. P. 3.3.9

(Negative sign indicates that x = 35 lies to the left side of mean (μ))

$$\therefore z_2 = \frac{x_1 - \mu}{\sigma} = \frac{63 - \mu}{\sigma}$$

From figure $A(\mu < x < \infty) = 0.5$

$A(x > 63) = 0.11$

$\therefore A(0 < z < z_2) = A(\mu < x < 63) = 0.5 - 0.11 = 0.39$

and $A(-\infty < x < \mu) = 0.5$

$A(x > 35) = 0.07$

$\therefore A(0 < z < z_1) = A(0 < x < 35) = 0.5 - 0.07 = 0.43$

$\therefore$ The point corresponding Area $0.43 \approx 1.48$

And the point corresponding Area $0.39 \approx 1.23$

(See the table value)

$$\therefore -z_1 = \frac{35 - \mu}{\sigma} \Rightarrow -1.48 = \frac{35 - \mu}{\sigma}$$

$$\text{and} \quad z_2 = \frac{63 - \mu}{\sigma} \Rightarrow 1.23 = \frac{63 - \mu}{\sigma}$$

Hence, $35 - \mu = -(1.48)\sigma \Rightarrow (1.48)\sigma - \mu = -35$

and $63 - \mu = (1.23)\sigma \Rightarrow -(1.23)\sigma - \mu = -63$

Solving these two equations simultaneously,

We get, $\sigma = 10.334$ and $\mu = 50.2915$

Hence, Mean $(\mu) = 50.2915$

and Standard deviation $(\sigma) = 10.3321$

Syllabus Topic : Fitting of binomial Distribution

3.4 Fitting of Binomial Distribution

Ex. 3.4.1 : Fit a binomial distribution to the following frequency distribution :

x :	0	1	2	3	4	5	6
f :	13	25	52	58	32	16	4

Soln. :

We know by binomial distribution probability of r success is,

$$P(x = r) = {}^nC_r\, p^r q^{n-r} \qquad \ldots(1)$$

Here $n = 6$ and $N = \Sigma f_i = 200$

x	f	$f_i x_i$
0	13	0
1	25	25
2	52	104
3	58	174
4	32	128
5	16	80
6	4	24
	$\Sigma f_i = 200$	$\Sigma f_i x_i = 535$

$$\text{Since mean} = \mu = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$\mu = \frac{535}{200} = 2.675$$

We know,

Mean $= \mu = np$

$2.675 = 6p$

$$\therefore p = \frac{2.675}{6} = 0.45$$

$q = 1 - p = 1 - 0.45$

$= 0.55$

x	Probability $P(X = r) = {}^nC_r\, p^r q^{n-r}$	Theoretical frequencies N × P
0	${}^6C_0\, p^0 q^6 = {}^6C_0 (0.45)^0 (0.55)^6$ $= (1)(1)(0.0277)$ $= 0.0277$	$200 \times 0.0277 = 5.54$
1	${}^6C_1\, p^1 q^5 = {}^6C_1 (0.45)(0.55)^5$ $= (6)(0.45)(0.0503)$ $= 0.1358$	$200 \times 0.1358 = 27.16$
2	${}^6C_2\, p^2 q^4 = {}^6C_2 (0.45)^2 (0.55)^4$ $= (15)(0.2025)(0.0915)$ $= 0.2779$	$200 \times 0.2779 = 55.58$
3	${}^6C_3\, p^3 q^3 = {}^6C_3 (0.45)^3 (0.55)^2$ $= (20)(0.0911)(0.1664)$ $= 0.3032$	$200 \times 0.3032 = 60.64$
4	${}^6C_4\, p^4 q^2 = {}^6C_4 (0.45)^4 (0.55)^2$ $= (15)(0.0410)(0.3025)$ $= 0.1860$	$200 \times 0.0611 = 12.22$
5	${}^6C_5\, p^5 q^1 = {}^6C_5 (0.45)^5 (0.55)$ $= (6)(0.0185)(0.55)$ $= 0.0611$	$200 \times 0.0611 = 12.22$
6	${}^6C_6\, p^6 q^0 = {}^6C_6 (0.45)^6 (0.55)^0$ $= (1)(0.0083)(1) = 0.0083$	$200 \times 0.0083 = 1.66$

Ex. 3.4.2 : Fit a binomial distribution for the following data and compare the theoretical frequencies with the actual ones :

x :	0	1	2	3	4	5
f :	2	14	20	34	22	8

Soln. :

We know by binomial distribution probability of r success is,

$$P(X = r) = {}^nC_r\, p^r q^{n-r} \qquad \ldots(1)$$

Here $n = 5$ $N = \Sigma f_i = 100$

x	f	$f_i x_i$
0	2	0
1	14	14
2	20	40
3	34	102
4	22	88
5	8	40
Σ	$\Sigma f_i = 100$	$\Sigma f_i x_i = 284$

$$\text{Since mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} \qquad \therefore \mu = \frac{284}{100} = 2.84$$

$\therefore$ Mean $= \mu = 2.84$

Also,

Mean $= \mu = np$ $\qquad \therefore 2.84 = 5 \times p$

$$\therefore p = \frac{2.84}{5} = 0.568$$

$p = 0.57$

$\therefore q = 1 - p = 1 - 0.57$

$q = 0.43$

Probability for x = r = 0

$= {}^5C_0\, p^0 q^5$ $\qquad \therefore$ Using formula (1)

$= 0\, {}^5C_0 (0.57)^0 (0.43)^5$

$= (1)(1)(0.0147)$

$p(X = 0) = 0.0147$

Similarly for x = 1, 2, 3, 4, 5, etc.

The successive terms in the expansion gives theoretical frequencies are :

x	Probability $P(X = r) = {}^nC_r\, p^r q^{n-r}$	Theoretical frequencies N × P Here N = 100
0	${}^5C_0\, p^0 q^5 = {}^5C_0 (0.57)^0 (0.43)^5$ $= (1)(1)(0.0147)$ $= 0.0147$	$100 \times 0.0147 = 1.47$
1	${}^5C_1\, p^1 q^4 = {}^5C_1 (0.57)(0.43)^4$ $= (5)(0.57)(0.0342)$ $= 0.09747$	$100 \times 0.09747 = 9.747 \approx 9.8$
2	${}^5C_2\, p^2 q^3 = {}^5C_2 (0.57)^2 (0.43)^3$ $= (10)(0.3249)(0.0795)$ $= 0.2583$	$100 \times 0.2583 = 25.83$
3	${}^5C_3\, p^3 q^2 = {}^5C_3 (0.57)^3 (0.43)^2$ $= (10)(0.1852)(0.1849)$ $= 0.3424$	$100 \times 0.3424 = 34.24$
4	${}^5C_4\, p^4 q^1 = {}^5C_4 (0.57)^4 (0.43)$ $= (5)(0.1056)(0.43)$ $= 0.2270$	$100 \times 0.2270 = 22.70$
5	${}^5C_5\, p^5 q^0 = {}^5C_5 (0.57)^4 (0.43)^0$ $= (1)(0.0602)(0.43)$ $= 0.0259$	$100 \times 0.0259 = 2.59$

Ex. 3.4.3 : The following data are the number of seeds germinating out of 10 on damp filter paper for 80 sets of seeds. Fit a binomial distribution of these data D :

x :	0	1	2	3	4	5	6	7	8	9	10
f :	6	20	28	12	8	6	0	0	0	0	0

Soln. :

We know by binomial distribution probability of r success is,

$$P(X = r) = {}^nC_r\, p^r q^{n-r} \qquad \ldots(1)$$

Here $n = 10$, and $N = \Sigma f_i = 80$

x	f	$f_i x_i$
0	6	0
1	20	20
2	28	56
3	12	36
4	8	32
5	6	30
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
Σ	$\Sigma f_i = 80$	$\Sigma f_i x_i = 174$

Since mean $= \mu = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$

$$= \frac{174}{80} = 2.175$$

Mean $= \mu = 2.175$

We know,

Mean $= \mu = np$

$\therefore\ 2.175 = 10\,p$

$\therefore\ p = \dfrac{2.175}{10} = 0.2175$

$p = 0.2175$

Also,

$\therefore\ q = 1 - p = 1 - 0.2175$

$= 0.7825$

The successive terms in the expansion gives theoretical frequencies are :

x	Probability $P(X = r) = {}^nC_r\, p^r q^{n-r}$	Theoretical frequencies $N \times P$ Here $N = 80$
0	${}^{10}C_0\, p^0 q^{10} = {}^{10}C_0 (0.2175)^0 (0.7825)^{10}$ $= (1)(1)(0.7825)^{10}$ $= 0.0861$	$80 \times 0.0861 = 6.888$
1	${}^{10}C_1\, p^1 q^9 = {}^{10}C_1 (0.2175)(0.7825)^9$ $= (10)(0.2175)(0.1100)$ $= 0.2393$	$80 \times 0.2393 = 19.144$
2	${}^{10}C_2\, p^2 q^8 = {}^{10}C_2 (0.2175)^2 (0.7825)^8$ $= (45)(0.0473)(0.1406)$ $= 0.2993$	$80 \times 0.2993 = 23.944$
3	${}^{10}C_3\, p^3 q^7 = {}^{9}C_3 (0.2175)^3 (0.7825)^7$ $= (120)(0.0103)(0.1796)$ $= 0.2220$	$80 \times 0.2220 = 17.76$
4	${}^{10}C_4\, p^4 q^6 = {}^{10}C_4 (0.2175)^4 (0.7825)^6$ $= (210)(0.0022)(0.2295)$ $= 0.1060$	$80 \times 0.1060 = 8.48$
5	${}^{10}C_5\, p^5 q^5 = {}^{10}C_5 (0.2175)^5 (0.7825)^5$ $= (252)(0.0005)(0.2934)$ $= 0.0370$	$80 \times 0.0370 = 2.96$
6	${}^{10}C_6\, p^6 q^4 = {}^{10}C_6 (0.2175)^6 (0.7825)^4$ $= (210)(0.0001)(0.3749)$ $= 0.0079$	$80 \times 0.0079 = 0.632$
7	${}^{10}C_7\, p^7 q^3 = {}^{10}C_7 (0.2175)^7 (0.7825)^3$ $= (120)(0.0000)(0.4791)$ $= 0$	$80 \times 0 = 0.$
8	${}^{10}C_8\, p^8 q^2 = {}^{9}C_8 (0.2175)^8 (0.7825)^2$ $= 0$	$80 \times 0 = 0$
9	${}^{10}C_9\, p^9 q^1 = {}^{9}C_9 (0.2175)^9 (0.7825)$ $= 0$	$80 \times 0 = 0$
10	${}^{10}C_{10}\, p^{10} q^0 = {}^{9}C_{10} (0.2175)^{10} (0.7825)$ $= 0$	$80 \times 0 = 0$

Syllabus Topic : Properties of Binomial Distribution

3.5 Properties of Binomial Distribution

(i) Mean

For the Binomial Distribution, $P(r) = {}^nC_r\, q^{n-r} p^r$

$$\text{Mean } \mu = \sum_{r=0}^{n} r\,P(r) = \sum_{r=0}^{n} r \cdot {}^nC_r\, q^{n-r} p^r$$

Using ${}^nC_r = \dfrac{n(n-1)(n-2)\ldots(n-(r-1))}{r!}$

$$\therefore\ \mu = 0 + 1 \cdot {}^nC_1\, q^{n-1} p + 2 \cdot {}^nC_2\, q^{n-2} p^2 + 3 \cdot {}^nC_3\, q^{n-3} p^3 + \ldots + n \cdot {}^nC_n\, p^n$$

$$= n q^{n-1} p + 2 \cdot \frac{n(n-1)}{2.1} q^{n-2} p^2 + 3 \cdot \frac{n(n-1)(n-2)}{3.2.1} q^{n-3} p^3 + \ldots + n \cdot p^n$$

$$= n q^{n-1} p + n(n-1) q^{n-2} p^2 + \frac{n(n-1)(n-2)}{2.1} q^{n-3} p^3 + \ldots + n p^n$$

Common out np from R.H.S.

$$= np\left[q^{n-1} + (n-1) q^{n-2} p + \frac{(n-1)(n-2)}{2.1} q^{n-3} p^2 + \ldots + p^{n-1}\right]$$

Again using ${}^nC_r = \dfrac{n(n-1)(n-2)\ldots(n-(r-1))}{r!}$

we can write,

$$\therefore\ \mu = np\,[{}^{n-1}C_0\, q^{n-1} + {}^{n-1}C_1\, q^{n-2} p + {}^{n-1}C_2\, q^{n-3} p^2 + \ldots + {}^{n-1}C_{n-1}\, p^{n-1}]$$

$$\therefore\ \mu = np\,(q+p)^{n-1} = np \qquad (\because\ p + q = 1)$$

Hence the mean of the Binomial Distribution in np.

(ii) Variance

$$\text{Variance } \sigma^2 = \sum_{r=0}^{n} r^2 P(r) - \mu^2$$

$$= \sum_{r=0}^{n} [r - r + r^2]\,P(r) - \mu^2 \text{ (Note the adjustment)}$$

$$= \sum_{r=0}^{n} [r + r(r-1)]\,P(r) - \mu^2$$

Separate the summation, we get,

$$\sigma^2 = \sum_{r=0}^{n} rP(r) + \sum_{r=0}^{n} r(r-1)\,P(r) - \mu^2$$

$$= \mu + \sum_{r=2}^{n} r(r-1)\,{}^nC_r\, q^{n-r} p^r - \mu^2$$

(since the contribution due to r = 0 and r = 1 is zero)

$$= \mu + [2.1 \cdot {}^nC_2\, q^{n-2} p^2 + 3.2 \cdot {}^nC_3\, q^{n-3} p^3 + \ldots + n(n-1)\,{}^nC_n\, p^n] - \mu^2$$

$$= \mu + \left[2.1 \cdot \frac{n(n-1)}{2.1} q^{n-2} p^2 + 3.2 \cdot \frac{n(n-1)(n-2)}{3.2.1} q^{n-3} p^3 + \ldots + n(n-1) p^n\right] - \mu^2$$

$$= \mu + [n(n-1) q^{n-2} p^2 + n(n-1)(n-2) q^{n-3} p^3 + \ldots + n(n-1) p^n] - \mu^2$$

$$= \mu + n(n-1) p^2 [q^{n-2} + (n-2) q^{n-3} p + \ldots + p^{n-2}] - \mu^2$$

$$= \mu + n(n-1) p^2 [{}^{n-2}C_0\, q^{n-2} + {}^{n-2}C_1\, q^{n-3} p + \ldots + {}^{n-2}C_{n-2}\, p^{n-2}] - \mu^2$$

$$= \mu + n(n-1) p^2 (q+p)^{n-2} - \mu^2 = \mu + n(n-1) p^2 - \mu^2 \qquad [\because\ q + p = 1]$$

$$= np + n(n-1) p^2 - n^2 p^2 \qquad [\because\ \mu = np]$$

$$= np\,[1 + (n-1) p - np] = np\,[1 - p] = npq$$

Hence the variance of the Binomial Distribution is npq.

(iii) Standard deviation (S.D.)

We know variance $= (\text{S.D.})^2$

Hence standard deviation of Binomial Distribution is $\sqrt{npq}$

(iv) Coefficient of skewness and Kurtosis

$$\gamma_1 = \frac{q - p}{\sqrt{npq}} = \frac{1 - 2p}{\sqrt{npq}}$$

gives the measure of skewness of the Binomial Distribution.

If $P < \dfrac{1}{2}$, skewness is positive ;

if $P > \dfrac{1}{2}$; skewness is negative and $P = \dfrac{1}{2}$; skewness is zero.

$\beta_2 = 3 + \dfrac{1 - \sigma pq}{npq}$ gives a measure of kurtosis

Table 3.5.1 : Properties of Binomial Distribution

Mean	$\mu = np$
Variance	$\sigma^2 = npq$
Standard deviation	$\sigma = \sqrt{npq}$
Coefficient of skewness	$\dfrac{q - p}{\sqrt{nqp}}$
Coefficient of kurtosis	$3 + \dfrac{1 - \sigma pq}{npq}$
Moment generating function	$M(t) = (q + pe^t)^n$
Characteristic function	$\phi(\omega) = (q + pe^{i\omega})^n$

3.5.1 Applications of Binomial Distribution :

(i) Binomial Distribution is Number of defectives in a sample from production line,

(ii) Estimation of reliability of systems.

(iii) Number of rounds fired from a gun hitting a target.

(iv) Radar detection.

1. Quality control measures and sampling process in industries to classify items as defectives or non-defective.
2. Medical applications such a success or failure, cure or no-cure.

Syllabus Topic : Properties of Poisson Distribution

3.6 Properties of Poisson Distribution

(i) Mean :

For the Poisson Distribution, $P(r) = \dfrac{\lambda^r e^{-\lambda}}{r!}$

$$\text{Mean } \mu = \sum_{r=0}^{\infty} r\,P(r) = \sum_{r=0}^{\infty} r \cdot \frac{\lambda^r e^{-\lambda}}{r!}$$

$$= e^{-\lambda} \sum_{r=1}^{\infty} \frac{\lambda^r}{(r-1)!} \qquad \text{since } r! = (r-1)! \times r$$

$$= e^{-\lambda}\left(\lambda + \frac{\lambda^2}{1!} + \frac{\lambda^3}{2!} + \ldots\ldots\right)$$

$$\therefore\ \mu = \lambda e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots\ldots\right)$$

Using series of $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \ldots$

$$= \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda$$

Thus, the mean of the Poisson Distribution is equal to the parameter λ.

Also we can prove from Binomial Distribution

Since Mean of Binomial Distribution is np.

$\therefore$ Mean of Poisson Distribution $= \underset{n \to \infty}{\text{Lt}}\, np = \underset{n \to \infty}{\text{Lt}}\, \lambda = \lambda$

(ii) Variance

Variance $\sigma^2 = \sum_{r=0}^{\infty} r^2 P(r) = \mu^2$

$$= \sum_{r=0}^{\infty} r^2 \cdot \frac{\lambda^r e^{-\lambda}}{r!} - \lambda^2 = e^{-\lambda} \sum_{r=1}^{\infty} \frac{r^2 \lambda^r}{r!} - \lambda^2$$

$$= e^{-\lambda}\left[\frac{1^2 \cdot \lambda}{1!} + \frac{2^2 \cdot \lambda}{2!} + \frac{3^2 \lambda^3}{3!} + \frac{4^2 \lambda^4}{4!} + \ldots\ldots\right] - \lambda^2$$

$$= \lambda e^{-\lambda}\left[1 + \frac{2\lambda^2}{1!} + \frac{3\lambda^2}{2!} + \frac{4\lambda^3}{3!} + \ldots\ldots\right] - \lambda^2$$

$$= \lambda e^{-\lambda}\left[1 + \frac{(1+1)\lambda}{1!} + \frac{(1+2)\lambda^2}{2!} + \frac{(1+3)\lambda^3}{3!} + \ldots\ldots\right] - \lambda^2$$

$$= \lambda e^{-\lambda}\left[\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots\ldots\right) + \left(\frac{\lambda}{1!} + \frac{2\lambda^2}{2!} + \frac{3\lambda^3}{3!} + \ldots\right)\right] - \lambda^2$$

$$= \lambda e^{-\lambda}\left[e^{\lambda} + \lambda\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots\ldots\right)\right] - \lambda^2$$

$$= \lambda e^{-\lambda}[e^{\lambda} + \lambda e^{\lambda}] - \lambda^2$$

$$= \lambda e^{-\lambda} \cdot e^{\lambda}(1 + \lambda) - \lambda^2 = \lambda(1 + \lambda) - \lambda^2 = \lambda$$

Hence, the variance of the Poisson Distribution is also λ.

Also we can prove from Binomial Distribution.

Since,

Variance of Binomial Distribution of $npq = np(1 - p)$

$\therefore$ Variance of Poisson Distribution $= \underset{n \to \infty}{\text{Lt}}\, np(1 - p)$

$$= \underset{n \to \infty}{\text{Lt}}\, \lambda\left(1 - \frac{\lambda}{n}\right) = \lambda$$

(iii) Standard deviation (S.D.)

We know variance $= (\text{S.D.})^2$

Hence standard deviation of Poisson's Distribution is $\sqrt{\lambda}$

Table 3.6.1 : Properties of Poisons Distribution

Mean	$\mu = \lambda$
Variance	$\sigma^2 = \lambda$
Standard deviation	$\sigma = \sqrt{\lambda}$
Coefficient of skewness	$1/\sqrt{\lambda}$
Coefficient of kurtosis	$3 + (1/\sqrt{\lambda})$
Moment generating function	$M(t) = e^{\lambda(e^t - 1)}$
Characteristic function	$\phi(\omega) = e^{\lambda(e^{i\omega} - 1)}$

Applications of Poisson's Distribution :

Poisson's Distribution is using in

(i) Quality control statistics to count the number of defects of an item

(ii) Arrival pattern of 'defective vehicles in a workshop', 'patients in a hospital' or 'telephone calls'.

(iii) Determining the number of details in district in a given period. By rare disease.

(iv) Number of fragments from a shell hitting a target.

(v) The number of plants infected with a particular disease in plot of filed.

(vi) In biology, to count the number of bacteria.

(vii) Number of weeds in particular species in different plots of a field.

Relation between Binomial and Poisson Distribution

Poisson's Distribution can be derived as a limiting case of the Binomial Distribution by making n very large and p very small, keeping up fixed (= m, say).

The probability of r successes in a Binomial-Distribution is

$$P(r) = {}^nC_r\, p^r q^{n-r} = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!} p^r q^{n-r}$$

$$= \frac{np(np-p)(np-2p)\ldots(np-r-1p)}{r!}(1-p)^{n-r}$$

As $n \to \infty$, $p \to 0$ $(np = m)$, we have

$$P(r) = \frac{m^r}{r!} \underset{n \to \infty}{\text{Lt}} \frac{(1 - m/n)^n}{(1 - m/n)^r} = \frac{m^r}{r!} e^{-m}$$

What the probabilities of 0, 1, 2, ..., r, ... successes in a Poisson Distribution are given by

$$e^{-m}, m e^{-m}, \frac{m^2 e^{-m}}{2!}, \ldots\ldots, \frac{m^2 e^{-m}}{r!}, \ldots\ldots$$

The sum of these probabilities is unity as it should be.

Constants of the Poisson Distribution

These constants can easily be derived from the corresponding constants of the Binomial Distribution simply by making $n \to \infty$, $p \to 0$, $(q \to 1)$ and noting that $np = m$

Mean $= \text{Lt}(np) = m$

$\mu_2 = \text{Lt}(npq) = m\,\text{Lt}(q) = m$

Standard deviation $= \sqrt{m}$

Also $\mu_3 = m$, $\mu_4 = m + 3m^2$

Syllabus Topic : Properties of normal Distribution

3.7 Properties of normal Distribution

I. The normal curve is bell-shaped and is symmetrical about its mean.

II. As it is symmetrical, its mean, median and modes are the same. Mean, median, and mode of the Distribution are coincide i.e., Mean = Median = Mode = μ

III. It has only one mode at $x = \mu$ (i.e., unimodal)

IV. Ordinates decreasing rapidly on both sides of the mean. The maximum ordinate occurs at $x = \mu$ and its value is $= \frac{1}{\sigma\sqrt{(2\pi)}}$

V. Its points of inflexion (four by putting $d^2y/dx^2 = 0$ and verifying that at these points $d^3y/dx^3 \neq 0$) are given by $x = \mu \pm \sigma$, these points are equidistant from the mean on either side.

VI. Mean deviation from the mean μ

$$= \int_{-\infty}^{\infty} |x - \mu| \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \qquad [\text{Put } z = (x - \mu)]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| e^{-z^2/2} dz$$

$$= \frac{\sigma}{\sqrt{2\pi}}\left[\int_{-\infty}^{0} -z e^{-z^2/2} dz + \int_{-0}^{\infty} z e^{-z^2/2} dz\right]$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-z^2/2} dz$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \left| -e^{-z^2/2} \right|_0^{\infty} = -\sqrt{\left(\frac{2}{\pi}\right)}\, \sigma(0 - 1)$$

$$= 0.7979\,\sigma = (4/5)\,\sigma$$

VII. Moments about the mean

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x - \mu)^{2n+1} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} dz$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} e^{-z^2/2} dz$$

where $z = (x - \mu)/\sigma$

$= 0$, since the integral is an odd function.

Thus, all odd order moments about the mean vanish.

$$\mu_{2n} = \int_{-\infty}^{\infty} (x - \mu)^{2n} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} z^{2n-1} e^{-z^2/2} \cdot z dz$$

[Integrated by parts]

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}}\left[\left| z^{2n-1} e^{-z^2/2} \right|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} (2n-1) z^{2n-2} e^{-z^2/2} dz\right]$$

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} (0 - 0) + (2n - 1)\sigma^2 \mu_{2n-2}$$

– Repeated application of this reduction formula, gives :

$$\mu_{2n} = (2n - 1)(2n - 3) \ldots\ldots 3 \cdot 1\, \sigma^{2n}$$

In particular, $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$

Hence $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$ and $\beta_2 = \frac{\mu_4}{\mu_2^2} = 3$

i.e. the coefficient of skewness is zero (i.e. the curve is symmetrical) and the Kurtosis is 3. The is the basis for the choice of the value 3 in the definitions of platykurtic and leptokurtic.

VIII. The probability of x lying between x_1 and x_2 is given by the area under the normal curve from x_1 to x_2, i.e. $P(x_1 \leq x \leq x_2)$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{x_1}^{x_2} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_{z_1}^{z_2} e^{-z^2/2} dz$$

where $z = (x - \mu)/\sigma$, $dz = dx/\sigma$ and $z_1 = (x_1 - \mu)/\sigma$,

$z_2 = (x_2 - \mu)/\sigma$

$$= \frac{1}{\sqrt{(2\pi)}}\left\{\int_0^{z_2} e^{-z^2/2} dz - \int_0^{z_1} e^{-z^2/2} dz\right\} = P_2(z) - P_1(z)$$

The values of each of the above integrals can be found from the table VI-Appendix 2, which gives the values of

$$P(z) = \frac{1}{\sqrt{(2\pi)}} \int_0^{z_2} e^{-z^2/2} dz$$

for various values of z. This integral is called the probability integral or the error function due to its use in the theory of sampling and the theory of errors.

Using this table, we see that the area under the normal curve from $z = 0$ to $z = 1$, i.e. from $x = \mu$ to $\mu + \sigma$ is 0.3413.

$\therefore$ (i) The area under the normal curve between the ordinates $x = \mu - \alpha$ and $x = \mu + \sigma$ is 0.6826, ~ 68 % nearly. Thus approximately 2/3 of the values lie within these limits.

(ii) The area under the normal curve between $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$ is 0.9544 – 95.5 %, which implies that about $4\frac{1}{2}$% of the values lie outside these limits.

(iii) 99.73% of the values lie between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$ i.e. only a quarter % of the whole lies outside these limits.

Syllabus Topic : Relation between Binomial and Normal Distribution

3.8 Relation between Binomial and Normal Distribution

- If n is large and if neither p nor q is too close to zero, the binomial distribution can be closely approximated by a normal distribution with standardized random variable given by

$$Z = \frac{X - np}{\sqrt{npq}}$$

- Here X is the random variable giving the number of successes in n Bernoulli trials and p is probability of success.
- The approximation becomes better with increasing n and is exact in the limiting case.
- If np and nq both are greater than 5 then approximation is very good.
- The binomial distribution approaches the normal distribution can be described as

$$\lim_{n \to \infty} P\left(a \le \frac{X - np}{\sqrt{npq}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz$$

Syllabus Topic : Relation between Poisson and Normal Distribution

3.9 Relation between Poisson and Normal Distribution

- We know there is a relation between the binomial and normal distributions. Also a relation between binomial and poisson distributions.
- If X is the poisson random variable of

$P(X = r) = \frac{\lambda^r e^{-\lambda}}{r!}$ and $\frac{X - \lambda}{\sqrt{\lambda}}$ is the corresponding standardized random variable then

$$\lim_{\lambda \to \infty} P\left(a \le \frac{X - \lambda}{\sqrt{\lambda}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz$$

i.e. the Poisson distribution approaches the normal distribution as $\lambda \to \infty$ or $\frac{X - \lambda}{\sqrt{\lambda}}$ is asymptotically normal.

Applications of Normal Distribution

Normal distribution is using

(i) to calculation of errors made by chance in experimental measurements.

(ii) Computation of hit probability of a shot.

(iii) Statistical inference in almost every branch of science.

Syllabus Topic : Importance of Normal Distribution

3.10 Importance of Normal distribution

Normal distribution plays a very important role in statistical theory because of the following reasons :

(i) A large number of RVs, such as binomial and Poisson, occurring in many applications have a distribution closely resembling the normal distribution.

(ii) Many of the distributions of sample statistics, such as sample mean and sample variance, tend to normality for samples of large size. In particular, the sampling distributions like Student's t, Snedecor's F and Chi-square distributions tend to normality when the size of the sample is large.

(iii) Tests of significance for small samples are based on the assumption that samples have been drawn from normal populations.

(iv) Even if a variable is not normally distributed, it can sometimes be converted into a normal variable by simple transformation of the variable.

(v) Normal distribution is applied to a large extent in statistical Quality Control in industry.

Multiple Choice Questions

Q. 1 If probability of success and failure are p and q respectively then for n number of trials, mean is _____.

(a) nq (b) np (c) npq (d) $\sqrt{npq}$

Ans. : (b)

Q. 2 If probability of success and failure are p and q respectively then for n number of trial, standard deviation (S.D.) is _____.

(a) nq (b) np (c) npq (d) $\sqrt{npq}$

Ans. : (d)

Q. 3 y the Binomial distribution, probability of r success in a series of n trial is

P(r) = _____.

(a) ${}^nC_r p^r q^{n-r}$ (b) ${}^nC_r (p+q)^r$

(c) ${}^nC_r p^{n-r} q^r$ (d) ${}^nP_r p^r q^{n-r}$

Ans. : (a)

Q. 4 By the Poission's distribution probability of r success with mean m in a series of n trial is P(r) = _____.

(a) $\frac{r e^{-m}}{m!}$ (b) $\frac{r^m e^{-m}}{r!}$ (c) $\frac{m^r e^{-m}}{r!}$ (d) $\frac{r^m e^{-m}}{m!}$

Ans. : (c)

Q. 5 In a Poisson distribution mean is m then S.D. of this distribution is _____.

(a) mp (b) mpq (c) $\sqrt{m}$ (d) m

Ans. : (c)

Q. 6 The normal distribution curve with mean μ and standard derivation σ is _____.

(a) $\frac{\sigma^2}{2\pi} e^{-(x-\mu)^2}$ (b) $\frac{\sigma}{\sqrt{2\pi}} e^{-(x-\mu)^2}$

(c) $\frac{\sigma}{2\pi} e^{-(x-\mu)^2/2\sigma^2}$ (d) $\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

Ans. : (d)

Q. 7 If probability of success p = 0.7 then probability of failure q = _____.

(a) 0.7 (b) 1.7 (c) – 0.7 (d) 0.3

Ans. : (d)

Explanation : $p + q = 1 \therefore q = 0.3$

Q. 8 The probability of strike a target is $\frac{1}{5}$. If six bombs are there then probability that exactly two will strike the target is _____.

(a) 0.5 (b) 0.2 (c) 0.738 (d) 0.246

Ans. : (d)

Q. 9 The mean of Binomial probability distribution is,

(a) $\sqrt{np}$ (b) np (c) npq (d) $\sqrt{npq}$

Ans. : (b)

Q. 10 Variance of Binomial distribution is,

(a) $\sqrt{np}$ (b) np (c) npq (d) $\sqrt{npq}$

Ans. : (c)

Q. 11 Standard deviation of binomial probability distribution is,

(a) $\sqrt{np}$ (b) $\sqrt{npq}$ (c) np (d) npq

Ans.: (b)

Q. 12 What is the probability of getting two heads, if an unbiased coin is thrown four tones.

(a) $\frac{5}{16}$ (b) $\frac{16}{5}$ (c) 16 (d) $\frac{3}{8}$

Ans. : (d)

Explanation :

Given : r = 2; $p = \frac{1}{2}$

$q = \frac{1}{2}$; n = 4

By Binomial distribution :

$$P(r = 2) = {}^nC_r P^r q^{n-r} = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2$$

$$= 6 \times \frac{1}{4} \times \frac{1}{4} = \frac{3}{8}$$

Q. 13 Out of 800 families with 4 children each, then find the number of families having no girls.

(a) 100 (b) 125 (c) 50 (d) 9

Ans. : (c)

Explanation :

Given, N = 800; n = 4; $p = \frac{1}{2}$, $q = \frac{1}{2}$, r = 0;

$$p(r = 0) = {}^nC_r p^r q^{n-r} = {}^4C_0 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)^4$$

$$= 1 \times 1 \times \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

$\therefore$ Number of families having no girl $= 800 \times \frac{1}{16} = 50$

Q. 14 Given, the probability of a man now aged 60 years will live 70 years, is 0.65. What is the probability that out of 10 men, (60 years old) 2 men will live upto 70 years.

(a) 0.004281 (b) 0.003281

(c) 0.23 (d) 0.4281

Ans. : (a)

Explanation :

Here, P = 0.65 ; q = 1 – p = 0.35; r = 2 ; n = 10;

P(only two men will live upto 70 years)

$= {}^nC_r P^r q^{n-r}$

$= {}^{10}C_2 (0.65)^2 (0.35)^8$

$= 0.004281$

Q. 15 Given, the probability that a person hit target in shooting practice is 0.4. If he shoots 10 times then the probability that 'the man' hits the targets is,

(a) $1 - (0.6)^{10}$ (b) $(0.6)^{10}$

(c) $1 - (0.4)^{10}$ (d) (0.4)10 4

Ans. : (a)

Explanation : Here : p = 0.4 ; q = 1 – p = 0.6 ;

n = 10 ; r = 0 (No hits the target)

$\therefore$ p (r = 0 ; no hits the target)

$= {}^nC_r P^r q^{n-r} = {}^{10}C_3 (0.4)^{10} (0.6)^{10}$

$= 1 \times 1 \times (0.6)^{10} = (0.6)^{10}$

Hence, the probability that the man hits the target.

$= 1 - P(r = 0) = 1 - (0.6)^{10}$

Q. 16 A box contains 100 bulbs, out of which 10 are defective. A sample of 5 bulbs are drawn. What is the probability of, none is defective.

(a) $(10)^5$ (b) $(9)^5$ (c) $\left(\frac{9}{10}\right)^5$ (d) $\left(\frac{10}{9}\right)^5$

Ans. : (c)

Explanation : Given, $P = \frac{10}{100} = 0.1; q = 1 - p = 0 - 9$

$n = 5;\ r = 0$ (none is defective)

By Binomial distribution,

$P(r = 0) = {}^nC_r\, P^r q^{n-r} = {}^5C_0\,(0.1)0\,(0.9)^5$

$= 1 \times 1 \times (0.9)^5 = (0.9)^5 = \left(\frac{9}{10}\right)^5$

Q. 17 In 100 set of tosses of a coin, the number of cases you expect 7 head and 3 tail are ,

(a) 10 (b) 20 (c) 9 (d) 12

Ans. (d)

Explanation : Given, $N = 100$;

$n = 10;\ P = \frac{1}{2};\ q = \frac{1}{2};\ r = 7;$

$p(r = 7) = N\,[{}^nC_r\, p^r q^{n-r}] = 100\left[{}^{10}C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3\right]$

$= 100\left[120 \times \frac{1}{2^7} \times \frac{1}{2^3}\right] = 100\left[120 \times \frac{1}{2^{10}}\right]$

$= 11.72 \approx 12$

Q. 18 20% of bolts produced by machine are defective. The mean and standard deviation of defective bolts in total of 900 bolts are respectively.

(a) 180 and 12 (b) 12 and 100

(c) 12 and 180 (d) 9 and 0.8

Ans. : (a)

Explanation : Given . $p = 20\% = \frac{20}{100} = 0.2$;

$q = 0.8;\ n = 900$

By Binomial distribution,

Mean $= np = 900 \times 0.2 = 180$ and

standard deviation $= \sqrt{npq}$

$= \sqrt{(900) \times (0.2) \times (0.8)}$

$= 12$

Q. 19 The mean and standard deviation of a binomial distribution are $\frac{3}{4}$ and $\frac{7}{8}$ respectively then the probability of success in a single trials is,

(a) 0.12 (b) 0.8 (c) 0.2 (d) 0.75

Ans. : (c)

Explanation : Given : mean $(np) = \frac{3}{4}$; standard deviation $= \sqrt{npq} = \frac{7}{9}$

$\therefore\ np = \frac{3}{4}$ and $npq = \frac{49}{81}$

$\therefore \left(\frac{3}{4}\right)q = \frac{49}{81} \Rightarrow q = \frac{49}{81} \times \frac{4}{3} = \frac{196}{243}$

$q = 0.8$

$\therefore p = 1 - q = 0.2$

Q. 20 The mean and variance of binomial probability distribution are 6 and 4 respectively then number of trials 'n' is,

(a) 10 (b) 18 (c) 10 (d) 9

Ans. : (b)

Explanation : Given : mean $= np = 6$; and variance $= npq = 4$

$\therefore\ npq = 4 \qquad \Rightarrow q = \frac{4}{6} = \frac{2}{3}$

$p = 1 - q = 1 - \frac{4}{6} = \frac{2}{6} = \frac{1}{3}$

$\therefore\ np = 6$

$\Rightarrow n\left(\frac{1}{3}\right) = 6 \qquad \Rightarrow n = 18$

Number of trials = 18

Q. 21 The mean and standard deviation of a Binomial distribution are 25 and 5 respectively. Number of trials 'n' is,

(a) 42 (b) 40 (c) 9 (d) 44

Ans. : (a)

Explanation : Given, Mean $= np = 25$; and standard deviation $= \sqrt{npq} = 3$

$\Rightarrow\ npq = 9$

$\therefore\ q = \frac{9}{25} = 0.4$

$\Rightarrow\ p = 0.6 \qquad \because\ np = 25$

$n = \frac{25}{P} = \frac{25}{0.6} \approx 42$

Q. 22 The mean and standard deviation of Binomial probability distribution are 6 and $\sqrt{2}$ respectively then what is the value of $p(r \geq 2)$,

(a) 0.88 (b) 0.99 (c) 0.1 (d) 0.55

Ans.: (b)

Explanation : Given, mean $(np) = 6$;

Standard derivation $(\sqrt{npq}) = \sqrt{2}$;

$(npq) = 2$

Also given $r \geq 2$

$\therefore\ np = 6$ and $npq = 2$

$\Rightarrow q = \frac{1}{3}$ and $p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3}$

And $n = \frac{6}{p} = \frac{6}{(2/3)} = 9$

$p(r = 0, 1) = {}^nC_r\, p^r q^{n-r}$

$= {}^9C_0\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^9 + {}^9C_1\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^8$

$= 1 \times 2 \times \frac{1}{3^9} + 9 \times \frac{2}{3} \times \frac{1}{3^8}$

$= \frac{2}{3^9} + 9\frac{2}{3^9} = 10\left(\frac{2}{3^9}\right)$

$\therefore\ p(r \geq 2) = 1 - P(r = 0.1) = 1 - 10\left(\frac{2}{3^9}\right)$

$= 1 - 0.001 = 0.999$

Q. 23 The probability of strike a target is $\frac{1}{5}$. If six bombs are these then probability that at least one will strike the target is _____.

(a) 0.57 (b) 0.234 (c) 0.738 (d) 0.246

Ans. : (c)

Explanation :

Probability of strike a target $p = \frac{1}{5} = 0.2$

$\therefore q = 0.8$ and $n = 6$

By Binomial distribution

Probability that at least one will strike the target = $P(1) + P(2) + \ldots + P(6)$

$= 1 - [P(0)] = 1 - {}^6C_0\, p^0 q^6 = 0.738$

Q. 24 20 % of bolts produced by a machine are defective; out of 4 bolts chosen at random probability of 1 defective bolt is _____.

(a) 0.5212 (b) 0.6834 (c) 0.4096 (d) 0.9245

Ans. : (c)

Q. 25 20 % of bolts produced by a machine are defective; out of 4 bolts chosen at random probability of at least 1 is defective is _____.

(a) 0.5904 (b) 0.9331

(c) 0.3015 (d) 0.7282

Ans. : (a)

Explanation : $n = 4, p = 20\% = 0.2$

$\therefore q = 1 - p = 0.8$

By Binomial distribution, Probability of at least 1 is defective $= P(x \geq 1)$

$= P(1) + P(2) + P(3) + P(4) = 1 - P(0)$

$= 1 - [{}^4C_0\, p^0 q^4] = 0.5904$

Q. 26 20 % of bolts produced by a machine are defective, out of 4 bolts chosen at random probability of at most 2 are defective is _____.

(a) 0.2728 (b) 0.5721 (c) 0.7915 (d) 0.9728

Ans. : (d)

Q. 27 If 10 % of riverts produced by a machine are defective. Out of 5 riverts chosen at random the probability of none will be defective is _____.

(a) 0.7217 (b) 0.5905 (c) 0.9381 (d) 0.1232

Ans. : (b)

Q. 28 If 10 % of riverts produced by a machine are defective Out of 5 riverts chosen at random the probability of one will be defective is _____.

(a) 0.3280 (b) 0.9215 (c) 0.7242 (d) 0.1251

Ans. : (a)

Q. 29 If 10 % of riverts produced by a machine are defective Out of 5 riverts chosen at random the probability of at least two will be defective is _____.

(a) 0.4102 (b) 0.7391 (c) 0.3202 (d) 0.0815

Ans. : (d)

Q. 30 If 10 % of riverts produced by a machine are defective out of 5 riverts chosen at random the probability of at most two will be defective is _____.

(a) 0.1234 (b) 0.3801 (c) 0.9914 (d) 0.2115

Ans. : (c)

Explanation : $n = 5,\ p = 10\% = 0.1 \quad \therefore q = 0.9$

By Binomial distribution

Probability of at most two will be defective $= P(x \leq 2)$

$= P(0) + P(1) + P(2) = {}^5C_0\, p^0 q^5 + {}^5C_1\, p^1 q^4 + {}^5C_2\, p^2 q^3$

$= 0.9914$

Q. 31 The probability that a student who is freshman will graduate is 0.4. Out of 5 students, the probability of none will be graduate is _____.

(a) 0.3164 (b) 0.6231 (c) 0.0777 (d) 0.2052

Ans. : (c)

Q. 32 The probability that a student who is freshman will graduate is 0.4. Out of 5 students the probability of at least one will be graduate is _____.

(a) 0.3242 (b) 0.9223 (c) 0.7392 (d) 0.1275

Ans. : (b)

Q. 33 The probability that a student who is freshman will graduate is 0.4, Out of 5 students, the probability of all will be graduate is _____.

(a) 0.0102 (b) 0.3412 (c) 0.7271 (d) 0.4256

Ans. : (a)

Explanation : $n = 5$, $p = 0.4$ $\therefore q = 1 - p = 0.6$

By Binomial Distribution

The Probability of all will be graduate

$= P(x = 5)$

$= P(5) = {}^5C_5\, p^5 q^0 = 0.0102$

Q. 34 On an average the box containing 10 articles is likely to have 2 defectives. In a consignment of 100 boxes, the total number of boxes expected to have exactly one defective are _____.

(a) 72 (b) 89 (c) 35 (d) 29

Ans. : (d)

Explanation : $n = 10$, $p = 0.2$ $\therefore q = 1 - p = 0.8$

By Binomial Distribution

The Probability of one defective $= {}^{10}C_1\, p^1 q^9$

$= 0.2684$

$\therefore$ Total no. of boxes $= 100 \times 0.268 = 26.8 \approx 29$.

Q. 35 On an average the box containing 10 articles is likely to have 2 defectives. In a consignment of 100 boxes, the total number of boxes expected to have at least one defective are _____.

(a) 29 (b) 72 (c) 89 (d) 35

Ans. : (c)

Explanation : $n = 10$, $p = 0.2$ $\therefore q = 1 - p = 0.8$

By Binomial Distribution

The Probability of at least one defective $= P(x \geq 1)$

$= P(1) + P(2) + \ldots + P(10)$

$= 1 - [P(0)] = 1 - [{}^{10}C_0\, p^0 q^{10}] = 0.8926$

$\therefore$ Total no. of boxes $= 100 \times 0.8926 = 89.26 \approx 89$.

Q. 36 A group of 20 aeroplanes are sent on an operational flight. The chances that an aeroplane fails to return from the flight is 5%, then probability that no plane return is _____.

(a) 0.3584 (b) 0.9215 (c) 0.2738 (d) 0.5748

Ans. : (a)

Explanation : $n = 20$, $p = 0.05$ $\therefore q = 0.95$

By Binomial Distribution $= P(x = 0)$

The Probability of no plane return $= {}^{20}C_0\, p^0 q^{20}$

$= 0.3584$

Q. 37 A group of 20 Aeroplanes are sent on an operational flight the chances that an aeroplane fails to return from the flight is 5%, then probability that at least one plane do not return is _____.

(a) 0.8225 (b) 0.9187 (c) 0.6416 (d) 0.3758

Ans. : (c)

Explanation : $n = 20$, $p = 0.05$ $\therefore q = 0.95$

By Binomial Distribution

The Probability of at least one plane does not return

$= P(x \geq 1)$

$= P(1) + P(2) + \ldots + P(20) = 1 - P(0)$

$= 1 - [{}^{20}C_0\, p^0 q^{20}] = 0.6416$

Q. 38 A group of 20 aeroplanes are sent on an operational flight the chances that an aeroplane fails to return from the flight is 5% then probability that exactly one plane does not return is _____.

(a) 0.9178 (b) 0.6448 (c) 0.3773 (d) 0.4232

Ans. : (c)

Explanation : $n = 20$, $p = 0.05$ $\therefore q = 0.95$

By Binomial Distribution

The Probability of exactly one plane does not return

$= P(x = 1)$

$= P(1) = {}^{20}C_1\, p^1 q^{19} = 0.3773$

Q. 39 The probability of 5 success in a series of 10 trial with $p = 0.2$ is _____.

(a) 0.3183 (b) 0.0975 (c) 0.5738 (d) 0.0264

Ans. : (d)

Explanation : $n = 10$, $p = 0.2$ $\therefore q = 1 - p = 0.8$

By Binomial distribution

The probability of 5 success $= {}^{10}C_5\, p^5 q^5 = 0.0264$

Q. 40 In 20 trials, the probability of 2 success with $p = 0.3$ is _____.

(a) 0.028 (b) 0.21 (c) 0.09 (d) 0.95

Ans. : (a)

Explanation : $n = 20$, $p = 0.3$ $\therefore q = 1 - p = 0.7$

By Binomial distribution

The probability of 2 success $= P(2) = {}^{20}C_2\, p^2 q^{18}$

$= 0.028$

Q. 41 Probability of man now aged 60 years will live up to 70 years of age is 0.65. The probability of out of 10 men of 60 years old 2 or more will live up to age of 70 years is _____.

(a) 0.888 (b) 0.999 (c) 0.777 (d) 0.222

Ans. : (b)

Explanation : $n = 10$, $p = 0.65$, $q = 1 - p = 0.35$

By Binomial Distribution, $= P(x \geq 2)$

$P(r \geq 2) = 1 - [P(0) + P(1)]$

$= 1 - \left[{}^{10}C_0\, p^0 q^{10} + {}^{10}C_1\, p^1 q^9\right]$

$= 0.999$

Q. 42 In 100 sets of 10 tosses of a coin, the probability to get 7 heads and 3 tails is ___.

(a) 12 (b) 8 (c) 9 (d) 7

Ans. : (a)

Explanation : $n = 10$

Probability to get head $= p = 0.5$ $\therefore q = 0.5$

By Binomial Distribution,

$P(7) = {}^{10}C_7\, p^7 q^3 = 0.117$

$\therefore$ Required $= 100 \times 0.117 = 11.7 \approx 12$

Q. 43 In 10 tosses of a coin the probability to get at least 8 heads is _____.

(a) 0.231 (b) 0.5 (c) 0.39 (d) 0.055

Ans. : (d)

Explanation : $n = 10$

Probability to get head $= p = 0.5$ $\therefore q = 0.5$

By Binomial Distribution

$P(r \geq 8) = P(8) + P(9) + P(10)$

$= {}^{10}C_8\, p^8 q^2 + {}^{10}C_9\, p^9 q + {}^{10}C_{10}\, p^{10} q^0 = 0.055$

Q. 44 In a Binomial distribution mean = 2 and mean + variance = 3 then $P(r \geq 2)$ is_____.

(a) 0.531 (b) 0.688 (c) 0.259 (d) 0.317

Ans. : (b)

Explanation :

mean $= np = 2$ and variance $= npq$

$\therefore$ mean + variance $= 3 \Rightarrow$ variance $= 1$

$\therefore \dfrac{npq}{np} = \dfrac{1}{2} \Rightarrow q = 0.5$

$\therefore p = 0.5$ and $n = 4$

By Binomial Distribution,

$P(r \geq 2) = P(2) + P(3) + P(4) = 1 - [P(0) + P(1)]$

$= 1 - \left[{}^4C_0\, p^0 q^4 + {}^4C_1\, p^1 q^3\right]$

$P(r \geq 2) = 0.688$

Q. 45 In a Binomial distribution mean = 2 and mean + variance = 3 then $P(r \leq 2)$ is ____.

(a) 0.531 (b) 0.317 (c) 0.688 (d) 0.259

Ans. : (c)

Explanation : mean $= np = 2$ and variance $= npq$

$\therefore$ mean + variance $= 3 \Rightarrow$ variance $= 1$

$\therefore \dfrac{npq}{np} = \dfrac{1}{2} \Rightarrow q = 0.5$

$\therefore p = 0.5$ and $n = 4$

By Binomial Distribution, $P(r \leq 2) = P(0) + P(1) + P(2)$

$= {}^4C_0\, p^0 q^4 + {}^4C_1\, pq^3 + {}^4C_2\, p^2 q^2 = 0.688$

Q. 46 One percent of articles from a certain machine are defective then the probability of one or more defective in a sample of 100 is _____.

(a) 0.63 (b) 0.36 (c) 0.71 (d) 0.5

Ans. : (a)

Explanation : Given : $n = 100$, $p = 1\% = 0.01$; mean $= m = np = 1$

By Poission's Distribution, $P(r) = \dfrac{m^r e^{-m}}{r!}$

$\therefore$ The probability of one or more defective $= P(x \geq 1)$

$= P(1) + P(2) + \ldots\ldots + P(100) = 1 - [P(0)]$

$= 1 - \left[\dfrac{m^0 e^{-m}}{0!}\right] = 0.632$

Q. 47 The average number of misprints per page of a book is 1.5 then probability of book free from misprints using Poission's distribution is _____.

(a) 0.3141 (b) 0.7165 (c) 0.5482 (d) 0.2231

Ans. : (d)

Explanation : mean $(m) = 1.5$

By Poission's Distribution,

The probability of a book free from misprints

$P(0) = \dfrac{m^0 e^{-m}}{0!} = 0.2231$

Q. 48 The average number of misprints per page of a book is 1.5 then number of pages containing more than one misprints if book contains 900 pages using Poission's distribution are,

(a) 802 (b) 705 (c) 699 (d) 578

Ans. : (c)

Explanation : mean (m) = 1.5

By Poission's Distribution,

The probability of a book containing more than one misprints = P(1) + P(2) + ...

$$= 1 - P(0) = 1 - \left[\frac{m^0 e^{-m}}{0!}\right] = 0.7768$$

No. of pages = 900 × 0.78 = 699.12 ≈ 699

Q. 49 Number of road accidents follows a Poission's distribution with mean 5. In a certain month on the highway the probability of less than 3 accidents is _____.

(a) 0.315 (b) 0.231 (c) 0.721 (d) 0.125

Ans. : (d)

Q. 50 Number of road accidents follows a Poission's distribution with mean 5. In a certain month on the highway the probability of more than 2 accidents is _____

(a) 0.875 (b) 0.357 (c) 0.523 (d) 0.125

Ans. : (a)

Explanation : Mean = m = 5

Probability of more than 2 accidents

$$= P(3) + P(4) + ... = 1 - [P(0) + P(1) + P(2)]$$

$$= 1 - \left[\frac{m^0 e^{-m}}{0!} + \frac{m e^{-m}}{1!} + \frac{m^2 e^{-m}}{2!}\right] = 0.875$$

Q. 51 Mean and variance of binomical distribution are 6 and 2 respectively. Then p(r = 1) is _____.

(a) 0.253 (b) 0.111 (c) 0.999 (d) 0.789

Ans. : (c)

Explanation : mean = m = np = 6 and variance = $(S.D)^2 = npq = 2$

$$\therefore \frac{npq}{np} = \frac{2}{6} \Rightarrow q = \frac{1}{3}$$

$$\therefore p = 1 - q = \frac{2}{3} \text{ and } n = \frac{6}{p} = 9$$

$$\therefore P(r \geq 1) = 1 - P(r = 0) = 1 - [{}^9C_0\, p^0 q^9] = 0.999$$

Q. 52 In a certain factory manufacturing condensers there, is a chance that 1% of the condenses are defective. The condenser are packed in a boxes of 100 then using poission's distribution the number of boxes contains 3 faulty condensers.

(a) 6 (b) 7 (c) 3 (d) 1 **Ans. : (a)**

Explanation :

p = 1% = 0.01 and n = 100

$\therefore$ mean = m = np = 1

By Poission's Distribution

$$\therefore P(r = 3) = \frac{m^3 e^{-m}}{3!} = 0.0613$$

Number of boxes containing 3 faulty condensers = 0.0613 × 100 = 6.13 ≈ 6

Q. 53 In a certain factory manufacturing condensers there is a chance that 1% of the condenses are defective. The condensers are packed in a boxes of 100 then using poisson distribution number of boxes contains more than 2 faulty condensers = _____.

(a) 80 (b) 8 (c) 4 (d) 3

Ans. : (b)

Explanation : P = 1% = 0.01 and n = 100

mean = m = np = 1

By Poission's Distribution,

$$P(r > 2) = P(3) + P(4) + ... = 1 - [P(0) + P(1) + P(2)]$$

$$= 1 - \left[\frac{m^0 e^{-m}}{0!} + \frac{m e^{-m}}{1!} + \frac{m^2 e^{-m}}{2!}\right]$$

$$= 1 - e^{-m}\left[1 + m + \frac{m^2}{2}\right] = 0.08030$$

Number of boxes containing more than 2 faulty condensers = 0.08030 × 100 = 8.03 ≈ 8

Q. 54 If the probability of bad reaction from a certain injection is 0.001 then the chance that out of 2000 individuals more than 2 will get a bad reaction is _____.

(a) 0.999 (b) 0.231 (c) 0.677 (d) 0.3233

Ans. : (d).

Q. 55 If a random variable has a Poission's distribution such that P(1) = P(2) then mean of the distribution is _____.

(a) 2 (b) 3 (c) 4 (d) 5

Ans. : (a)

Q. 56 In a random variable has a Poission Distribution such that P(1) = P(2) then P(4) is _____.

(a) 0.20 (b) 0.71 (c) 0.09 (d) 0.01

Ans. : (c)

Q. 57 In a Poission distribution p(r = 2) is two third of P(r = 1) then P(r = 0) is _____.

(a) 0.0321 (b) 0.7287 (c) 0.9132 (d) 0.2635

Ans. : (d)

Q. 58 A certain screw making machine produces on average of 2 defective screws in 100 and packs them in boxes of 500 then the probability that a box contains 15 defective screws is _____.

(a) 0.5 (b) 0.035 (c) 0.3 (d) 0.7

Ans. : (b)

Explanation : $p = \frac{2}{100} = 0.02$

$\therefore$ mean = m = np

$\therefore$ mean = m = 500 × 0.02 = 10

By poission's Distribution,

$$P(r = 15) = \frac{m^{15} e^{-m}}{15!} = 0.035$$

Q. 59 If x is a normal variable with mean 30 and S.D. 5, the probability of (26 ≤ x ≤ 40) is _____.

[Given : z = 0.8; A = 0.2881; z = 2; A = 0.4772]

(a) –0.7653 (b) –0.1891

(c) 0.7653 (d) 0.1891

Ans. : (c)

Explanation :

Given mean = μ = 30 and S.D. = σ = 5

$$\therefore P(26 \leq x \leq 40) = P\left(\frac{26-\mu}{\sigma} \leq z = \frac{x-\mu}{\sigma} \leq \frac{40-\mu}{\sigma}\right)$$

$$= P(-0.8 \leq z \geq 2)$$

$$= P(-0.8 \leq z \leq 0) + P(0 \leq z \leq 2)$$

$$= 0.2881 + 0.4772 = 0.7553$$

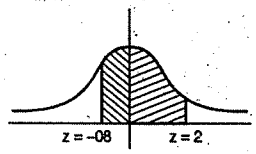


Fig. Q. 59

Q. 60 If x is a normal variable with mean 30 and S.D. 5 the probability of (x ≥ 45) is _____.

[Given : z = 3; A = 0.4986]

(a) 0.4986 (b) 0.1532

(c) 0.0014 (d) 0.32 **Ans. : (c)**

Explanation :

Given mean μ = 30 and S.D. = σ = 5

$$P(x \geq 45) = P\left(z \geq \frac{45-\mu}{\sigma}\right) \quad \left(\because z = \frac{x-\mu}{\sigma}\right)$$

$$= P(z \geq 3) = 0.5 - P(0 \leq z \geq 3)$$

$$= 0.5 - 0.4986 = 0.0014$$

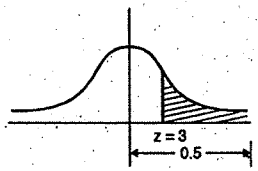


Fig. Q. 60

Chapter Ends...

□□□

CHAPTER 4

UNIT IV

Correlation

Syllabus

Syllabus

Introduction, Types of Correlation, Correlation and Causation, Methods of studying correlation, Karl Pearson's Correlation Coefficient, Spearman's Rank Correlation Coefficient, Properties of Karl Pearson's correlation coefficient and Spearman's rank correlation coefficient, probable Errors.

4.1 Introduction

- Statistics is the study of how to collect, organize, analyze and interpret the information about one or more population (s) under investigation.
- Recently, the use of statistics has been increased in human activity, particularly in Engineering and Science. Everything dealing with the collection processing, analysis and interpretation of numerical data which is used in statistical survey. It is believed that statistics is in use from the time when man began to count and measure.
- It is considered that the word statistics derived from the Italian word "Statista" or the Greek word "Statistika". In olden days there was a king and his kingdom is so large.
- He puts the records of land occupied, land revenue, manpower, roads, crops required and so many thing. Such records is considered as statistics.
- Statistics is nothing but the branch of science where we learn the various methods of collecting data, processing and different methods of analyzing it and the interpreting the data. (Collecting → Processing / analyzing → Interpreting) which are useful for developing the modern statistical techniques.

In statistics, we use the terms like,

1. Variable / Variate
2. Ungrouped data (row data)
3. Grouped data (classified data)

There are two methods of classifications of data

(i) Classification by class intervals or frequency distribution.

(ii) Classification by attributes.

4. **Frequency :** Number of observations in a particular class interval is called frequency of that class.

After arranging the data in ascending or descending order, to make it more compact, it is presented in a tabular form consisting of columns headed by symbols x and f. The column headed by x consists of various observations recorded and the column headed by f consists of frequency.

4.1.1 Measures of Central Tendency

Central Tendency

- In general, frequency distribution gives the data around at some central values. These central values represents the whole group.
- So it is very essential to find out these central values. Many physical processes generates numbers in the form of tabular data, graphs. (Different methods gives different averages which are called measure of central tendency).
- Generally, the central values are mean, median, mode, geometric mean and harmonic mean.
- Naturally, we select a representative observation from the central part, this is referred to as an **average** or **measure of central tendency.**
- It is important that all the major properties of the observations in the data should be represented in the average. Average is a value around which most of the observations are clustered, hence this single value itself gives clear idea regarding phenomenon (the concept) under study.

Averages

- Mathematical
  - Arithmetic mean (Weighted)
  - Geometric mean
  - Harmonic mean
- Others
  - Median (positional)
  - Mode

4.1.2 Arithmetic Mean (or Simply) Mean (A. M.)

We denote mean by letter $\bar{x}$.

(i) Let, $x_1, x_2, x_3, \ldots, x_n$, be the set of observations.

$$\text{Mean}\left(\bar{x}\right) = \frac{\text{sum of observations}}{\text{total number of observations}}$$

$$\therefore \bar{x} = \frac{\sum x_i}{n}$$

(ii) For grouped data : (Discrete frequency distribution)

Let, $x_1, x_2, \ldots, x_n$ be 'n' observations

And $f_1, f_2, f_3, \ldots, f_n$ be the corresponding frequencies

$$\therefore \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{N}$$

(iii) **For continuous frequency distribution :** In this case frequency is associated to the entire class and not to any specific single value. But for calculation purpose, we make a reasonable assumption that the frequency is associated to the midpoint of class or equivalently we say that the frequency is distributed over the respective class uniformly.

Let $d_i = x_i - a$ is called deviation of x_i from 'a'; Here 'a' is any point from x_i (generally we take mid value of x_i).

$$\text{Mean}\left(\bar{x}\right) = a + \frac{\sum f_i d_i}{N}$$

(iv) **Step-Deviation Method :** If 'h' be the length (width) of interval of class

$$\text{Mean}\left(\bar{x}\right) = a + h\left[\frac{\sum f_i u_i}{N}\right] \text{ where, } u_i = \frac{x_i - a}{h}$$

☞ **Mean**

- The mean or average of data is the sum of given numbers divided by total count.
- The mean of a set of numbers need not be one of the number in the set.
- $\left(\bar{x} = \frac{2+3+5}{3} = 3.333\right)$. In computing mean, some numbers may occurs more than once in the data.
- The algebraic sum of the deviations of all the variables about their mean is zero. $\left(\sum f_i\left(x_i - \bar{x}\right) = 0\right)$.
- Sometimes, the higher and lower numbers are deleted and the other remaining numbers are averaged.
- The other measure of centres are median, mode, quartile ...etc. The mean gives the centre.

4.1.3 Effect of Change the Origin and Scale

1. If we add (or subtract) a constant from each observations, then we say that the origin is changed. Similarly, if each observation is multiplied (or divided) by some constant then we say that the scale is changed.
2. Let $\bar{x}_1$ be the arithmetic mean of first group of size n_1 and $\bar{x}_2$ be the arithmetic mean of second group of size n_2 then the combined mean $\bar{x}_c$ is

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

The above result can be generalized.

4.1.4 Dispersion

- We have seen that average value condenses (or merge) into a single value. However average is not sufficient to describe the distribution completely.
- There may be two distributions with same means but the distributions may not be identical.

Ex.							A.M.
1.	A :	48	49	50	51	52	50
2.	B :	30	35	50	65	70	50
3.	C :	0	15	45	95	95	50

- Here, we observe that the average marks of all A, B, C are same but they are different in variation. Clearly, A is more consistent than B because A. M. is very close to other values.
- Whereas, B is more consistent than C cause observations in C are scattered or dispersed from central value. This variation is called 'dispersion'.
- Average remains good representative if dispersion is less (i.e. if observations are close to it). Thus dispersion decides the reliability of average. **(When deviation is small number, the average is good.)**

To measure the scatteredness of data from mean, we use

(i) Range (ii) Quartile deviation

(iii) Mean deviation (iv) Standard deviation

4.1.5 Standard Deviation (S.D)

- It is very important and powerful technique to measure the data.
- It is defined as 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 symbol 'σ'.

(i) Standard deviation $(\sigma) = \sqrt{\frac{1}{n}\sum\left(x_i-\bar{x}\right)^2} \leftarrow$ for individual observation

OR

$$= \sqrt{\frac{\sum x_i^2}{n} - \left(\bar{x}\right)^2}$$

(ii) $\sigma = \sqrt{\frac{1}{N}\sum f_i\left(x_i-\bar{x}\right)^2} \leftarrow$ for frequency distribution

OR

$$= \sqrt{\frac{\sum f_i x_i^2}{N} - \left(\bar{x}\right)^2}$$

(iii) $$\sigma = \sqrt{\left(\frac{\sum f_i d_i^2}{N}\right) - \left(\frac{\sum f_i d_i}{N}\right)^2}$$

(iv) $$\sigma = h\sqrt{\left(\frac{\sum f_i u_i^2}{N}\right) - \left(\frac{\sum f_i u_i}{N}\right)^2}$$

where, $u_i = \frac{d_i}{h}$ and $d_i = x_i - a$, $N = \sum f_i$

The square of standard deviation is called variance and it is denoted by σ^2.

- Moreover, among all the measures of dispersions, Standard deviation is affected by least fluctuations of sampling and which satisfies most of the requisites of good measure.
- [Greater standard deviation means greater the dispersion of the values of data (means, data spread out more). **The group having less standard deviation, is more reliable]**

4.1.6 Coefficient of Variation

Prof. Karl Pearson suggested the relative measure of standard deviation. It is called co-efficient of variation (C.V.)

Co-efficient of variation (C.V.) $= \frac{S.D.}{|A.M.|} \times 100$

$$= \frac{\sigma}{|\bar{x}|} \times 100\%$$

Co-efficient of variation (C.V.) is always expressed in percentage

- 100 number is multiplied for convenience.
- For comparison of the variability of two objects, we calculate the Co-efficient of variation (C.V.) of each object. The object having less Co-efficient of variation (C.V.) is more consistent than object having greater Co-efficient of variation (C.V.).

  (i.e. use Co-efficient of variation (C.V) to compare two different data).
- Almost all the industries reduced the Co-efficient of variation (C.V.) of their goods to considerable extent in last 50 years because of competition.
- Earlier the Japanese industrial product and American industrial product have same average quality, however there was considerable difference in Co-efficient of variation (C.V.)
- Co-efficient of variation (C.V.) of Japanese goods was less than 5 times than that of American goods.
- As a result of low co-efficient of variation (C.V.), the Japanese goods were more popular. In pharmaceutical industries, Co-efficient of variation (C.V.) is as low as 1 or less than 1. The variation in weight of tablets is almost negligible.

Note : For solving the examples, we can use any S.D. formula. (As per convenience)

Syllabus Topic : Correlation and Causation

4.2 Correlation and Causation :

- Correlation is a statistical tool which helps to measure and analyze the degree of relationship between two variables.
- The degree of relationship between the variables under consideration is measure through the correlation analysis.
- Many times we come across situations where two variables are inter-related.
- Correlation analysis deals with the association between two or more variables.
- Causation means cause & effect relation. If two variables vary in such a way that movement in one are affected by movement in other, these variables are called cause and effect relationship.

☞ **Ex. 1**

Sachin always gets good score in mathematics test when it is sunny day. One day he says to his friend, 'I am going to appear mathematics test because today is sun shine clearly'. His friend replied , Sachin , not because of sunny day, but you studied mathematics that's why you are going for mathematics test and getting good score. Outside sunny atmosphere nothing to do with it. Means, sunniness does not cause you to get good score in mathematics test.

Here, two events sunniness and Sachin giving test, occurs together without one causing the other. In other words, the two events are correlated in some way but there is no causal relation between them. Simply because two relations are correlated does not mean that one causal the other.

☞ **Ex. 2**

The students with higher grades in colleges have higher grades in high school.

☞ **Ex. 3**

People who takes vitamin C , recover more quickly from cold.

These are correlations. i.e. there is correspondence between these events.

In case of example.2, many times, the students having higher grades in high school, not getting more grades in colleges or vice versa. Now it is incorrect that earning high grades in high school always causes someone to earn higher grades in colleges.(in some cases ,it is correct). Same condition with example 3.

So we agree that there are cases of correlation without causation.

Causation always implies correlation but converse may not be true. In bivariate distribution x and y changes such that the change in one variable affects a change in the other variables, then these variables are said to be correlated, and this relationship is called correlation. e.g.: Relation between income and expenditure; relations between price and demand etc.

Note : Remember (x_i, y_i) is an ordered pair. First component in every pair is a observation on variable x and second component in every pair is a observation on variable y. Also the components x_i and y_i are inseparable, i.e. we cannot rearrange the pairs like (x_i, y_j).

Syllabus Topic : Types of Correlation

4.3 Types of Correlation

There are some types of Correlation especially relate to different types of causation :

(i) **Positive correlation :** If an increase (or decreases) in one variable corresponds to an increases (or decrease) in other variables, then the correlation is said to be positive or directive correlation. In this case events frequently occurs together.

(ii) **Negative correlation :** If increase (or decrease) in one variable corresponds to the decreases (or increases) in other variables then the correlation is called negative or inverse correlation. In this case when one event occurs, the other will not occurs.

If there is no relations between two variables, then these variables are said to be independent.

(iii) **Perfect correlation :** If two variables, vary in such a way that their ratio is always constant, then the correlation is called perfect correlation.

Other types such as, non-linear or curvilinear correlation, simple, partial and multiple correlation.

The graphical representation of the bivariate distribution (x_i, y_i) on a graph by means of points, is called, scatter or DOT diagram.

(iv) **Linear correlation :** Correlation is called linear when the ratio of amount of change in one variable with the amount of change in the other variable is a constant .

The graph of the variables having a linear relationship will form a straight line. Ex X = 1, 3, 5, 7 , Y = 5, 9, 13, 17, Y = 3 + 2x

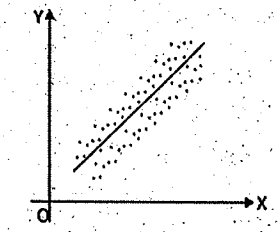


Fig. 4.3.1

(v) **Non Linear correlation :** The correlation is said to be non linear if the change in one variable does not a constant ratio to the change in the other variable.

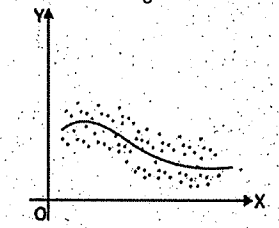


Fig. 4.3.2

4.3.1 Covariance.

Let us consider n pair of observations of two series X and Y, such as $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$, where $x_1, x_2, x_3 \ldots x_n$ are observed values of X and $y_1, y_2, y_3, \ldots, y_n$ are observed values of Y.

The deviation of X-values from their mean $\bar{x}$ are $\left(x_1-\bar{x}\right), \left(x_2-\bar{x}\right), \left(x_3-\bar{x}\right), \ldots, \left(x_n-\bar{x}\right)$.

Similarly, the deviation of Y - values from their mean $\bar{y}$ are, $\left(y_1-\bar{y}\right), \left(y_2-\bar{y}\right), \left(y_3-\bar{y}\right), \ldots, \left(y_n-\bar{y}\right)$

Then the sum of n products divided by n, is defined as covariance of x and y and is denoted by Cov (x, y).

$\therefore$ Cov (x, y)

$$= \frac{\left(x_1-\bar{x}\right)\left(y_1-\bar{y}\right) + \left(x_2-\bar{x}\right)\left(y_2-\bar{y}\right) + \ldots + \left(x_n-\bar{x}\right)\left(y_n-\bar{y}\right)}{n}$$

$$= \frac{1}{n}\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right) \quad \ldots(4.3.1)$$

$$= \frac{1}{n}\left[\sum\left(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y}\right)\right]$$

$$= \frac{1}{n}\sum x_i y_i - \bar{y}\frac{\sum x_i}{n} - \bar{x}\frac{\sum y_i}{n} + \frac{\bar{x}\bar{y}(n)}{(n)}$$

$$= \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y}$$

$$\boxed{\text{Cov}(x, y) = \frac{1}{n}\sum (x_i y_i) - \bar{x}\bar{y}} \quad \ldots(4.3.2)$$

4.3.1.1 Properties of Covariance

(i) Cov (x, x) = Var (x)

(ii) Effect of change of origin :

Cov (x – a, y – b) = Cov (x, y),

where a, b are constants.

(iii) Effect of change of scale.

$$\text{Cov}\left(\frac{x-a}{h}, \frac{y-b}{k}\right) = \frac{1}{hk}\text{Cov}(x, y),$$

where a, b, h, k are constants.

(iv) Cov (aX + b, cY + d) = ac Cov (x, y)

(v) Cov (X + Y, Z) = Cov (X, Z) + Cov (Y, Z)

(vi) If x and y are independent variables then Cov (x, y) = 0

(vii) Cov (x, y) = Cov (y, x)

(viii) Cov (x, y) may be negative value.

(ix) Co-variance is a joint central moment of order (1, 1) of (X, Y)

$\therefore \mu_{11} = \text{Cov}(x, y)$

Syllabus Topic : Methods of Studying Correlation

4.4 Methods of Studying Correlation :

When the change in one variable (x) affects change in other variable (y). We say that x causes y. For example, does the burning fossil fuel causes global warming ? We can determine correlation by different methods.

(i) Scatter diagram.

(ii) Karl Pearson's Co-efficient of Correlation.

(iii) Spearman's Rank Correlation Co-efficient,

(iv) Method of Least Square

4.4.1 Scatter Diagram Method :

It is a very simplest method to find the correlation between two variables. Let us consider n pair of values (xi , yi) of variable of X and Y. In this method we can plot the points on XY planes as dots, diagram so obtained is called scatter diagram.

(i)

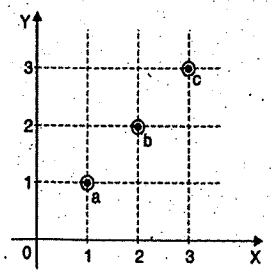


Fig. 4.4.1

$$\bar{x} = \frac{1+2+3}{3} = \frac{6}{3} = 2$$

$$\bar{y} = \frac{1+2+3}{3} = \frac{6}{3} = 2$$

$$\text{Cov}(x, y) = \frac{1}{3}\sum_{1}^{3}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{3}\Big[(x_1-\bar{x})(y_1-\bar{y}) + (x_2-\bar{x})(y_2-\bar{y}) + (x_3-\bar{x})(y_3-\bar{y})\Big]$$

$$= \frac{1}{3}[(1-2)(1-2) + (2-2)(2-2) + (3-2)(3-2)]$$

$$= \frac{1}{3}[(-1)(-1) + (0)(0) + (1)(1)]$$

$$\text{Cov}(x, y) = \frac{1}{3}[1+0+1] = \frac{2}{3} \quad \textbf{Positive Correlation.}$$

(ii)

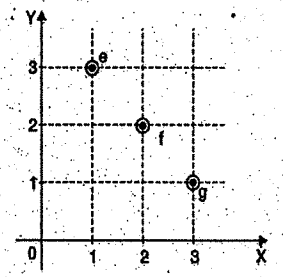


Fig. 4.4.2

$$\boxed{\begin{aligned}\bar{x} &= 2\\ \bar{y} &= 2\end{aligned}}$$

$$\text{Cov}(x, y) = \frac{1}{3}\sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{3}\Big[(x_1-\bar{x})(y_1-\bar{y}) + (x_2-\bar{x})(y_2-\bar{y}) + (x_3-\bar{x})(y_3-\bar{y})\Big]$$

$$= \frac{1}{3}[(1-2)(3-2) + (2-2)(2-2) + (3-2)(1-2)]$$

$$= \frac{1}{3}[(-1(1) + (0)(0) + (1)(-1)]$$

$$\text{Cov}(x, y) = \frac{1}{3}[-1+0-1]$$

$$\text{Cov}(x, y) = -\frac{2}{3} \quad \textbf{Negative Correlation.}$$

(iii)

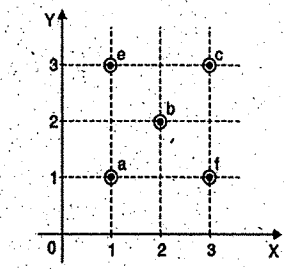


Fig. 4.4.3

$$\boxed{\begin{aligned}\bar{x} &= 2\\ \bar{y} &= 2\end{aligned}}$$

$$\text{Cov}(x, y) = \frac{1}{5}\sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{5}\Big[(x_1-\bar{x})(y_1-\bar{y}) + (x_2-\bar{x})(y_2-\bar{y}) + (x_3-\bar{x})(y_3-\bar{y}) + (x_4-\bar{x})(y_4-\bar{y}) + (x_5-\bar{x})(y_5-\bar{y})\Big]$$

$$= \frac{1}{5}[(1-2)(1-2) + (2-2)(2-2) + (3-2)(3-2) + (1-2)(3-2) + (3-2)(1-2))$$

$$= \frac{1}{5}[(-1)(-1) + (0)(0) + (1)(1) + (-1)(1) + (1(-1))]$$

$$\text{Cov}(x, y) = \frac{1}{5}[1+0+1-1-1] = 0$$

There is no correlation.

Syllabus Topic : Karl Pearson's Correlation Co-efficient

4.5 Karl-Pearson's Correlation Co-efficient

Correlation co-efficient is calculated to study the extent or the degree of relationship between two variables x and y.

It is usually denoted by r (x, y) or r_{xy} or corr (x, y).

$\therefore$ Karl Pearson's co-efficient of correlation is defined as,

$$r(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \cdot \sigma_y} \quad \ldots(4.5.1)$$

(i.e. r (x, y) = the ratio of the average product of deviations of X and Y to the product of their standard deviations)

$$\because \sigma_x^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{1}{n}\Big[\sum\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)\Big] = \frac{1}{n}\sum x_i^2 - \frac{2\bar{x}}{n}\sum x_i + \frac{\sum n\bar{x}^2}{n}$$

$$\sigma_x^2 = \frac{1}{n}\sum x_i^2 - (\bar{x})^2 \quad \ldots(4.5.2)$$

Similarly, $\sigma_y^2 = \frac{1}{n}\sum y_i^2 - (\bar{y})^2 \quad \ldots(4.5.3)$

$\therefore$ By using Equation (4.2.1) and definitions of σ_x and σ_y, Equation (4.5.1) becomes,

$$r(x, y) = \frac{\frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2 \times \frac{1}{n}\sum(y_i - \bar{y})^2}}$$

$$= \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\sqrt{\left[\frac{1}{n}\sum x_i^2 - (\bar{x})^2\right]\left[\frac{1}{n}\sum y_i^2 - (\bar{y})^2\right]}}$$

(by using Equations (4.4.1), (4.4.2) and (4.4.3))

Or $$r(x, y) = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum x_i^2 - n(\bar{x})^2\right)\left(\sum y_i^2 - n(\bar{y})^2\right)}}$$

Remember

(i) The denominator of r (x, y) is always positive, so the value of r (x, y) is positive or negative according as covariance.

Correlation coefficient r (x, y) does not change in magnitude under the change of origin and scale.

i.e. $r\left(\frac{x-a}{h}, \frac{y-b}{k}\right) = r(x, y)$; if both h and k have same sign.

$= -r(x, y)$; if h and k have opposite signs.

(ii) Let, $u_i = \frac{x_i - a}{h}$; $v_i = \frac{y_i - b}{k}$

If both h and k have same algebraic sign and let h = k = 1

$\therefore u_i = x_i - a$ and $v_i = y_i - b$

$\text{Cov}(x_i - a, y_i - b) = \text{Cov}(u_i, v_i)$

$= \frac{1}{n}\sum u_i v_i - \bar{u}\bar{v}$ ..... by the property.

$$\sigma_u^2 = \frac{1}{n}\sum u_i^2 - (\bar{u})^2$$

and $\sigma_v^2 = \frac{1}{n}\sum v_i^2 - (\bar{v})^2$, then

$$r(u, v) = \frac{\text{Cov}(u, v)}{\sigma_u \cdot \sigma_v}$$

(iii) For frequency distribution, i.e. if frequency is given then,

(1) $$\text{Cov}(x, y) = \frac{1}{N}\sum f_i(x_i - \bar{x})(y_i - \bar{y})$$

$$\sigma_x = \sqrt{\frac{1}{N}\sum f_i(x_i - \bar{x})^2} \; ;$$

$$\sigma_y = \sqrt{\frac{1}{N}\sum f_i(y_i - \bar{y})^2}$$

(2) Also, in terms of u_i and v_i.

$$\text{Cov}(u, v) = \frac{1}{N}\sum f_i u_i v_i - \bar{u}\,\bar{v}$$

$$\sigma_u = \sqrt{\frac{1}{N}\left(\sum f_i u_i^2\right) - (\bar{u})^2}$$

$$\sigma_v = \sqrt{\frac{1}{N}\left(\sum f_i v_i^2\right) - (\bar{v})^2}$$

Syllabus Topic : Properties of Karl Pearson's Correlation Coefficient

4.6 Properties of Karl Pearson's Correlation Coefficient

(i) $r(x, y)$ lies between -1 to 1 i.e. $-1 \le r \le +1$

(ii) If $r(x, y) > 0$, then correlation is positive.

(iii) If $r(x, y) < 0$, then correlation is negative and if $r(x, y) = 0$, then there is no correlation.

(iv) It determines a single value which summarizes the extent of linear relationships.

(v) A positive "linear" correlation between two variables x and y implies that high values of x are associated with high values of y, and that low values of x are associated with low values of y.

(vi) A negative correlation implies the opposite; high values of x are associated with low values of y.

(vii) By definition a "correlation coefficient" close to zero indicates little or no linear relationship between two variables.

(viii) Correlation coefficients close to 1 and -1 denote a strong positive or negative relationship between two variables.

Applications :

1. Correlation and regression is used in Agriculture (via fertilizers, irrigation etc.), in business and trades, share trading, in medical science, in economics, in industry to measure the work done and in many more fields.
2. It is used to measure the hardness of material.

4.6.1 Working Rule for Finding Correlation Coefficient

Step I : First write the formula (Write appropriate formula as per our convenience)

$$r(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} \quad \text{OR} \quad \frac{\text{cov}(u, v)}{\sigma_u \cdot \sigma_v}$$

Step II : Write $\text{Cov}(x, y) = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})$

or $\text{Cov}(x, y) = \frac{1}{n}\left(\sum x_i y_i\right) - \bar{x}\,\bar{y}$

or $\text{Cov}(x, y) = \frac{1}{N}\sum f_i (x_i - \bar{x})(y_i - \bar{y})$

or $\text{Cov}(x, y) = \frac{1}{N}\left(\sum f_i u_i v_i\right) - (\bar{u}\,\bar{v})$

or $\text{Cov}(x, y) = \frac{1}{n}\left(\sum u_i v_i\right) - \bar{u}\,\bar{v}$

Step III :

Write $\sigma_x = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$

or $\sigma_x = \sqrt{\frac{1}{n}\sum x_i^2 - (\bar{x})^2}$

Or $\sigma_x = \sqrt{\frac{1}{N}\sum f_i (x_i - \bar{x})^2}$

or $\sigma_u = \sqrt{\frac{1}{N}\sum f_i u_i^2 - (\bar{u})^2}$

or $\sigma_u = \sqrt{\frac{1}{n}\sum u_i^2 - (\bar{u})^2}$

Similarly, write the formula for σ_y or σ_v by replacing x by y; $\bar{x}$ by $\bar{y}$ and u_i by v_i in above formula

Step IV : Find $\bar{x} = \frac{\sum x_i}{n}$; $\bar{y} = \frac{\sum y_i}{n}$ ...if data is individual

or $\bar{x} = \frac{\sum f_i x_i}{N}$; $\bar{y} = \frac{\sum f_i y_i}{N}$ ...if frequency is given

or $\bar{u} = \frac{\sum f_i u_i}{N}$; $\bar{v} = \frac{\sum f_i v_i}{N}$ ...in terms of u and v if frequency given

or $\bar{u} = \frac{\sum u_i}{n}$; $\bar{v} = \frac{\sum v_i}{n}$ ...if data is individual

Where, $N = \sum f_i$

and n = number of observations / items

Step V : Then construct table according to your formula and find the appropriate values.

Solved Examples

Ex. 4.6.1 : Find the Karl-Pearson's coefficient of correlation for the data :

$n = 25$, $\sum x_i = 100$, $\sum y_i = 125$, $\sum x_i^2 = 250$, $\sum y_i^2 = 500$, $\sum x_i y_i = 522$.

Soln. :

Step I : The Karl-Pearson's coefficient of correlation

$$r(x, y) = \frac{\left(\sum x_i y_i - n\bar{x}\bar{y}\right)}{\sqrt{\left(\sum x_i^2 - n(\bar{x})^2\right) \times \left(\sum y_i^2 - n(\bar{y})^2\right)}}$$

($\because$ Frequency is not given)

Step II : Here, $\bar{x} = \frac{\sum x_i}{n} = \frac{100}{25} = 4$

$\bar{y} = \frac{\sum y_i}{n} = \frac{125}{25} = 5$

Step III : $\therefore r(x, y) = \frac{522 - 25 \times 4 \times 5}{\sqrt{(250 - 25 \times 16)(500 - 25 \times 25)}}$

$= \frac{522 - 500}{\sqrt{(-150)(-125)}} = \frac{22}{\sqrt{18750}} = \frac{22}{136.9306}$

$\boxed{r(x, y) = 0.160665}$ ✓ ...Ans.

It is a Positive correlation.

Ex. 4.6.2 : Find the correlation coefficient between x and y, given that :

$n = 50$, $\sum(x_i - 40) = 30$; $\sum(y_i - 20) = 70$; $\sum(x_i - 40)^2 = 170$; $\sum(y_i - 20)^2 = 165$; $\sum(x_i - 40)(y_i - 20) = 140$.

Soln. :

Step I :

Given : $n = 50$, $\sum(x_i - 40) = 30$; $\sum(y_i - 20) = 70$;

$\sum(x_i - 40)^2 = 170$; $\sum(y_i - 20)^2 = 165$;

$\sum(x_i - 40)(y_i - 20) = 140$

Step II : Here, data is given in the form of $\sum(x_i - a)$ and $\sum(x_i - b)$.

So we use deviation formula

Let $u_i = x_i - 40$; $v_i = y_i - 20$

$\therefore \bar{u} = \frac{\sum u_i}{n} = \frac{30}{50} = 0.6$;

$\bar{v} = \frac{\sum v_i}{n} = \frac{70}{50} = 1.4$

Step III : Coefficient of correlation

$$r(x, y) = \frac{\sum(u_i v_i) - n\bar{u}\bar{v}}{\sqrt{\left[\sum u_i^2 - n(\bar{u})^2\right] \times \left[\sum v_i^2 - n(\bar{v})^2\right]}}$$

($\because$ Frequency is not given)

$$= \frac{140 - 50 \times (0.6)(1.4)}{\sqrt{[(170) - 50(0.6)^2][(165) - 50(1.4)^2]}}$$

$$= \frac{98}{\sqrt{152 \times 67}} = \frac{98}{\sqrt{10184}} = \frac{98}{100.9158}$$

$\boxed{r(x, y) = 0.9711}$ ✓ ...Ans.

Ex 4.6.3 : Calculate the coefficient of correlation from the following information :

$n = 10$; $\sum x = 40$; $\sum x^2 = 190$; $\sum y_i^2 = 200$, $\sum xy = 150$; $\sum y = 40$

Soln. :

Step I : Given : $n = 10$; $\sum x^2 = 190$; $\sum y_i^2 = 200$, $\sum xy = 150$; $\sum y = 40$; $v\sum x = 40$

Step II : Here, $\bar{x} = \frac{\sum x_i}{n} = \frac{40}{10} = 4$

and $\bar{y} = \frac{\sum y_i}{n} = \frac{40}{10} = 4$

Step III : Since, the co-efficient of correlation

$$r(x, y) = \frac{\sum xy - n\bar{x}\,\bar{y}}{\sqrt{(\sum x_i^2) - n(\bar{x})^2} \cdot \sqrt{(\sum y_i^2) - n(\bar{y})^2}}$$

$$r(x, y) = \frac{150 - 10 \times (4 \times 4)}{\sqrt{(190) - 10 \times (4)^2} \times \sqrt{(200) - 10 \times (4)^2}}$$

$$= \frac{-10}{\sqrt{30} \times \sqrt{40}} = \frac{-10}{20\sqrt{3}} = \frac{-1}{2\sqrt{3}} = \frac{-1}{3.4641}$$

$= -0.2887$ ✓ ...Ans.

Here, the correlation is negative

Ex. 4.6.4 : If $r = 0.87$; $\sum x_i y_i = 90$; $\sigma_x = 4.3$, $\sum y_i^2 = 98$; Find n if x and y are deviations from arithmetic mean.

Soln. :

Step I : Given, x and y are deviations from arithmetic mean

i.e. given, $x_i = x_i - \bar{x}$ And $y_i = y_i - \bar{y}$, $r = 0.87$, $\sum x_i y_i = 90$, $\sigma_x = 4.3$, $\sum y_i^2 = 98$

Step II : Since $\sigma_y^2 = \frac{1}{n}\sum(y_i - \bar{y})^2 = \frac{1}{n}\sum y_i^2 = \frac{98}{n}$

(Since, $y_i - \bar{y} = y_i$)

$$\boxed{\sigma_y^2 = \frac{98}{n}}$$

Step III : Also,

$$r(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \cdot \sigma_y} = \frac{\frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \cdot \sigma_y}$$

$$r = \frac{\sum(x_i y_i)}{n\,\sigma_x \cdot \sigma_y}$$

Squaring both side,

$$r^2 = \frac{\sum(x_i y_i)^2}{n^2 \sigma_x^2 \cdot \sigma_y^2}$$

$$(0.87)^2 = \frac{(90)^2}{n^2 (4.3)^2 \left(\frac{98}{n}\right)}$$

$$\Rightarrow n = \frac{(90)^2}{(0.87)^2 (98)(4.3)^2} = \frac{8100}{1371.5179380}$$

$n = 5.9058$ $\boxed{n \approx 6}$ ✓ ...Ans.

Ex. 4.6.5 : Find the coefficient of correlation for the following values

x_i	10	14	18	22	26	30
y_i	18	12	24	6	30	36

Soln. :

Step I : We know, Karl Pearson's coefficient of correlation r (x, y)

$$r(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

$$= \frac{\frac{1}{n}\sum(x_i y_i) - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum x_i^2 - (\bar{x})^2}\sqrt{\frac{1}{n}\sum y_i^2 - (\bar{y})^2}}$$

and $\bar{x} = \frac{\sum x_i}{n}$; $\bar{y} = \frac{\sum y_i}{n}$ Here, $n = 6$

Step II : Construct the table as per the formula

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
10	18	180	100	324
14	12	168	196	144
18	24	432	324	576
22	6	132	484	36
26	30	780	676	900
30	36	1080	900	1296
$\sum x_i = 120$	$\sum y_i = 126$	$\sum x_i y_i = 2772$	$\sum x_i^2 = 2680$	$\sum y_i^2 = 3276$

Step III :

$$\therefore \quad \bar{x} = \frac{120}{6} = 20 \; ; \qquad \bar{y} = \frac{126}{6} = 21$$

$$r(x,y) = \frac{\frac{1}{6}(2772) - (20)(21)}{\sqrt{\frac{1}{6}(2680) - (20)^2}\sqrt{\frac{1}{6}(3276) - (21)^2}}$$

$$= \frac{462 - 420}{\sqrt{46.66} \times \sqrt{105}} = \frac{42}{\sqrt{4899.3}} = \frac{42}{69.9994}$$

$$\boxed{r(x,y) = 0.60} \checkmark$$

...Ans.

Ex. 4.6.6 : Calculate the coefficient of correlation for the following distribution :

x_i	5	9	15	19	24	28	32
y_i	7	9	14	21	23	29	30
f_i	6	9	13	20	16	11	7

Soln. :

Step I :

Here, frequency is given, so we use,

$$r(x,y) = \frac{Cov(u,v)}{\sigma_u \cdot \sigma_v}$$

$$= \frac{\frac{1}{N}\sum(f_i u_i v_i) - \bar{u}\bar{v}}{\sqrt{\frac{1}{N}\sum f_i u_i^2 - (\bar{u})^2}\sqrt{\frac{1}{N}\sum f_i v_i^2 - (\bar{v})^2}}$$

$$r(x,y) = \frac{\sum f_i u_i v_i - N\bar{u}\bar{v}}{\sqrt{\sum f_i u_i^2 - N(\bar{u})^2}\sqrt{\sum f_i v_i^2 - N(\bar{v})^2}}$$

And $\bar{u} = \frac{\sum f_i u_i}{N}$; $\bar{v} = \frac{\sum f_i v_i}{N}$; $N = \sum f_i$

$u_i = x_i - a$ and $v_i = y_i - b$

Step II : Now construct the table:

x_i	y_i	f_i	$u_i = x_i - a = x_i - 19$	$v_i = y_i - b = y_i - 21$	u_i	v_i	$f_i u_i$	$f_i v_i$	$f_i u_i^2$	$f_i v_i^2$	$f_i u_i v_i$
5	7	6	−14	−14	196	−84	−84	1176	1176	1176	
9	9	9	−10	−12	120	−90	−108	900	1296	1080	
15	14	13	−4	−7	28	−52	−91	208	637	364	
19	21	20	0	0	0	0	0	0	0	0	
24	23	16	5	2	10	80	32	400	64	160	
28	29	11	9	8	72	99	88	891	704	792	
32	30	7	13	9	117	91	63	1183	567	819	
		$\sum f_i = 82$				$\sum f_i u_i = 44$	$\sum f_i v_i = -100$	$\sum f_i u_i^2 = 4758$	$\sum f_i v_i^2 = 4444$	$\sum f_i u_i v_i = 4391$	

Step II : Put these table values in the formulae, we get,

$$\bar{u} = \frac{44}{82} = 0.5366 \; ;$$

$$\bar{v} = \frac{-100}{82} = -1.2195$$

$$r(x,y) = \frac{(4391) - (82)(0.5366)(-1.2195)}{\sqrt{[(4758) - 82(0.5366)^2]} \times \sqrt{[(4444) - 82(-1.2195)^2]}}$$

$$= \frac{(4391) + (53.6595)}{\left[\sqrt{(4758) - (23.6110)}\right] \times \left[\sqrt{(4444) - (121.9488)}\right]}$$

$$= \frac{(4444.6595)}{(68.8069) \times (65.7423)}$$

$$= \frac{4444.6595}{4523.5239} = 0.9826 \checkmark$$

...Ans.

(Here we can use another formula with frequency)

Ex. 4.6.7 : Calculate the coefficient of correlation for the following distribution :

x_i	35	34	40	43	56	20	38
y_i	32	30	31	32	53	20	33

Soln. :

Step I : Since $r(x,y) = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$

(Here, frequency is not given. We can use formula in terms of x and y)

OR $r(x,y) = \frac{Cov(u,v)}{\sigma_u \cdot \sigma_v}$

$$= \frac{\sum u_i v_i - n\bar{u}\bar{v}}{\sqrt{\left[\sum u_i^2 - n(\bar{u})^2\right] \times \left[\sum v_i^2 - n(\bar{v})^2\right]}}$$

$\bar{u} = \frac{\sum u_i}{n}$; $\bar{v} = \frac{\sum v_i}{n}$;

$u_i = x_i - a$; $v_i = y_i - b$; $n = 7$

Step II : Now construct the table :

x_i	y_i	$u_i = x_i - 43$	$v_i = y_i - 32$	$u_i v_i$	u_i^2	v_i^2
35	32	−8	0	0	64	0
34	30	−9	−2	18	81	4
40	31	−3	−1	3	9	1
43	32	0	0	0	0	0
56	53	13	21	273	169	441
20	20	−23	−12	276	529	144
38	33	−5	+1	−5	25	1
		$\sum u_i = -35$	$\sum v_i = 7$	$\sum u_i v_i = 565$	$\sum u_i^2 = 877$	$\sum v_i^2 = 591$

Step III : $\bar{u} = \frac{-35}{7} = -5$; $\bar{v} = \frac{7}{7} = 1$

$$r(x,y) = \frac{(565) - 7(-5)(1)}{\sqrt{[(877) - 7(-5)^2] \times [(501) - 7(1)^2]}}$$

$$= \frac{600}{\sqrt{702 \times 584}} = \frac{600}{640.2874}$$

$$\boxed{r(x,y) = 0.9370} \checkmark$$

...Ans.

Ex. 4.6.8 : Following are the values of import of row material and export of finished product in suitable units

Export material	10	11	14	14	20	22	16	12	15	13
Import material	12	14	15	16	21	26	21	15	16	14

Calculate coefficient of correlation between the import and export values.

Soln. :

Step I :

Since, $r(x,y) = \frac{Cov(u,v)}{\sigma_u \cdot \sigma_v}$

$$= \frac{\sum u_i v_i - n\bar{u}\bar{v}}{\sqrt{\left[\sum u_i^2 - n(\bar{u})^2\right] \times \left[\sum v_i^2 - n(\bar{v})^2\right]}}$$

$\bar{u} = \frac{\sum u_i}{n}$; $\bar{v} = \frac{\sum v_i}{n}$; $n = 10$

$u_i = x_i - a$; $v_i = y_i - b$

$= x_i - 20 = y_i - 21$

Step II : Now construct the table :

x_i	y_i	$u_i = x_i - 20$	$v_i = y_i - 21$	$u_i v_i$	u_i^2	v_i^2
10	12	−10	−9	90	100	81
11	14	−9	−7	63	81	49
14	15	−6	−6	36	36	36
14	16	−6	−5	30	36	25
20	21	0	0	0	0	0
22	26	2	5	10	4	25
16	21	−4	0	0	16	0
12	15	−8	−6	48	64	36
15	16	−5	−5	25	25	25
13	14	−7	−7	49	49	49
		$\sum u_i = -53$	$\sum v_i = 40$	$\sum u_i v_i = 351$	$\sum u_i^2 = 411$	$\sum v_i^2 = 326$

Step III : $\bar{u} = \frac{-53}{10} = -5.3$; $\bar{v} = \frac{-40}{10} = -4$

$$r(x,y) = \frac{(351) - 10(-5.3)(-4)}{\sqrt{[(411) - 10(-5.3)^2] \times [(326) - 10(-4)^2]}}$$

$$= \frac{351 - 212}{\sqrt{130.1 \times 166}} = \frac{139}{146.9578}$$

$$\boxed{r(x,y) = 0.945849} \checkmark$$

...Ans.

Ex. 4.6.9 : From a group of 10 students, marks obtained by each in two papers X and Y are given below :

x	23	28	42	17	26	35	29	37	16	46
y	25	22	38	21	27	39	24	32	18	44

Calculate Karl Pearson's co-efficient of correlation.

Soln. :

Step I : Since, Karl Pearson's co-efficient of correlation is

$$\therefore\ r(x, y) = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\sum x_i y_i - n(\bar{x}\bar{y})}{\sqrt{\sum x_i^2 - n(\bar{x})^2} \cdot \sqrt{\sum y_i^2 - n(\bar{y})^2}}$$

$$\bar{x} = \frac{\sum x_i}{n}\ ;\quad \bar{y} = \sum \frac{y_i}{n}\ ;\quad n = 10$$

Step II : Now construct the table :

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
23	25	575	529	625
28	22	616	784	484
42	38	1596	1764	1444
17	21	357	289	441
26	27	702	676	729
35	39	1365	1225	1521
29	24	696	841	576
37	32	1184	1369	1024
16	18	288	256	324
46	44	2024	2116	1936
$\sum x_i = 299$	$\sum y_i = 290$	$\sum x_i y_i = 9403$	$\sum x_i^2 = 9849$	$\sum y_i^2 = 9104$

Step III : $\bar{x} = \frac{299}{10} = 29.9$ and $\bar{y} = \frac{290}{10} = 29$

$$\therefore\ r(x, y) = \frac{(9403) - 10\,(29.9)\,(29)}{\sqrt{(9849) - 10\,(29.9)^2} \cdot \sqrt{(9104) - 10\,(29)^2}}$$

$$= \frac{9403 - 8671}{\sqrt{908.9} \cdot \sqrt{694}} = \frac{732}{\sqrt{630776.6}}$$

$\therefore\ r(x, y) = 0.92167$ ✓ ...Ans.

(We may use formula in terms of u and v)

Ex. 4.6.10 : Find the co-efficient of correlation between x and y from the table.

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

Soln. :

Step I : Karl Pearson's coefficient of correlation r (x, y) is,

$$\therefore\ r(x, y) = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum x_i^2 - (\bar{x})^2} \cdot \sqrt{\frac{1}{n}\sum y_i^2 - (\bar{y})^2}}$$

$$= \frac{\sum x_i y_i - n(\bar{x}\bar{y})}{\sqrt{\sum x_i^2 - n(\bar{x})^2} \cdot \sqrt{\sum y_i^2 - n(\bar{y})^2}}$$

and $\bar{x} = \frac{\sum x_i}{n}\ ;\quad \bar{y} = \frac{\sum y_i}{n}$; Here $n = 8$

Step II : Now construct the table :

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	1	1	1	1
3	2	6	9	4
4	4	16	16	16
6	4	24	36	16
8	5	40	64	25
9	7	63	81	49
11	8	88	121	64
14	9	126	196	81
$\sum x_i = 56$	$\sum y_i = 40$	$\sum x_i y_i = 364$	$\sum x_i^2 = 524$	$\sum y_i^2 = 256$

Step III : $\bar{x} = \frac{\sum x_i}{n} = \frac{56}{8} = 7$;

and $\bar{y} = \frac{\sum y_i}{n} = \frac{40}{8} = 5$

$$\therefore\ r(x, y) = \frac{364 - 8\,(7)\,(5)}{\sqrt{(524) - 8\,(7)^2} \cdot \sqrt{(256) - 8\,(5)^2}}$$

$$= \frac{364 - 280}{\sqrt{524 - 392} \cdot \sqrt{256 - 200}} = \frac{84}{\sqrt{132} \cdot \sqrt{56}}$$

$\therefore\ r(x, y) = 0.9770$ ✓ ...Ans.

Ex. 4.6.11 : Obtain the correlation coefficient between population density (per square miles) and death rate (per thousand persons) from the following data :

Population density	200	500	400	700	300
Death rate	12	18	16	21	10

Soln. :

Step I : Let, x_i = Population density

and y_i = Death rate

$\therefore$ Since, Karl Pearson's coefficient of correlation r (x, y) is,

$$r(x, y) = \frac{Cov(u, v)}{\sigma_u \cdot \sigma_v}$$

$$= \frac{\sum u_i v_i - n(\bar{u}\bar{v})}{\sqrt{\sum u_i^2 - n(\bar{u})^2} \cdot \sqrt{\sum v_i^2 - n(\bar{v})^2}}$$

$$\bar{u} = \frac{\sum u_i}{n}\ ;\quad \bar{v} = \frac{\sum v_i}{n}\ ;\quad n = 5$$

$u_i = x_i - a$; $v_i = y_i - b$

$a = 400$; $b = 16$

Step II : Now construct the table :

x_i	y_i	$u_i = x_i - a = x_i - 400$	$v_i = y_i - b = y_i - 16$	u_i^2	v_i^2	$u_i v_i$
200	12	−200	−4	40,000	16	800
500	18	100	2	10,000	4	200
400	16	0	0	0	0	0
700	21	300	5	90,000	25	1500
300	10	−100	−6	10,000	36	600
		$\sum u_i = 100$	$\sum v_i = -3$	$\sum u_i^2 = 1,50,000$	$\sum v_i^2 = 81$	$\sum u_i v_i = 3100$

Step III : $\bar{u} = \frac{100}{5} = 20$;

$\bar{v} = \frac{-3}{5} = -0.6$

$$\therefore\ r(x, y) = \frac{3100 - 5 \times (2) \times (-0.6)}{\sqrt{150000 - 5 \times (20)^2} \cdot \sqrt{(81) - 5 \times (-0.6)^2}}$$

$$= \frac{3160}{\sqrt{148000} \cdot \sqrt{79.2}}$$

$\therefore\ r(x, y) = 0.92298$ ✓ ...Ans.

Ex. 4.6.12 : Find coefficient of correlation for the following data :

X	6	2	10	4	8
Y	9	11	5	8	7

Soln. :

Step I : Since, the coefficient of correlation

$$r(x, y) = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\sum x_i y_i - n(\bar{x}\bar{y})}{\sqrt{\sum x_i^2 - n(\bar{x})^2} \cdot \sqrt{\sum y_i^2 - n(\bar{y})^2}}$$

Step II :

$$\bar{x} = \frac{\sum x_i}{n}\ ;\quad \bar{y} = \frac{\sum y_i}{n}\ ;\quad n = 5$$

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
$\sum x_i = 30$	$\sum y_i = 40$	$\sum x_i^2 = 220$	$\sum y_i^2 = 340$	$\sum x_i y_i = 214$

Step III :

Now, $\bar{x} = \frac{30}{5} = 6$; $\bar{y} = \frac{40}{5} = 8$

$$r(x, y) = \frac{214 - 5\,(6 \times 8)}{\sqrt{212 - 5(6)^2} \cdot \sqrt{(340 - 5\,(8)^2)}}$$

$$= \frac{220 - 240}{\sqrt{212 - 180} \cdot \sqrt{(340 - 320)}}$$

$$= \frac{-20}{(5.6568)\,(4.4721)} = \frac{-20}{25.2977}$$

$r(x, y) = -0.7905$ ✓ ...Ans.

Ex. 4.6.13 : Given $r = 0.9$, $\sum xy = 70$, $\sigma_y = 3.5$, $\sum x^2 = 100$. Find the number of item, if x and y are deviations from arithmetic mean.

Soln. :

Step I : Given, $r = 0.9$; $\sum xy = 70$;

$\sigma_y = 3.5$; $\sum x^2 = 100$

Also, x and y are deviations from arithmetic mean

i.e. given, $x_i = x_i - \bar{x}$

And $y_i = y_i - \bar{y}$

Step II :

$\therefore$ Since, $\sigma_x^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 = \frac{100}{n}$

(Since, $x_i - \bar{x} = x_i$)

$$\sigma_x^2 = \frac{100}{n}$$

Step IIII : and

$$r(x, y) = \frac{cov(x, y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \cdot \sigma_y}$$

$$\Rightarrow r = \frac{\sum (x_i y_i)}{n\,\sigma_x \cdot \sigma_y}$$

(Since, $x_i - \bar{x} = x_i$ and $y_i - \bar{y} = y_i$)

Step IV : Squaring both sides,

$$r^2 = \frac{\sum (x_i y_i)^2}{n^2 \sigma_x^2 \cdot \sigma_y^2}$$

$$\Rightarrow (0.9)^2 = \frac{(70)^2}{n^2\left(\frac{100}{n}\right)(3.5)^2}$$

$$\Rightarrow n = \frac{(70)^2}{(0.9)^2 \times 100 \times (3.5)^2} = \frac{(70)^2}{992.25}$$

$$\boxed{n = 4.9382 \approx 5} \checkmark \quad \text{...Ans.}$$

Syllabus Topic : Spearman's Rank Correlation Coefficient

4.7 Spearman's Rank Correlation Co-efficient

- Sometimes we come across the situation where data cannot be measure quantitatively but qualitative assessment is feasible.
- In general, the ranks in the two characteristics are different . When the qualitative data values are ranked in order of characteristics like beauty, intelligence, size, importance, smartness, honesty etc, using the numbers 1, 2, 3, ..., n. then use rank correlation.
- Spearman's rank correlation is one of the method to measure the correlation between two variables (characteristics) x and y . It gives linear relationship between two characteristics.
- When the corresponding values of characteristics x and y are assigned ranks serially and the correlation between the ranks of n items in two characteristics (say age and heights) is called **rank correlation.**
- The Spearman's rank correlation Co-efficient is denoted by r_{rank} or r_s and is defined as

$$r_s = r_{rank} = 1 - \left[\frac{6\sum d_i^2}{n(n^2-1)}\right] \quad \text{...(1)}$$

where, d_i = difference between ranks of corresponding two series x and y (i.e. $d = x_i - y_i$)

n = Number of observations.

Let $x_1, x_2, x_3, \ldots, x_n$ and $y_1, y_2, y_3, \ldots, y_n$ be the values of series x and y respectively.

Let 1, 2, 3, ..., n are the ranks assigned to n individual x_i and y_i (may be from lowest to largest or vice versa).

Since, ranks are integer value from 1 to n, therefore,

$$\text{Mean } (\bar{x}) = \frac{1+2+3+\ldots+n}{n} = \left[\frac{\frac{n(n+1)}{2}}{n}\right] = \left(\frac{n+1}{2}\right)$$

$$\left(\because \text{Standard result : } \sum_{i=1}^{m} i = 1+2+3+\ldots+m = \frac{m(m+1)}{2}\right)$$

and variance in x is

$$\sigma_x^2 = \frac{1}{n}\sum x_i^2 - (\bar{x})^2 = \left[\frac{1^2+2^2+3^2+\ldots+n^2}{n}\right] - \left[\frac{n+1}{2}\right]^2$$

Use standardization result :

$$\sum_{i=1}^{m} i^2 = 1^2+2^2+3^2+\ldots+n^2 = \frac{m(m+1)(2m+1)}{6}$$

$$\therefore \sigma_x^2 = \left[\frac{\cancel{n}(n+1)(2n+1)}{6\cdot\cancel{n}}\right] - \left[\frac{n+1}{2}\right]^2$$

$$= \left[\frac{2n^2+n+2n+1}{6}\right] - \left[\frac{n^2+2n+1}{4}\right]$$

$$= \frac{4(2n^2+3n+1) - 6(n^2+2n+1)}{24}$$

$$= \frac{8n^2+12n+4-6n^2-12n-6}{24}$$

$$\sigma_x^2 = \frac{2n^2-2}{24} = \frac{n^2-1}{12}$$

Hence, $\bar{x} = \frac{n+1}{2}$; $\quad \sigma_x^2 = \frac{n^2-1}{12}$

Similarly

$\bar{y} = \frac{n+1}{2}$; $\quad \sigma_y^2 = \frac{n^2-1}{12}$

Let, x_i and y_i be the ith ranks of individual x and y

$\therefore$ $d_i = (x_i - y_i)$ be the deviations between the ranks

$$\sum x_i = \sum y_i = 1+2+3+\ldots+n = \frac{n(n+1)}{2}$$

$$\sum x_i^2 = \sum y_i^2 = 1^2+2^2+3^2+\ldots+n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$d_i = (x_i - y_i) = (x_i - \bar{x}) - (y_i - \bar{y}) \quad \text{...}(\because \bar{x} = \bar{y})$$

$$\therefore \frac{1}{n}\sum d_i^2 = \frac{1}{n}\sum [(x_i-\bar{x}) - (y_i-\bar{y})]^2$$

$$= \frac{1}{n}\left[\sum (x_i-\bar{x})^2 + \sum (y_i-\bar{y})^2 - 2\sum (x_i-\bar{x})\cdot(y_i-\bar{y})\right]$$

$$= \frac{1}{n}\sum (x_i-\bar{x})^2 + \frac{1}{n}(y_i-\bar{y})^2 - \frac{2}{n}\sum (x_i-\bar{x})\cdot(y_i-\bar{y})$$

$$\frac{1}{n}\sum d_i^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y \quad \text{...}\left(\because r = \frac{\text{cov}(x,y)}{\sigma_x \cdot \sigma_y}\right)$$

$$\frac{1}{n}\sum d_i^2 = 2\sigma_x^2 - 2r\sigma_x^2 \quad (\because \sigma_x = \sigma_y)$$

$$= 2(1-r)\sigma_x^2$$

$$= 2(1-r)\frac{(n^2-1)}{12}$$

$$\frac{\sum d_i^2}{n} = (1-r)\frac{(n^2-1)}{6}$$

$$\Rightarrow 1 - r = \frac{6\sum d_i^2}{n(n^2-1)}$$

Hence $r = 1 - \left[\frac{6\sum d_i^2}{n(n^2-1)}\right]$

is called Spearman's rank correlation coefficients

$$\sum d_i = \sum (x_i - y_i) = \sum x_i - \sum y_i$$

$$\sum d_i = 0 \quad \text{(always)}$$

- For Spearman's rank correlation co-efficient, function should be monotonic.
- (Monotonically increasing or decreasing. i.e. positive or negative correlation .)

Syllabus Topic : Properties of Spearman's Rank Correlation Co-efficient

4.8 Properties of Spearman's Rank Correlation Co-efficient

(i) It can be used for any data.

(ii) Rank correlation co-efficient lies between –1 to +1. i.e. $-1 \le r \le +1$.

(iii) If $r_s = +1$, then both ranks are move in same direction.

(iv) If $r_s = -1$, then both ranks are move in opposite direction.

(v) If $r_s = 0$, then there is no correlation.

(vi) $\sum d_i = 0$ always.

(vii) It is easy to apply as compare with Pearson's Coefficient.

(viii) It is useful for qualitative data when we can assign rank.

(ix) It can't be used in grouped frequency distribution.

4.8.1 Working Rule to Solve the Problems on Spearman's Rank Correlation Co-efficient

If rank of each series variable is given then

(i) Calculate the difference ' d ' of two ranks. i. e. find $d_i = x_i - y_i$

(ii) Square the difference and calculate the sum. i.e. calculate $\sum d_i^2$.

(iii) Substitute the obtained values in Spearman's Rank Correlation Co-efficient formula.

Note :

If Ranks are not given and all series values are distinct, then we need to assign rank to the given data (series) The lowest value in the series can be assign rank 1 and 2 to the next smallest and so on up to a rank of n assigned to the largest. (we may assign ranks in reverse order also.)

i.e. we can assign ranks to the items in ascending or descending order.

If equal ranks or tie in ranks are given then for such cases, average ranks should be assign to each of that individual.

If the ranks of two items are same (say 4), then we can consider their rank as average of 4 and 5. i.e. $\frac{4+5}{2} = 4.5$ and next item is assigned rank 6 and so on.

Solved Examples

Ex. 4.8.1 : A material used in the construction industry contains an impurity suspected of having an adverse effect upon the performance of material. Percentage of impurity and performance indexes for 5 specimans of this material are as follows :

Specimens	1	2	3	4	5
% Impurity (x_i) rank	2	4	1	3	5
Performance (y_i) rank	3	2	1	5	4
d_i	–1	2	0	–2	1
d_i^2	1	4	0	4	1

Soln. :

Step 1 :

Here $\sum d_i = -1+2-2+1 = 0$;

$\sum d_i^2 = 1+4+0+4+1 = 10$;

$\therefore$ By Rank correlation co-efficient

$$r_s = 1 - \left[\frac{6\sum d_i^2}{n(n^2-1)}\right] = 1 - \left[\frac{6(10)}{5(5^2-1)}\right]$$

$$= 1 - \left[\frac{60}{5(25-1)}\right] = 1 - \left[\frac{60}{120}\right]$$

$$= 1 - \frac{1}{2} = \frac{1}{2} = 0.5 \checkmark \quad \text{...Ans.}$$

Ex. 4.8.2 : Two boys, Sachin and Saurabh were asked to rank 7 different types of perfumes. The ranks given by them were as follows :

Perfume type	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Sachin (x_i)	2	4	1	3	5	7	6
Saurabh (y_i)	1	3	2	5	4	7	6

Calculate rank correlation co-efficients

Soln. :

Step I : Spearman's rank correlation coefficients is

$$r_s = 1-\left[\frac{6\sum d_i^2}{n(n^2-1)}\right] \quad ...(1)$$

where, d_i = difference between the ranks from each individuals

n = number of items = 7

Step II : Construct table :

Perfume	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Sachin	2	4	1	3	5	7	6
Saurabh	1	3	2	5	4	7	6
$d_i = x_i - y_i$	1	1	−1	−2	1	0	0
d_i^2	1	1	1	4	1	0	0

Here, $\sum d_i = 0$

$$\sum d_i^2 = 1+1+1+4+1 = 8$$

Step III : Equation (1) becomes

$$r_s = 1-\left[\frac{6\times 8}{7(7^2-1)}\right] = 1-\left(\frac{48}{7\times 48}\right)$$

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

$$r_s = \frac{6}{7} = 0.8571 \checkmark \quad ...\text{Ans.}$$

Ex. 4.8.3 : The marks secured by 9 students in the online and theory exam are as follows :

Students	1	2	3	4	5	6	7	8	9
Online exam (x)	10	15	12	17	13	16	24	14	22
Theory exam (y)	30	42	45	46	33	34	40	35	39

Calculate the rank correlation coefficient.

Soln. :

Step I : Since, the co-efficient of Spearman's rank correlation is

$$r_s = 1-\left[\frac{6\sum d_i^2}{n(n^2-1)}\right] \quad ...(1)$$

$d_i = (x_i - y_i)$ = is the difference between ranks

n = number of students = 9

Step II : Construct table :

Students	1	2	3	4	5	6	7	8	9
X	10	15	12	17	13	16	24	14	22
Y	30	42	45	46	33	34	40	35	39
Ranks in X	9	5	8	3	7	4	1	6	2
Ranks in Y	9	3	2	1	8	7	4	6	5
d_i	0	2	6	2	−1	−3	−3	0	−3
d_i^2	0	4	36	4	1	9	9	0	9

(Here, ranks are given as highest number 1 and so on, we may give lowest number as 1)

Here, $\sum d_i = 0$

$$\sum d_i^2 = 4+36+4+1+9+9+9$$

$$= 72$$

Step III : Substitute all these values in Equation (1)

$$\therefore\ r_s = 1-\frac{6\sum d_i^2}{n(n^2-1)} = 1-\frac{6(72)}{9(9^2-1)}$$

$$= 1-\frac{6(72)}{9(80)} = 1-\left(\frac{432}{720}\right) = 1-0.6$$

$$r_s = 0.4 \checkmark \quad ...\text{Ans.}$$

Ex. 4.8.4 : Ten participants in a beauty contest are ranked by two judges as follows :

x	1	6	5	10	3	2	4	9	7	8
y	6	4	9	8	1	2	3	10	5	7

Calculate rank correlation coefficient.

Soln. :

Step I :

Since, Rank correlation coefficients

$$r_s = 1-\frac{6\sum d_i^2}{n(n^2-1)} \quad ...(1)$$

where, d_i = difference between ranks

n = 10

Step II : Construct table :

x	1	6	5	10	3	2	4	9	7	8
y	6	4	9	8	1	2	3	10	5	7
d_i	−5	2	−4	2	2	0	1	−1	2	1
d_i^2	25	4	16	4	4	0	1	1	4	1

Step III

Here, $\sum d_i = -5+2-4+2+2+1-1+2+1$

$$= 0$$

$$\sum d_i^2 = 25+4+16+4+4+1+1+4+1$$

$$= 60$$

$\therefore$ Coefficient of rank correlation is

$$r_s = 1-\frac{6\times 60}{10(100-1)}$$

$$= 1-\frac{360}{10\times 99} = 1-\frac{360}{990}$$

$$r_s = 0.6364 \checkmark \quad ...\text{Ans.}$$

Ex 4.8.5 : Given $r_s = 0.5$; $n = 10$. Later, it was noticed that the difference in ranks by one of the persons was wrongly taken as 3 instead of 7. What is the correct value of r_s.

Soln. :

Step I :

Given : $n = 10$; $r_s = 0.5$;

Since,

$$r_s = 1-\left[\frac{6\sum d_i^2}{n(n^2-1)}\right]$$

Therefore, $$0.5 = 1-\frac{6\sum d_i^2}{10(10^2-1)} = 1-\frac{6\sum d_i^2}{10(99)}$$

$$0.5 = 1-\frac{6\sum d_i^2}{990}$$

$$\frac{6\sum d_i^2}{990} = 1-0.5 = 0.5$$

$$\Rightarrow\ \Sigma d_i^2 = \frac{(0.5)\times(990)}{6} = 82.5$$

Step II : Given, one of rank difference was wrongly taken as 3 instead of 7

Modified, $$\sum d_i^2 = (82.5)-(3)^2+(7)^2$$

$$= (82.5)-(9)+49$$

$$= 122.5$$

Step III : Hence, corrected rank coefficient

$$r_s = 1-\left[\frac{6\sum d_i^2}{n(n^2-1)}\right] = 1-\left[\frac{(122.5)\times 6}{10(10^2-1)}\right]$$

$$= 1-\left[\frac{735}{10\times 99}\right] = 1-0.742 = 0.258$$

For Repeated Ranks :

- If any two or more individuals have same rank then the rank correlation coefficients (r_s) fails and we need small adjustment. In such case, each individual gives average rank.
- If two individual of same series are ranked equal at 4th place. They would have the 4th and 5th ranks. The common rank $= \frac{4+5}{2} = 4.5$
- If three individuals are ranked equal at 3rd place, they should have the 3rd, 4th, 5th ranks. Their common rank $= \frac{3+4+5}{3} = \frac{12}{3} = 4$
- So for such cases, add $\frac{1}{12}m(m^2-1)$ to $\sum d_i^2$;

Where, m = number of times an individual number repeated

Note : For each repeated items, add this adjustment factor

$$\therefore\ r = 1-6\left[\frac{\sum d_i^2+\frac{1}{12}m(m^2-1)+\frac{1}{12}m(m^2-1)+...}{n(n^2-1)}\right]$$

Ex. 4.8.6 : Obtain the rank correlation co-efficient for the following data :

X	70	64	82	55	64	90	82	50	60	64
Y	70	62	75	52	85	65	75	45	60	80

Soln. :

Step I : Since, Spearman's rank correlation coefficient (r_s) is

$$\therefore\ r_s = 1-6\left[\frac{\sum d_i^2+\frac{1}{12}m(m^2-1)+\frac{1}{12}m(m^2-1)+...}{n(n^2-1)}\right]$$

Step II : Construct table, rank can be given from highest number 1 and so on.

X	70	64	82	55	64	90	82	50	60	64
Y	70	62	75	52	85	65	75	45	60	80
Rank (x_i)	4	6	2.5	9	6	1	2.5	10	8	6
Rank (y_i)	5	7	3.5	10	1	6	3.5	9	8	2
d_i	−1	−1	−1	−1	5	−5	−1	1	0	4
d_i^2	1	1	1	1	25	25	1	1	0	16

Step III : $\Sigma d_i^2 = 72$; $\Sigma d_i = 0$; $n = 10$

Here, in X-series : the number 82 occurs twice. It's common rank is, $\frac{2+3}{2} = 2.5$

The value of 64 occurs thrice

Its rank 5, 6, 7. Common rank $= \frac{5+6+7}{3} = 6$

Similarly in Y-series, the value 75 occurs two times. They would have given ranks 3 and 4, the common rank given to them is

$$\frac{3+4}{2} = 3.5$$

Here, m = 2, 3, 2

$$\therefore r = 1 - \frac{6\left\{\sum d_i^2 + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1)\right\}}{n(n^2-1)}$$

$$= 1 - \frac{6\left\{72 + \frac{1}{12}2(2^2-1) + \frac{1}{12}3(3^2-1) + \frac{1}{12}2(2^2-1)\right\}}{10(10^2-1)}$$

$$= 1 - \frac{6 \times 75}{990} = 1 - \frac{430}{990}$$

$= 0.5656$ ✓ ...Ans.

Exercise

Ex. 1 : Students got the following percentage of marks in electronics and physics

Students	1	2	3	4	5	6	7	8	9	10
Marks in electronics	78	36	98	25	75	82	90	62	65	39
Marks in physics	84	51	91	60	68	62	86	58	63	47

Calculate the rank correlation coefficient

Ans. : $r_s = 0.84$

Ex. 2 : Calculate the rank correlation co-efficient for the following ages of husbands and wives :

Husbands Age (x)	23	27	28	28	29	30	31	33	35	36
Wife's age (y)	18	20	22	27	21	29	27	29	28	29

Ans. : 0.8

Ex. 3 : A sample of 12 father's and their eldest sons gave the following data about their heights in inches.

Father	65	63	67	64	68	62	70	66	68	67	69	71
Son	68	66	68	65	69	66	68	65	71	67	68	70

Calculate the coefficient of rank correlation

Ans. : 0.722

Ex. 4 : Two girls were asked to rank 7 different types of hair pins. The ranks given by them were :

Hair pins	H_1	H_2	H_3	H_4	H_5	H_6	H_7
Girl 1	2	4	1	3	5	7	6
Girl 2	1	3	2	5	4	7	6

Calculate rank correlation co-efficient.

Ans. : $r_s = 0.8571$

Ex. 5 : In practical exam of computer subject, three examiner judges the students and had given the rank to them as follows :

Students	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
Examiner - 1	6	1	4	8	7	5	3	10	9	2
Examiner - 2	3	5	6	1	8	2	4	10	9	7
Examiner - 3	1	2	3	4	5	6	7	8	1	10

Which pair of examiners have nearest approach towards subject **Ans. :** for $E_{12} = r_{s12} = 0.309$

For $E_{23} = r_{s23} = 0.527$

For $E_{13} = r_{s13} = 0.2$

Since r_{s23} is maximum

So the pair E_{23} i.e. examiner 2 and 3 is having the nearest approach towards subject.

Ex. 6 : Following table gives the values of export of mango(x) and the values of imports of mango product (y) in Lakhs of Rs.

x	42	44	58	54	90	99	62
y	78	60	65	70	71	80	70

Find spearman's rank correlation coefficient between x and y **Ans. :** 0.795

Ex. 7 : The coefficient of rank correlation of the marks obtained by 10 students in maths and chemistry was found to be 0.9. it was noticed later that the difference in ranks in the two subjects obtained for one of the student was wrongly taken as 5

Syllabus Topic : Probable Errors

4.9 Probable Error of Karl Person's Coefficient of Correlation :

- Probable Error (PE) helps in determining the and reliability of observed value of Karl Person's Coefficient of Correlation (r).
- For N pair of observations, if r is the Karl Person's Coefficient of Correlation, then the probable error (PE) is :

$$\text{Probable Error(PE)} = 0.6745\left(\frac{1-r^2}{\sqrt{N}}\right)$$

Where, r – is the Karl Person's Coefficient of Correlation.

N – Number of observations

1. If r < PE then the value of r is not significance. i.e. here is no correlation between the variables
2. The value of the probable error is bounds in – 1 and + 1 ($-1 \le r \le +1$).
3. If r > 6 PE , then the value of r is significance. i.e. there is certain correlation between the given variables.
4. By adding and subtracting the value of P.E from the value of 'r,' we get the upper limit and the lower limit, respectively within which the correlation of coefficient is expected to lie.
5. The probable Error can be used only for normal distribution curve.
6. Probable error can be calculated I the sample items are taken off in an unbiased manner and must remain independent of each other's value
7. Also, probable error can be calculated if the data is taken from the sample only.

Solved Examples

Ex. 4.9.1 : Find the probable error with correlation coefficient 0.8 for 25 pairs of samples.

Soln. :

Step I : Given, r = 0.8 and N = 25.

We know that, Probable Error(PE) $= 0.6745\left(\frac{1-r^2}{\sqrt{N}}\right)$

Step II : So, on putting the values in formula:

Probable Error $= 0.674 \times \{(1-(0.8)2)/\sqrt{25}\}$

$= 0.674 \times \{(1-0.64)/5\} = 0.674 \times (0.36/5)$

≈ 0.0486

Therefore, the probable error is: 0.0486. ✓ ...Ans.

Ex. 4.9.2 : If co-efficient of correlation r = 0.7 and n = 64, then find the Probable Error of the correlation of coefficient. Furthermore, find the limits for the population correlation coefficient.

Soln. :

Step I : Given, r = 0.7 and N = 64.

Here, we have to calculate the probable error We know that,

Step II :

$$\text{Probable Error(PE)} = 0.6745\left(\frac{1-r^2}{\sqrt{N}}\right)$$

$= 0.6745 \times \{(1-(0.7)2)/\sqrt{64}\}$

$= 0.6745 \times 0.06375$

≈ 0.043

Therefore, the Probable Error(PE) is 0.043.

Step III : Now, we have to calculate the limits of the population correlation coefficient.

We can use formula , Probable Limit-

ρ (rho) = r ± Probable Error

Hence, we get, (0.7 ± 0.043) i.e. (0.743, 0.657). ✓ ...Ans.

Exercise

Ex. 1 : The following marks have been obtained by a class of students in two papers of economics :

Paper I	45	55	56	58	60	65	68	70	75	80	85
Paper II	56	50	48	60	62	64	65	70	74	82	90

Calculate the coefficient of correlation for the above data.

Ex. 2 : Obtain the correlation coefficient between population density (per square miles) and death rate (per thousand persons) from the following data :

Population density	200	500	400	700	300
Death rate	12	18	16	21	10

Ex. 3 : Calculate the coefficient of correlation from the following data :

$n = 20$, $\sum x_i = 40$, $\sum x_i^2 = 190$, $\sum y_i^2 = 200$,

$\sum x_i y_i = 150$, $\sum y_i = 40$

Ex. 4 : Find the number of items if x and y are deviations from arithmetic mean given :

$r(x, y) = 0.9$, $\sum xy = 70$; $\sigma_y = 3.5$, $\sum x_i^2 = 100$

Ex. 5 : Calculate the co-efficient of correlation for the following data :

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Ex. 6 : Calculate the coefficient of correlation between the marks obtained by 10 students in Statistics and Economics, from the following table.

Students	1	2	3	4	5	6	7	8	9	10
Marks in Economics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Obtain regression equations and estimate marks in statistics if marks in Economics are 30.

Ex. 7 : Find the coefficient of correlation for the following data ;

x	10	14	19	26	30	34	39
y	12	16	18	26	29	35	38

Multiple Choice Questins

Q. 1 The Karl-Pearson's correlation coefficient r (x, y) between two variables x and y is _____.

(a) $\frac{\sum(xy)}{\sum x \sum y}$ (b) $\frac{\sigma_x}{\sum x} \cdot \frac{\sigma_y}{\sum y}$

(c) $\frac{1}{n}\sum uv - \bar{u}\,\bar{v}$ (d) $\frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$

Ans. : (d)

Explanation : Karl Pearson's co-efficient of correlation is defined as, $r(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \cdot \sigma_y}$

Q. 2 The value of covariance of u, v is _____

(a) $\frac{1}{n}\left(\sum u_i v_i - \bar{u}\,\bar{v}\right)$

(b) $\frac{1}{n}\sum u_i v_i - \bar{u}\,\bar{v}$

(c) $\left(\sum u_i - \bar{u}\right)\left(\sum v_i - \bar{v}\right)$

(d) $\frac{1}{n}\left(\sum u_i - \bar{u}\right)\frac{1}{n}\left(\sum v_i - \bar{v}\right)$

Ans. : (b)

Explanation : $\text{Cov}(x, y) = \frac{1}{n}\left(\sum u_i v_i\right) - \bar{u}\,\bar{v}$

Q. 3 The range of value of correlation coefficient r is _____.

(a) 1 (b) $-1 \le r \le 1$

(c) $0 \le r \le 1$ (d) -1

Ans. : (b)

Explanation :

The range of value of correlation coefficient r is $-1 \le r \le 1$

Q. 4 Probable error of co-efficient of correlation r is,

(a) $0.6745\left(\frac{1-r^2}{\sqrt{N}}\right)$ (b) $0.6745(1 + r^2)$

(b) $0.6745\left(\frac{1-r^2}{N}\right)$ (d) $0.6745\left(\frac{1-r}{N}\right)^2$

Ans. : (a)

Explanation : For N pair of observations, if r is the Karl Person's Coefficient of Correlation, then the probable error (PE) is :

Probable Error(PE) $= 0.6745\left(\frac{1-r^2}{\sqrt{N}}\right)$

Q. 5 For the data : n = 10; $\bar{u} = -5.1$, $\bar{v} = -10$, $\sum u_i v_i = 1242$, $\sum u_i^2 = 1169$, $\sum v_i^2 = 1694$ the value of coefficient of correlation r is _____.

(a) 0.74 (b) 0.92 (c) 0.65 (d) -0.89

Ans. : (b)

Explanation : $n = 10, \bar{u} = 5.1, \bar{v} = -10,$

$\sum u_i v_i = 1242$

$\sum u_i^2 = 1169, \quad \sum v_i^2 = 1694$

$r = \frac{\text{cor}(u,v)}{\sigma_u \sigma_v}$ and

$\text{cov}(u,v) = \frac{1}{n}\sum u_i v_i - \bar{u}\,\bar{v} = 73.2$

$\sigma_u^2 = \frac{1}{n}\sum u_i^2 - \bar{u}^2 = 90.89 \Rightarrow \sigma_u = 9.53$

$\sigma_v^2 = \frac{1}{n}\sum v_i^2 - \bar{u}^2 = 69.4 \Rightarrow \sigma_v = 8.33$

$\therefore \quad r = \frac{73.2}{(9.53)(8.33)} = 0.92$

Q. 6 The value of coefficient of correlation r for the following data :

$n = 100, \bar{x} = 62, \bar{y} = 53, \sigma_x = 10, \sigma_y = 12, \Sigma\left(x - \bar{x}\right)\left(y - \bar{y}\right) = 8000$ is _____.

(a) 0.92 (b) 0.87 (c) 0.66 (d) -0.78

Ans. : (c)

Explanation :

$n = 100, \bar{x} = 62, \bar{y} = 53, \sigma_x = 10, \sigma_y = 12$

$\sum(x - \bar{x})(y - \bar{y}) = 8000;$

We have, $r = \frac{\text{Cor}(x,y)}{\sigma_x \sigma_y}$

$\text{Cor}(x,y) = \frac{1}{n}\sum(x - \bar{x})(y - \bar{y}) = \frac{8000}{100} = 80$

$\therefore \quad r = 0.666$

Q. 7 Given n = 25, $\Sigma x = 75$, $\Sigma y = 100$, $\sum x^2 = 250$, $\sum y^2 = 500$, $\sum xy = 325$ then value of coefficient of correlation r is _____.

(a) -0.7 (b) 0.8 (c) 0.5 (d) 0.9

Ans. : (c)

Explanation : Given n = 25, $\Sigma x = 75$, $\sum y = 100$, $\sum x^2 = 250$ $\Sigma y^2 = 500$, $\sum xy = 325$

$r = \frac{\text{cor}(x,y)}{\sigma_x \sigma_y}; \quad \bar{x} = \frac{\sum x}{n} = 3; \quad \bar{y} = \frac{\sum y}{n} = 4$

$\text{Cor}(x,y) = \frac{1}{n}\sum xy - \bar{x}\,\bar{y} = 13 - 12 = 1$

$\sigma_x^2 = \frac{1}{n}\sum x^2 - \bar{x}^2 = 10 - 9 = 1 \Rightarrow \sigma_x = 1$

$\sigma_y^2 = \frac{1}{n}\sum y^2 - \bar{y}^2 = 20 - 16 = 4 \Rightarrow \sigma_y = 2$

$\therefore \quad r = 0.5$

Q. 8 The value of probable error with correlation coefficient 0.8 for 25 pairs of samples is

(a) 0.0486 (b) 0.8 (c) 0.05 (d) 0.9

Ans. : (a)

Explanation : Given, r = 0.8 and N = 25.

We know that, Probable Error(PE) $= 0.6745\left(\frac{1-r^2}{\sqrt{N}}\right)$

So, on putting the values in formula:

Probable Error $= 0.674 \times \{(1 - (0.8)2)/\sqrt{25}\}$

$= 0.674 \times \{(1 - 0.64)/5\} = 0.674 \times (0.36/5)$

≈ 0.0486

Therefore, the probable error is: 0.0486

Q. 9 If n = 10 and $\sum d_i^2 = 60$, then the value of Coefficient of rank correlation is

(a) 4.86 (b) 0.8 (c) 0.5 (d) 0.6364

Ans. : (d)

Explanation :

Coefficient of rank correlation is

$r_s = 1 - \frac{6 \times 60}{10(100 - 1)}$

$= 1 - \frac{360}{10 \times 99} = 1 - \frac{360}{990}$

$r_s = 0.6364$

Q. 10 If n = 7 and $\sum d_i^2 = 8$, then the value of Coefficient of rank correlation is

(a) 4.86 (b) 0.8571 (c) 0.5 (d) 0.6364

Ans. : (b)

Explanation :

$\sum d_i^2 = 8$

Equation (1) becomes

$r_s = 1 - \left[\frac{6 \times 8}{7(7^2 - 1)}\right] = 1 - \left(\frac{48}{7 \times 48}\right)$

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

$r_s = \frac{6}{7} = 0.8571$

Q. 11 If Spearman's Rank Correlation Co-efficient $r_s = 0$, then

(a) there is no correlation

(b) there is positive correlation

(c) there is perfect correlation

(d) there is negative correlaton

Ans (a)

Explanation : If $r_s = 0$, then there is no correlation.

Q. 12 The Spearman's rank correlation Co-efficient is defined as

(a) $1 - \left[\frac{6\sum d_i^2}{n(n^2 - 1)}\right]$ (b) $\frac{\text{cor}(x,y)}{\sigma_x \sigma_y}$

(c) $\left[\frac{6\sum d_i^2}{n(n^2 - 1)}\right]$ (d) $\frac{1}{n}\left(\sum x_i y_i\right) - \bar{x}\,\bar{y}$

Ans : (a)

Explanation : The Spearman's rank correlation Co-efficient is denoted by r_{rank} or r_s and is defined as

$r_s = r_{rank} = 1 - \left[\frac{6\sum d_i^2}{n(n^2 - 1)}\right]$

Chapter Ends...

□□□

CHAPTER 5

UNIT V

Linear Regression Analysis

Syllabus

Introduction, Linear and Non-linear Regression, Lines of Regression, Derivations of Regression Lines of Y on X and X on Y , Angle between the Regression Lines, Coefficient of Regression, Theorems on Regression Coefficient, Properties of Regression Coefficient .

Syllabus Topic : Introduction

5.1 Introduction

In this chapter we will study the statistical relationships between two of more random variables. Here, study the regression curves of dependent variable Y on independent variable X .Conversely, we can also find regression curves of dependent variable X on independent variable Y.

Regression analysis is a statistical device to estimate one of the variables (dependant variable) from the other variable (independent variable).When two variables are correlated, then we can use the correlation to estimate the one variable, if the value of other variable is given. The method of prediction ,on the basis of correlation is known as Regression Analysis.

Syllabus Topic : Linear Regression and Non- Linear Regression

5.2 Linear Regression

Regression may be simple (involving two variables) or multiple (more than two variables)and the relationship could be linear (straight line) or non-linear .

Linear Regression

- Linear regression is a linear approach to find the relationship between a dependent variable (response) and one or more independent variable (input).
- If it is one input variable then it is called simple linear regression. Linear regression focuses on conditional probability distribution of the output variable.
- Linear regression gives two straight lines, as there are two variables.

Non-Linear Regression

- Regression is called nonlinear regression because the relationships between the dependent and independent variables are not linear.
- Non-linear regression is a regression in which the response (dependent) or criterion variables are appear as a non-linear function of one or more independent variables.
- There are several non-linear regressions models, such as Asymptotic Regression/Growth Model, Logistic Population Growth Model, Decay Model etc.
- Non-linear regression estimates the parameters by minimizing the sum of squared errors.
- It is more flexible to fit curves than linear regression.

Syllabus Topic : Lines Of Regression

5.3 Lines of Regression

- By using least square principle, what we studied in previous topic, we can find the lines of regressions of Y on X and vice versa from the given data.
- The best line should be that line, which is very close to all the point in scatter diagram.
- The equation of line so obtained is called as least square regression line.

- Suppose $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots\ldots, (x_n, y_n)$ are some n finite pair of observations of random variables X and Y.

Syllabus Topic : Derivation of Lines of Regression of Y on X

5.3.1 Derivation of Lines of Regression of Y on X

- Let us suppose that the variable Y is dependent and the variable X is independent so, we can find Y for the known values of X.
- Let us consider the equation of straight line $Y = mX + c$, where m and c are parameters.
- By putting the observed values of X, estimated values of Y and by least square principle, we get the lines of regression Y on X is

$$(y-\bar{y}) = \frac{r(x,y)\,\sigma_y}{\sigma_x}(x-\bar{x})$$

or $$(y-\bar{y}) = \frac{Cov(x,y)}{\sigma_x^2}(x-\bar{x})$$

or $$(y-\bar{y}) = b_{yx}(x-\bar{x})$$

where, $$b_{yx} = \frac{Cov(x,y)}{\sigma_x^2}$$

is called regression co-efficient of y on x.

Proof : Let us consider the equation of straight line

$$y = mx + c \qquad \ldots(5.3.1)$$

where, m -is the slope and c -is y intercept.

Here, we have to find the values of m and c, by using least square principle.

Let y_i – be the observed value of y and y_i' – be the estimated value of y corresponding x_i from equation

$$\therefore\ y_i' = m x_i + c \qquad \ldots(5.3.2)$$

Therefore, $y_i - y_i'$, be the error in the estimation. By least square principle, we have

$$(y_1-y_1')^2 + (y_2-y_2')^2 + (y_3-y_3')^2 + \ldots + (y_n-y_n')^2$$

be the sum of squares of errors.

Symbolically, we can write

$$F = \sum_{i=1}^{n}(y_i-y_i')^2$$

as a sum of squares of errors.

To minimize the sum of the squares of differences between y_i and y_i'. By the concept of maxima and minima, we can find,

$\frac{\partial F}{\partial c} = 0$ and $\frac{\partial F}{\partial m} = 0$; gives stationary values.

Since, $$F = \sum_{i=1}^{n}(y_i-y_i')^2$$

$$F = \sum_{i=1}^{n}(y_i-mx_i-c)^2$$

by using Equation (5.3.2) ...(5.3.3)

By the concept of DUIS rule, differentiating Equation (5.3.3), w.r.t. m , keeping c as constant

$$\frac{dF}{dm} = \frac{d}{dm}\sum_{i=1}^{n}(y_i-mx_i-c)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n}\frac{\partial}{\partial m}(y_i-mx_i-c)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n}2(y_i-mx_i-c)(-x_i) = 0$$

$$\Rightarrow -2\left[\sum_{i=1}^{n}x_iy_i - m\sum_{i=1}^{n}x_i^2 - c\sum_{i=1}^{n}x_i\right] = 0$$

$$\Rightarrow \sum x_iy_i - m\sum_{i=1}^{n}x_i^2 - c\sum_{i=1}^{n}x_i = 0$$

$$\boxed{\sum x_iy_i = m\sum_{i=1}^{n}x_i^2 + c\sum x_i} \qquad \ldots(5.3.4)$$

Again by DUIS concept, differentiating Equation (5.3.3), w.r.t c, keeping m as constant.

$$\frac{dF}{dc} = \frac{d}{dc}\sum_{i=1}^{n}(y_i-mx_i-c)^2 = 0$$

$$\sum_{i=1}^{n}\frac{\partial}{\partial c}(y_i-mx_i-c)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n}2(y_i-mx_i-c)(-1) = 0$$

$$\Rightarrow -2\sum_{i=1}^{n}(y_i-mx_i-c) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(y_i-mx_i-c) = 0$$

$$\sum_{i=1}^{n}y_i - m\sum_{i=1}^{n}x_i - nc = 0 \qquad \ldots\left(\text{Since, } \sum_{i=1}^{n}c = nc\right)$$

$$\boxed{\sum_{i=1}^{n}y_i = m\sum_{i=1}^{n}x_i + nc}$$

$$\Rightarrow c = \frac{\sum y_i}{n} - \frac{m\sum x_i}{n}$$

$$\boxed{c = \bar{y} - m\bar{x}} \quad ...(5.3.5)$$

By using Equation (5.3.5), Equation (5.3.4) becomes

$$\sum x_i y_i = m\sum x_i^2 + (\bar{y} - m\bar{x})\sum x_i$$

$$\sum x_i y_i = m\sum x_i^2 + \bar{y}\sum x_i - m\bar{x}\sum x_i$$

$$= m\sum x_i^2 + \bar{y}(n\bar{x}) - m\bar{x}(n\bar{x})$$

$$...\left(\because \frac{\sum x_i}{n} = \bar{x}\right)$$

$$\sum x_i y_i = m\sum x_i^2 + n\bar{x}\bar{y} - mn(\bar{x})^2$$

$$\sum x_i y_i - n\bar{x}\bar{y} = m\left[\sum x_i^2 - n\left(\bar{x}\right)^2\right]$$

OR $$\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y} = m\left[\frac{\sum x_i^2}{n} - (\bar{x})^2\right]$$

$$\text{Cov}(x, y) = m\sigma_x^2 \quad ...\text{by covariance and standard deviation formula.}$$

OR $$\boxed{m = \frac{\text{Cov}(x, y)}{\sigma_x^2}} \quad ...(5.3.6)$$

is a slope (regression co-efficient of y on x) of line Y on X.

Putting the values of m from Equation (5.3.6) and c from Equation (5.3.5), Equation (5.3.1) becomes

$$y = mx + \bar{y} - m\bar{x} \quad (\text{Since}, \bar{y} - m\bar{x})$$

$$(y - \bar{y}) = m(x - \bar{x})$$

$$\boxed{(y - \bar{y}) = \frac{\text{Cov}(x, y)}{\sigma_x^2}(x - \bar{x})}$$

OR $$\boxed{(y - \bar{y}) = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})} \quad \text{OR} \quad \boxed{(y - \bar{y}) = b_{yx}(x - \bar{x})}$$

Which is the lines of regression of Y on X.

Where, $b_{yx} = r\dfrac{\sigma_y}{\sigma_x} = \dfrac{\text{Cov}(x, y)}{\sigma_x^2}$ is called regression co-efficient of y on x

Syllabus Topic : Derivation of Lines of Regression of X on Y

5.3.2 Derivation of Lines of Regression of X on Y

Here, let us suppose that X is dependent variable and Y is independent variable, so we can find the values of X for known values of Y. we get the lines of regression of X on Y is

$$\boxed{(x - \bar{x}) = \frac{r(x, y)\sigma_x}{\sigma_y}(y - \bar{y})}$$

OR $$\boxed{(x - \bar{x}) = \frac{\text{Cov}(x, y)}{\sigma_y^2}(y - \bar{y})}$$

OR $$\boxed{(x - \bar{x}) = b_{xy}(y - \bar{y})}$$

where, $\boxed{b_{xy} = \dfrac{\text{Cov}(x, y)}{\sigma_y^2}}$ is called regression co-efficient of x on y.

Proof : Let us consider the equation of straight line

$$x = ay + b \quad ...(5.3.7)$$

where, a- is the slope and b-is y intercept.

Here, we have to find the values of a and b, by using least square principle.

Let x_i – be the observed value of x and x_i' – be the estimated value of x corresponding y_i from equation

$$\therefore x_i' = ay_i + b \quad ...(5.3.8)$$

Therefore, $x_i - x_i'$, be the error in the estimation. By least square principle, we have

$(x_1 - x_1')^2 + (x_2 - x_1')^2 + (x_3 - x_1')^2 + ...+ (x_n - x_n')^2$ be the sum of squares of errors.

Symbolically, we can write

$$F = \sum_{i=1}^{n}(x_i - x_i')^2 \quad \text{as a sum of squares of errors.}$$

To minimize the sum of the squares of differences between x_i and x_i',

By the concept of maxima and minima ,we can find,

$$\frac{\partial F}{\partial a} = 0 \quad \text{and} \quad \frac{\partial F}{\partial b} = 0;$$

gives stationary values.

Since, $$F = \sum_{i=1}^{n}(x_i - x_i')^2$$

$$F = \sum_{i=1}^{n}(x_i - ay_i - b)^2$$

by using Equation (5.3.2.2) ...(5.3.9)

By the concept of DUIS rule, differentiating Equation (5.3.9), w.r.t. a , keeping b as constant

$$\frac{dF}{da} = \frac{d}{da}\sum_{i=1}^{n}(x_i - ay_i - b)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n}\frac{\partial}{\partial a}(x_i - ay_i - b)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n}2(x_i - ay_i - b)(-y_i) = 0$$

$$\Rightarrow -2\left[\sum_{i=1}^{n}x_i y_i - a\sum_{i=1}^{n}y_i^2 - b\sum_{i=1}^{n}y_i\right] = 0$$

$$\Rightarrow \sum x_i y_i - a\sum_{i=1}^{n}y_i^2 - c\sum_{i=1}^{n}y_i = 0$$

$$\boxed{\sum x_i y_i = a\sum_{i=1}^{n}y_i^2 + b\sum y_i} \quad ...(5.3.10)$$

Again by DUIS concept, differentiating Equation (5.3.9), w.r.t. b, keeping a as constant.

$$\frac{dF}{db} = \frac{d}{db}\sum_{i=1}^{n}(x_i - my_i - b)^2 = 0$$

$$\sum_{i=1}^{n}\frac{\partial}{\partial b}(x_i - my_i - b)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n}2(x_i - ay_i - b)(-1) = 0$$

$$\Rightarrow -2\sum_{i=1}^{n}(x_i - ay_i - b) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(x_i - ay_i - b) = 0$$

$$\sum_{i=1}^{n}x_i - a\sum_{i=1}^{n}y_i - nb = 0$$

$$...(\text{Since}, \sum_{i=1}^{n}b = nb)$$

$$\boxed{\sum_{i=1}^{n}x_i = a\sum_{i=1}^{n}y_i + nb}$$

$$\Rightarrow b = \frac{\sum x_i}{n} - \frac{a\sum y_i}{n}$$

$$\boxed{b = \bar{x} - a\bar{y}} \quad ...(5.3.11)$$

By using Equation (5.3.11), Equation (5.3.10) becomes

$$\sum x_i y_i = a\sum y_i^2 + (\bar{x} - a\bar{y})\sum y_i$$

$$\sum x_i y_i = a\sum y_i^2 + \bar{x}\sum y_i - a\bar{y}\sum y_i$$

$$= a\sum y_i^2 + \bar{x}(n\bar{y}) - a\bar{y}(n\bar{y})$$

$$...\left(\because \frac{\sum y_i}{n} = \bar{y}\right)$$

$$\sum x_i y_i = a\sum y_i^2 + n\bar{x}\bar{y} - an(\bar{y})^2$$

$$\sum x_i y_i - n\bar{x}\bar{y} = a\left[\sum y_i^2 - n(\bar{y})^2\right]$$

OR $$\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y} = a\left[\frac{\sum y_i^2}{n} - (\bar{y})^2\right]$$

$$\text{Cov}(x, y) = a\sigma_y^2 \quad ...\text{by covariance and standard deviation definition.}$$

OR $$\boxed{a = \frac{\text{Cov}(x, y)}{\sigma_y^2}} \quad ..(5.3.12)$$

is a slope (regression co-efficient of X on Y) of line X on Y.

Putting the values of 'a' from Equation (5.3.12) and 'b' from Equation (5.3.11), Equation (5.3.7) becomes

$$x = ay + \bar{x} - a\bar{y} \quad (\text{Since}, b = \bar{x} - a\bar{y})$$

$$(x - \bar{x}) = a(y - \bar{y})$$

$$\Rightarrow \boxed{(x - \bar{x}) = \frac{\text{Cov}(x, y)}{\sigma_y^2}(y - \bar{y})}$$

OR $$\boxed{(x - \bar{x}) = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})}$$

$$\boxed{(x - \bar{x}) = b_{xy}(y - \bar{y})}$$

where, $b_{xy} = \dfrac{\text{Cov}(x, y)}{\sigma_y^2} = r\dfrac{\sigma_x}{\sigma_y}$ is called regression co-efficient of x on y.

Which is the lines of regression of X on Y.

In share market, the term $\beta = b_{xy}$ (beta index) is used. If $\beta > 1$, then the share is considered as aggressive / good, otherwise defensive.

If we write $u_i = x_i - a$, $v_i = y_i + b$

Then $\bar{x} = \bar{u} + a$

$\bar{y} = \bar{v} + b$

Syllabus Topic : Angle between the Regression Lines

5.4 Angle Between The Regression Lines

If θ is the acute angle between the two regression lines for two variables x and y, show that,

$$\tan\theta = \left|\left(\frac{1-r^2}{r}\right)\cdot\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right|.$$

Proof : Let us consider the lines of regression of y on x and x on y.

Step I : Let θ be the angle between two regression lines, as shown in Fig. P. 5.10.1

$$(y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \quad ...(1)$$

be the equation of lines of regression of y on x

$$\text{and} \quad (x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \quad ...(2)$$

be the equation of lines of regression of x on y.

Step II : From Equation (2),

$$\frac{\sigma_y}{r\sigma_x}(x-\bar{x}) = (y-\bar{y}) \quad ...(3)$$

Step III :

Since, if $y = m_1x + c_1$ and $y = m_2x + c_2$ are straight lines then m_1 and m_2 are their slopes and

$$\tan\theta = \left|\frac{m_1-m_2}{1+m_1m_2}\right|$$

From Equations (1) and (3).

$$m_1 = r\frac{\sigma_y}{\sigma_x} \quad \text{and} \quad m_2 = \frac{\sigma_y}{r\cdot\sigma_x}$$

Step IV :

$$\therefore \tan\theta = \left|\frac{m_1-m_2}{1+m_1m_2}\right| = \left|\frac{m_2-m_1}{1+m_1m_2}\right|$$

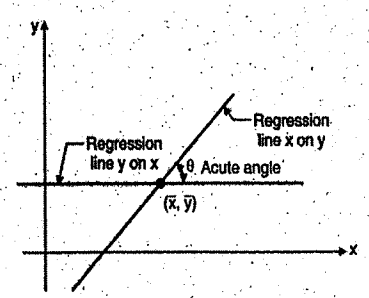


Fig. P. 5.4.1

$$= \left|\frac{\left(\frac{\sigma_y}{r\sigma_x}\right)-\left(r\frac{\sigma_y}{\sigma_x}\right)}{1+\frac{r\sigma_y}{\sigma_x}\frac{\sigma_y}{r\sigma_x}}\right| = \left|\frac{\left(\frac{1-r^2}{r}\right)\cdot\frac{\sigma_y}{\sigma_x}}{1+\frac{\sigma_y^2}{\sigma_x^2}}\right|$$

$$= \left|\frac{\left(\frac{1-r^2}{r}\right)\cdot\sigma_y}{\sigma_x\left(\frac{\sigma_x^2+\sigma_y^2}{\sigma_x^2}\right)}\right|$$

$$\boxed{\therefore \tan\theta = \left|\left(\frac{1-r^2}{r}\right)\cdot\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)\right|}$$

Syllabus Topic : Co-efficient of Regression

5.5 Co-efficient of Regression

I. For the lines of regression of Y on X

$$(y-\bar{y}) = \frac{\text{Cov}(x,y)}{\sigma_x^2}(x-\bar{x})$$

$$\text{OR} \quad (y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \quad \text{OR} \quad (y-\bar{y}) = b_{yx}(x-\bar{x})$$

$$\text{Where,} \quad b_{yx} = \frac{\text{Cov}(x,y)}{\sigma_y^2} = r\frac{\sigma_y}{\sigma_x}$$

is called regression co-efficient of Y on X.

In general, for the line $y = mx + c$, the slope 'm' is called regression co-efficient of Y on X.

II. For the lines of regression of X on Y

$$(x-\bar{x}) = \frac{\text{Cov}(x,y)}{\sigma_y^2}(y-\bar{y}) \quad \text{OR}$$

$$(x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \quad \text{OR} \quad (x-\bar{x}) = b_{xy}(y-\bar{y})$$

$$\text{where,} \quad b_{xy} = \frac{\text{Cov}(x,y)}{\sigma_y^2} = r\frac{\sigma_x}{\sigma_y}$$

is called regression co-efficient of X on Y.

In general, for the line $x = my + c$, the slope 'm' is called regression co-efficient of X on Y.

Syllabus Topic : Theorems on Co-efficient of Regression

5.6 Theorems on Co-efficient of Regression

Theorem 1

- Arithmetic mean of regression co-efficients is greater than the correlation co-efficient.

i.e. $\left|\frac{b_{xy}+b_{yx}}{2}\right| > r$. Also $(\sigma_x - \sigma_y)^2 > 0$

Theorem 2

- Co-efficient correlation is a geometric mean of regression co-efficient.

i.e.. $r(x,y) = \sqrt{b_{xy}\cdot b_{yx}}$

Theorem 3

- If one of the regression co-efficient is greater than unity, other must be less than unity.

i.e. If $b_{xy} > 1$ then $b_{yx} < 1$

Theorem 4

- Regression coefficients are invariants to the change of origin.

Theorem 5

- If r (x, y) is a geometric mean of b_{yx} and b_{xy}, then r (x, y) lies between b_{xy} and b_{yx}.

Theorem 6

- If the variables are independent, then the two lines of regression are perpendicular to each other

Theorem 7

- If the given two lines are coincides or parallel, then there is a perfect correlation between the two variables.

Syllabus Topic : Properties of Regression Co-efficient

5.7 Properties of Regression Co-efficient

1. Co-efficient of correlation and Co-efficient of regressions have same sign because all these terms have same numerator.
2. If both the regression co-efficient are negative then r is negative.
3. Arithmetic mean of regression co-efficients is greater than the correlation co-efficient.

   i.e. $\left|\frac{b_{xy}+b_{yx}}{2}\right| > r$. Also $(\sigma_x - \sigma_y)^2 > 0$
4. Co-efficient correlation is a geometric mean of regression co-efficient.

   i.e.. $r(x,y) = \sqrt{b_{xy}\cdot b_{yx}}$

   $$b_{xy}\cdot b_{yx} = \frac{\text{Cov}(x,y)}{\sigma_y^2}\cdot\frac{\text{Cov}(x,y)}{\sigma_x^2} = \left[\frac{\text{Cov}(x,y)}{\sigma_x\cdot\sigma_y}\right]^2$$

   $$\boxed{b_{xy}\cdot b_{yx} = r^2}$$

   is called the co-efficient of determination

   * chose positive square root if regression co-efficient are positive otherwise negative
5. $(b_{xy}\cdot b_{yx} = r^2) \le 1$
6. If one of the regression co-efficient is greater than unity, other must be less than unity.

   i.e. If $b_{xy} > 1$ then $b_{yx} < 1$
7. Regression coefficients are invariants to the change of origin.

   Let $u_i = \frac{x_i - a}{h}$; $v_i = \frac{y_i - b}{k}$ then

   $$b_{uv} = \frac{k}{h}b_{xy} \quad \text{and} \quad b_{vu} = \frac{h}{k}b_{yx}$$
8. $b_{yx} = r\cdot\frac{\sigma_y}{\sigma_x}$ and $b_{xy} = r\frac{\sigma_x}{\sigma_y}$
9. As r (x, y) is a geometric mean of b_{yx} and b_{xy}, this means r (x, y) lies between b_{xy} and b_{yx}.
10. If $r = \pm 1$ then $b_{xy} = \frac{1}{b_{yx}}$ or $b_{yx} = \frac{1}{b_{xy}}$
11. If $\sigma_x = \sigma_y$ then $b_{yx} = b_{xy} = r(x,y)$
12. Since, $\tan\theta = \left|\left(\frac{1-r^2}{r}\right)\cdot\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)\right|$
13. If $r = 0$; $\tan\theta \to \infty \Rightarrow \theta \to \pi/2$

    i.e. If the variables are independent, then the two lines of regression are perpendicular to each other i.e. the variables are more spread means not correlated.

    Also if $r = 0$, the regression lines are perpendicular to each other. These two lines of regression becomes $x = \bar{x}$ and $y = \bar{y}$ which are two straight lines parallel to x and y axis respectively.
14. If $r = \pm 1$, then $\theta \to 0 \Rightarrow \theta = 0$ or π

    i.e. The given two lines are coincides or parallel. There is a perfect correlation between the two variables.
15. The lines of regression are intersects at $(\bar{x}, \bar{y})$.
16. If r^2 is larger, then θ is smaller.

5.7.1 Working Rule to find Lines of Regression

Step I : If data is given then first find r(x, y), σ_x and σ_y and then use the equations of regression of y on x or x on y.

Step II : If directly straight lines are given then use following points.

Step III : Write the given equations of regression lines.

Step IV : Put $x = \bar{x}$ and $y = \bar{y}$ in the given equation and solve it, we get mean $\bar{x}$ and $\bar{y}$.

Step V : Consider the lines of regression of y on x then write as $y = m_1 x + c_1$ and $m_1 = b_{yx}$ OR regression lines of x on y then write as $x = m_2 y + c_2$ and $m_2 = b_{xy}$.

Step VI : Find $r = \sqrt{b_{xy} \cdot b_{yx}}$ such that $-1 \le r(x, y) \le 1$

Step VII : Sometimes use $b_{xy} = \dfrac{r\,\sigma_x}{\sigma_y}$

then find σ_x, if r, b_{xy} and σ_y given OR

find σ_y, if r, b_{yx} and σ_x given.

Note : 1. The lines of regressions

(i) x on y is : $(x - \bar{x}) = r\dfrac{\sigma_x}{\sigma_y}(y - \bar{y}) = b_{xy}(y - \bar{y})$

(ii) y on x is : $(y - \bar{y}) = r\dfrac{\sigma_y}{\sigma_x}(x - \bar{x}) = b_{yx}(x - \bar{x})$

2. $b_{xy} = \dfrac{r\,\sigma_x}{\sigma_y}$ and $b_{yx} = \dfrac{r\,\sigma_y}{\sigma_x}$

3. σ_x^2 = variance of x and σ_y^2 = variance of y

4. $r = \sqrt{b_{xy} \cdot b_{yx}}$

Ex. 5.7.1 : Find the correlation co-efficient from the following data :

$n = 10, \sum x_i = 300, \sum y_i = 250;$

$\sum (x_i - 30)^2 = 154; \sum (y_i - 25)^2 = 162$

$\sum (x_i - 30)(y_i - 25) = 144.$

Also find (i) Regression line Y on X and X on Y (ii) Angle between the lines

Soln. :

Step I :

Given : $n = 10, \sum x_i = 300; \sum y_i = 250;$

$\therefore \bar{x} = \dfrac{300}{10} = 30$ and $\bar{y} = \dfrac{250}{10} = 25$

Step II : The coefficient of correlation

$$r(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n}\sum (y_i - \bar{y})^2}}$$

$$r(x, y) = \frac{\frac{1}{10}\sum (x_i - 30)(y_i - 25)}{\sqrt{\frac{1}{10}\sum (x_i - 30)^2}\sqrt{\frac{1}{10}\sum (y_i - 25)^2}}$$

$$= \frac{\frac{1}{10}(144)}{\sqrt{\frac{1}{10}(154)} \times \sqrt{\frac{1}{10}(162)}}$$

$$= \frac{14.4}{\sqrt{15.4 \times 16.2}} = \frac{14.4}{15.79494}$$

$$\boxed{r(x, y) = 0.911684}$$

Step III : $\sigma_x = \sqrt{\frac{1}{n}\sum (x_i - 30)^2} = \sqrt{15.4} = 3.9243$

$\sigma_y = \sqrt{\frac{1}{n}\sum (y_i - 25)^2} = \sqrt{16.4} = 4.0249$

Step IV :

1. Lines of regression of y on x is

$$(y - \bar{y}) = r(x, y)\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$$\Rightarrow (y - 25) = \frac{(0.911684)(4.0249)}{(3.9243)}(x - 30)$$

$$(y - 25) = 0.93505(x - 30)$$

$$\boxed{y = 0.93505x - 3.0515}$$

Step V :

2. Lines of regressions of x on y

$$(x - \bar{x}) = r(x, y)\frac{\sigma_x}{\sigma_y}(y - \bar{y});$$

$$(x - 30) = \frac{(0.911684)(3.9423)}{(4.0249)}(y - 25)$$

$$x - 30 = 0.89297(y - 25)$$

$$\boxed{x = (0.89297)y + 7.67575}$$

Step VI :

Acute angle between the regression lines is given by,

$$\tan\theta = \left|\frac{(1 - r^2)}{r} \cdot \frac{\sigma_x \cdot \sigma_y}{(\sigma_x^2 + \sigma_y^2)}\right|$$

$$= \left[\frac{1 - (0.911684)^2}{(0.911684)}\right] \cdot \frac{(3.9243) \cdot (4.0249)}{[(3.9243)^2 + (4.0249)^2]}$$

$$= \left(\frac{0.168832}{0.911684}\right) \times \left(\frac{15.79491}{31.59995}\right)$$

$$\tan\theta = 0.09257$$

$$\Rightarrow \theta = \tan^{-1}(0.09257) = 5.29^\circ \checkmark \quad \text{...Ans.}$$

Ex. 5.7.2 : Two lines of regressions are $5y - 8x + 17 = 0$ and $2y - 5x + 14 = 0$. If $\sigma_y^2 = 16$.

Find : (i) The mean values of x and y,

(ii) σ_x^2,

(iii) Correlation coefficient r (x, y)

Soln. :

Step I :

Given : $\sigma_y^2 = 16 \Rightarrow \sigma_y = \pm 4$

The given regression lines are

$-8x + 5y = -17$ and $-5x + 2y = -14$

i.e. $8x - 5y = 17$ ...(1)

$5x - 2y = 14$ ...(2)

Step II :

(i) Let $(\bar{x}, \bar{y})$ be the point of intersection of lines (1) and (2),

Therefore, $(\bar{x}, \bar{y})$ satisfies the Equations (1) and (2)

$\therefore$ put $x = \bar{x}$ and $y = \bar{y}$ in Equations (1) and (2)

$\therefore 8\bar{x} - 5\bar{y} = 17$ and $5\bar{x} - 2\bar{y} = 14$

Solving simultaneously, we get

$\bar{x} = 4; \quad \bar{y} = 3$

mean $\bar{x} = 4$ and $\bar{y} = 3$

Step III :

(ii) Let us consider the lines of regression of y on x, from Equation (1)

$$y = \frac{8}{5}x - \frac{17}{5} = (1.6)x - \frac{17}{5}$$

and lines of regression of x on y from Equation (2)

$$x = \frac{2}{5}y + \frac{14}{5} = (0.4)y + \frac{14}{5}$$

$\therefore b_{yx} = 1.6$ and $b_{xy} = 0.4$

Step IV : $r(x, y) = \sqrt{b_{yx} \cdot b_{xy}} = 0.8$

$$\boxed{r(x, y) = 0.8}$$

Step V :

(iii) To find σ_x

$$\therefore \sigma_x = \frac{b_{xy} \times \sigma_y}{r(x, y)} = \frac{(0.4) \times (4)}{0.8} = 2$$

$$\sigma_x^2 = 4 \checkmark \quad \text{...Ans.}$$

Ex. 5.7.3 : The equation of two regression lines obtained in a correlation analysis are $4x - 5y + 33 = 0$ and $20x - 9y - 107 = 0$. If the variance of y is 16 then find.

(i) The mean values of x and y

(ii) The correlation coefficient between x and y

(iii) The variance of x

Soln. :

Step I : Given :

$\sigma_y^2 = 16 \Rightarrow \sigma_y = \pm 4$

Regression lines are

$4x - 5y = -33$ ...(1)

$20x - 9y = 107$ ...(2)

Step II :

(i) Let $(\bar{x}, \bar{y})$ be the point of intersection of given regression lines

$\therefore$ put, $x = \bar{x}$ and $y = \bar{y}$

$\therefore 4\bar{x} - 5\bar{y} = -33$

and $20\bar{x} - 9\bar{y} = 107$; solving these two equations.

$\bar{x} = 13; \quad \bar{y} = 17$

$\therefore$ mean values of x and y are $\bar{x} = 13$; $\bar{y} = 17$

Step III :

(ii) Let us consider the lines of regression of y on x is from Equation (1),

$$y = \frac{4}{5}x + \frac{33}{5} = 0.8x + \frac{33}{5}$$

$$\boxed{\therefore b_{yx} = 0.8}$$

and the lines of regression of x on y is from Equation (2),

$$x = \frac{9}{20}y + \frac{107}{20} = 0.45y + \frac{107}{20}$$

$$\boxed{\therefore b_{xy} = 0.45}$$

Step IV :

$\therefore$ Correlation co-efficient r(x, y)

$$r(x, y) = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{0.8 \cdot (0.45)}$$

$$\boxed{r = 0.6}$$

Step V :

(iii) Since, $\sigma_y = 4$

$$\sigma_x = \frac{b_{xy} \times \sigma_y}{r(x, y)} = \frac{(0.45)(4)}{0.6} = 3$$

$\therefore$ variance in x is σ_x^2

$$\boxed{\sigma_x^2 = 9} \checkmark \quad \text{...Ans.}$$

Ex. 5.7.4 : The equation of two lines are $2x = 8 - 3y$ and $2y = 5 - x$. Find the mean values of x and y. Obtain the value of correlation co-efficient.

Soln. :

Step I : Given, regression lines are

$2x + 3y = 8$ ...(1)

and $x + 2y = 5$ ...(2)

Step II : Since, the given lines are intersect at $(\bar{x}, \bar{y})$

$\therefore$ put $x = \bar{x}$ and $y = \bar{y}$ in Equations (1) and (2)

$\therefore 2\bar{x} + 3\bar{y} = 8$

and $\bar{x} + 2\bar{y} = 5$; solving, we get

$\therefore\ \bar{x} = 1\ ;\ \bar{y} = 2$

mean values of x and y are

$$\boxed{\bar{x} = 1\ ;\ \bar{y} = 2}$$

Step III : Let, the regression line of x on y is from Equation (1)

$$x = -\frac{3}{2}y + 4$$

$$\boxed{b_{xy} = -\frac{3}{2} = -1.5}$$

Step IV : and the regression line of y on x is from Equation (2),

$$y = -\frac{1}{2}x + \frac{5}{2}$$

$$\therefore\ \boxed{b_{yx} = -0.5}$$

Step V :

$$\therefore\ r(x, y) = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{(-1.5) \times (-0.5)}$$

$$\boxed{r(x, y) = 0.8680}\ \checkmark \qquad \text{...Ans.}$$

Ex. 5.7.5 : If two lines of regression are $9x + y - \lambda = 0$ and $4x + y = \mu$ and the means of x and y are 2 and – 3 respectively. Find the values of λ and μ and coefficient of correlation between x and y. Also find regression co-efficient b_{yx} and b_{xy}.

Soln. :

Step I :

Given : $\bar{x} = 2$ and $\bar{y} = -3$

The regression lines are

$$9x + y = \lambda \qquad \text{...(1)}$$

and $$4x + y = \mu \qquad \text{...(2)}$$

Step II : Let $(\bar{x}, \bar{y})$ be the point of intersection of given two regression lines

$\therefore$ put $x = \bar{x}$ and $y = \bar{y}$

$\therefore\ 9\bar{x} + \bar{y} = \lambda \Rightarrow 9(2) - 3 = \lambda \Rightarrow \lambda = 15$

and $4\bar{x} + \bar{y} = \mu \Rightarrow 4(2) - 3 = \mu \Rightarrow \mu = 5$

$\therefore$ Equations (1) and (2) becomes

$$9x + y = 15 \qquad \text{...(3)}$$

and $$4x + y = 5 \qquad \text{...(4)}$$

Step III : Let the regression line of y on x is from Equation (4),

$$y = -4x + 15$$

$\Rightarrow$ b_{yx} = regression coefficient of y on x

$$\boxed{b_{yx} = -4}$$

and the regression line of x on y is from Equation (3)

$$x = -\frac{y}{9} + \frac{15}{9}$$

$$\boxed{b_{xy} = -\frac{1}{9}} = -0.1111$$

is the coefficient of regression of x on y

Step IV : $\therefore\ r = \sqrt{4 \times 0.1111}$

$$\boxed{r(x, y) = 0.6666}\ \checkmark \qquad \text{...Ans.}$$

Ex. 5.7.6 : The regression equations are : $8x - 10y + 66 = 0$ and $40x - 18y = 214$. The value of variance of x is 9. Find :

(i) The mean values of x and y

(ii) The correlation co-efficient between x and y.

(iii) The Standard deviation of y

Soln. :

Step I : The given regression lines are ;

$$8x - 10y = -66 \qquad \text{...(1)}$$

and $$40x - 18y = 214 \qquad \text{...(2)}$$

Step II : Since, the regression lines (1) and (2) passes through $(\bar{x}, \bar{y})$

Therefore, from Equations (1) and (2)

$$8(\bar{x}) - 10(\bar{y}) = -66$$

and $$40(\bar{x}) - 18(\bar{y}) = 214$$

we get, $\bar{x} = 13$ and $\bar{y} = 17$

Hence, mean $(\bar{x}) = 13$ and $(\bar{y}) = 17$

Step III : Let us consider Equation (1) is the lines of regression of y on x and Equation (2) is the lines of regression of x on y.

Therefore, from Equation (1)

$$y = \frac{8}{10}x + \frac{66}{10} \qquad \text{...(3)}$$

and from Equation (2)

$$x = \frac{18}{40}y + \frac{214}{40} \qquad \text{...(4)}$$

Step IV :

(i) b_{yx} = Regression coefficient of y on x $= \frac{8}{10} = 0.8$

(ii) b_{xy} = Regression coefficient of x on y $= \frac{18}{40} = 0.45$

Step V : The correlation coefficient between x and y is

$$r(x, y) = \sqrt{b_{yx} \times b_{yx}}$$

$$= \sqrt{(0.8) \times (0.45)} = 0.6$$

(since both the regression coefficient are positive)

Step VI :

Given, $\sigma_x^2 = 9$, $\therefore\ \sigma_x = \pm 3$

Take, $\sigma_x = 3$

$$\because\ b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$$

$$\Rightarrow\ \sigma_y = \frac{b_{yx} \times \sigma_x}{r} = \frac{(0.8) \times (3)}{0.6} = 4$$

$\therefore\ \sigma_y = 4$

Hence standard deviation of y = 4 ✓ ...Ans.

Ex. 5.7.7 : Find the coefficient of correlation for distribution in which $\sigma_x = 4$, $\sigma_y = 1.8$ and $b_{yx} = 0.32$

Soln. :

Given : $\sigma_x = 4$, $\sigma_y = 1.8$, $b_{yx} = 0.32$

Since, $$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$\Rightarrow\ r(x, y) = \frac{b_{yx}\sigma_x}{\sigma_y} = \frac{(0.32)(4)}{(1.8)}$$

$$\boxed{r(x, y) = 0.711}\ \checkmark \qquad \text{...Ans.}$$

Ex. 5.7.8 : If $\bar{x} = 8.2$; $\bar{y} = 12.4$; $\sigma_x = 6.2$; $\sigma_y = 20$; $(x, y) = 0.9$, then find the lines of regression. Estimate the value of x for y = 10 and estimate y for x = 10.

Soln. :

Step I : Given, $\bar{x} = 8.2$; $\bar{y} = 12.4$; $\sigma_x = 6.2$; $\sigma_y = 20$; $r(x, y) = 0.9$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = (0.9)\frac{(20)}{(6.2)} = 2.9032$$

and $$b_{xy} = r\frac{\sigma_x}{\sigma_y} = (0.9)\frac{(6.2)}{(20)} = 0.279$$

Step II : Regression lines of y on x is

$$(y - \bar{y}) = r\frac{\sigma_y}{\sigma_x}(y - \bar{y})$$

$\Rightarrow\ y - (12.4) = (2.9032)(x - 8.2) = (2.9032)x - 23.80624$

$y = (2.903)x - 11.40624$

at $x = 10$

$$\boxed{y = 17.62576}$$

Step III : The regression line of x on y is

$$(x - \bar{x}) = r(x, y)\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$\Rightarrow\ (x - 8.2) = (0.279)(y - 12.4) = (0.279)y - 3.4596$

$$\Rightarrow\ \boxed{x = (0.279)y + 4.7404}$$

at $y = 10$

$$\boxed{x = 7.5304}\ \checkmark \qquad \text{...Ans.}$$

Ex. 5.7.9 : The two regression equations of the variables x and y are $x = 19.3 - 0.87y$; $y = 11.64 - 0.50x$. Find (i) $\bar{x}$, $\bar{y}$ (ii) The correction coefficient between x and y.

Soln. :

Step I : Given regression lines are,

$$x = 19.3 - 0.87y \qquad \text{...(1)}$$

and $$y = 11.64 - 0.50x \qquad \text{...(2)}$$

Step II : Equation (1) is regression line of x and y

$\therefore$ From Equation (1), $\boxed{b_{xy} = -0.87}$

and Equation (2) is regression line of y on x

$\therefore$ From Equation(2), $\boxed{b_{yx} = -0.50}$

Step III : Coefficient of correlation,

$$r(x, y) = \sqrt{b_{xy} \cdot b_{yx}}$$

$$= \sqrt{(-0.87) \cdot (-0.50)} = 0.6596$$

Step IV : Replace $x = \bar{x}$, $y = \bar{y}$ in Equations (1) and (2) then,

By solving Equations (1) and (2), we get

$\bar{x} = 16.2358$ and $\bar{y} = 3.5222$ ✓ ...Ans.

Ex. 5.7.10 : If $\Sigma x_i = 30$, $\Sigma y_i = 40$, $\Sigma x_i^2 = 220$; n = 5, $\Sigma y_i^2 = 340$, $\Sigma x_i y_i = 214$, then obtain the regression lines for this data.

Soln. :

Step I :

Given : $\Sigma x_i = 30$; $\Sigma y_i = 40$, $\Sigma x_i^2 = 220$; $\Sigma y_i^2 = 340$, n = 5, $\Sigma x_i y_i = 214$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{30}{5} = 6 \text{ and } \bar{y} = \frac{\Sigma y_i}{n} = \frac{40}{5} = 8$$

Step II : The coefficient of correlation

$$r(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\frac{1}{n}\Sigma x_i y_i - (\bar{x}\bar{y})}{\sqrt{\frac{\Sigma x_i^2}{n} - (\bar{x})^2} \cdot \sqrt{\frac{\Sigma y_i^2}{n} - (\bar{y})^2}}$$

$$= \frac{\frac{1}{5}(214) - (6 \times 8)}{\sqrt{\left(\frac{220}{5}\right) - (6)^2} \times \sqrt{\left(\frac{340}{5}\right) - (8)^2}}$$

$$= \frac{-5 \cdot 2}{\sqrt{8} \times \sqrt{4}} = \frac{-5 \cdot 2}{\sqrt{32}}$$

$$r(x, y) = -0.9193$$

$$\sigma_x = \sqrt{\frac{1}{n}(\Sigma x_i)^2 - (\bar{x})^2}$$

$$= \sqrt{\left(\frac{220}{5}\right) - (6)^2} = \sqrt{8}$$

$$\sigma_x = 2.8285$$

and $\sigma_y = \sqrt{\frac{1}{n}(\sum y_i)^2 - (\bar{y})^2} = \sqrt{\frac{1}{5}(340) - (8)^2}$

$= \sqrt{4} = 2$

Step III : Lines of regression of y on x is,

$$(y-\bar{y}) = r \cdot \frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

$$\Rightarrow \quad (y-8) = (-0.9193) \times \frac{(2)}{(2.8285)}(x-6)$$

$$(y-8) = -0.6500\,(x-6)$$

$$\therefore \quad y = (-0.65)\,x + 11.9$$

and lines of regression of x on y is

$$(x-\bar{x}) = r \cdot \frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$$(x-6) = (-0.9193) \times \frac{(2.8285)}{2}(y-8)$$

$$(x-6) = -(1.300)\,(y-8)$$

$$\Rightarrow \quad x = (-1.3)\,y + 16.4$$

The equation of two lines of regression obtained in a correlation analysis are the following :

$2x + 3y - 8 = 0$ and $x + 2y - 5 = 0$ ✓ ...Ans.

Ex. 5.7.11 : Obtain the value of the correlation co-efficient and the variance of y given that variance of x is 12.

Soln. :

(i) Correlation Co-efficient

Step I : Given, regression lines are

$2x + 3y = 8$ ...(1)

and $x + 2y = 5$ ...(2)

Step II :

Let, the regression line of x on y is from Equation (1)

$$x = -\frac{3}{2}y + 4$$

$$b_{xy} = -\frac{3}{2} = -1.5$$

Step III : and the regression line of y on x is from Equation (2),

$$y = -\frac{1}{2}x + \frac{5}{2} \qquad \therefore \quad b_{yx} = -0.5$$

Step III:

$$\therefore \; r(x,y) = -\sqrt{b_{xy} \cdot b_{yx}} = -\sqrt{(-1.5) \times (-0.5)}$$

(Since, both b_{xy} and b_{yx} are negative)

$$r(x,y) = -0.8680$$

Step IV :

(ii) variance of y :

Since, $\sigma^2_x = 12 \qquad \therefore \; \sigma_x = 3.4641$

$$b_{xy} = \frac{r(x,y)\,\sigma_x}{\sigma_y}$$

$$-\frac{3}{2} = \frac{(-0.8680)\,(3.4641)}{\sigma_y}$$

$$\sigma_y = 2$$

$\therefore$ variance in y is $= \sigma^2_y$

$\sigma^2_x = 4$ ✓ ...Ans.

Ex. 5.7.12 : Calculate the coefficient of correlation between the marks obtained by 8 students in mathematics and statistics from the following table. Also find the lines of regression.

	A	B	C	D	E	F	G	H
Marks in Maths (x)	25	30	32	35	37	40	42	45
Marks in Stats (y)	8	10	15	17	20	22	24	25

Soln. : PART- I

Step I : Since, we know

Regression line of y on x is

$$(y-\bar{y}) = \frac{r\,\sigma_y}{\sigma_x}(x-\bar{x}) \qquad ...(1)$$

And regression lines of x on y is ,

$$(x-\bar{x}) = r\,\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \qquad ...(2)$$

Step II :

where, $r = \frac{\text{cov}(x,y)}{\sigma_x \cdot \sigma_y}$

$$= \frac{\frac{1}{n}\sum x_i y_i - (\bar{x}\bar{y})}{\sqrt{\frac{1}{n}\sum x_i^2 - (\bar{x})^2} \cdot \sqrt{\frac{1}{n}\sum y_i^2 - (\bar{y})^2}}$$

i.e. $$r = \frac{\sum x_i y_i - n(\bar{x}\bar{y})}{\sqrt{\sum (x_i^2) - n(\bar{x})^2} \cdot \sqrt{\sum (y_i^2) - n(\bar{y})^2}}$$

Here , n = 8

PART-II

Step III : Construct the table as per formula:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
25	8	625	64	200
30	10	900	100	300
32	15	1024	225	480
35	17	1225	289	595
37	20	1369	400	740
40	22	1600	484	880
42	24	1764	576	1008
45	25	2025	625	1125
$\sum x_i = 286$	$\sum y_i = 141$	$\sum x_i^2 = 10532$	$\sum y_i^2 = 2763$	$\sum x_i y_i = 5328$

Step IV : $\bar{x} = \frac{\sum x_i}{n} = \frac{286}{8} = 35.75$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{141}{8} = 17.625$$

Step V :

$$r(x,y) = \frac{(5328) - 8\,(35.75)\,(17.625)}{\sqrt{(10532) - 8\,(35.75)^2} \times \sqrt{(2763) - 8\,(17.625)^2}}$$

$$\boxed{r(x,y) = 0.9826}$$

PART-III

Step VI : Lines of regression of y on x , from equation (1),

$$(y-\bar{y}) = r\,\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

$$(y-17.625) = (0.9825)\,\frac{\sqrt{\frac{1}{n}\sum y_i^2 - (\bar{y})^2}}{\sqrt{\frac{1}{n}\sum x_i^2 - (\bar{x})^2}} \times (x-35.75)$$

$$= (0.9825)\,\frac{\sqrt{\frac{1}{8}(2763) - (17.625)^2}}{\sqrt{\frac{1}{8}(10532) - (35.75)^2}} \times (x-35.75)$$

$$= (0.9825)\,\frac{(5.894)}{(6.1998)} \times (x-35.75)$$

$$(y-17.625) = 0.9341\,(x-35.75)$$

$$\boxed{\therefore \; y - 0.9341x + 15.769 = 0}$$

Step VII : Lines of regression of x on y is

$$(x-\bar{x}) = r\,\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$$(x-35.75) = (0.9825) \cdot \frac{(6.1998)}{(5.894)} \times (y-17.625)$$

$$= (1.034)\,(y-17.625)$$

$$\boxed{x - 1.034\,y - 17.5275 = 0}$$ ✓ ...Ans.

Ex. 5.7.13 : Obtain lines of regression for the following data.

x_i	6	2	10	4	8
y_i	9	11	5	8	7

Find y for x = 5 and x for y = 10.

Soln. :

Step I : Since,

Regression line of y on x is,

$$(y-\bar{y}) = \frac{r\,\sigma_y}{\sigma_x}(x-\bar{x}) \qquad ...(1)$$

And regression lines of x on y is ,

$$(x-\bar{x}) = r\,\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \qquad ...(2)$$

Step II :

where, $b_{yx} = \frac{\text{Cov}(x,y)}{\sigma^2_x}$ and $b_{xy} = \frac{\text{Cov}(x,y)}{\sigma^2_y}$

$$\sigma^2_x = \frac{1}{n}\sum x_i^2 - (\bar{x})^2 \;;$$

$$\sigma^2_y = \frac{1}{n}\sum y_i^2 - (\bar{y})^2$$

$$\bar{x} = \frac{\sum x_i}{n} \;; \qquad \bar{y} = \frac{\sum y_i}{n} \;;$$

Here $n = 5$; $\text{Cov}(x,y) = \frac{1}{n}\sum x_i y_i - (\bar{x}\bar{y})$

Step III : Now construct the table as per formula :

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
$\sum x_i = 30$	$\sum y_i = 40$	$\sum x_i^2 = 220$	$\sum y_i^2 = 340$	$\sum x_i y_i = 214$

Step IV : $\bar{x} = \frac{30}{5} = 6$; $\bar{y} = \frac{40}{5} = 8$

$\sigma^2_x = \frac{220}{5} - (6)^2 = 8$ and $\sigma^2_y = \frac{340}{5} - (8)^2 = 4$

$$\text{Cov}(x,y) = \frac{214}{5} - (6)\,(8) = -5.2$$

$$\therefore \; b_{xy} = \frac{(-5.2)}{4} = -1.3$$

and $$b_{yx} = \frac{(-5.2)}{8} = -0.65$$

Step V : Regression line of y on x is,

$$(y-\bar{y}) = b_{yx}(x-\bar{x}) = (-0.65)\,(x-6)$$

$$y - 8 = -0.65\,x + 3.9$$

$$\boxed{\text{at } x = 5 \;,\; y = 8.65}$$

Step VI : Regression line of x on y is,

$$(x-\bar{x}) = b_{xy}(y-\bar{y})$$

$$(x-6) = (-1.3)\,(y-8)$$

$$(x-6) = -(1.3)\,y + 10.4$$

$$\boxed{\text{at } y = 10 \;,\; x = 3.4}$$ ✓ ...Ans.

Ex. 5.7.14 : Calculate the regression equation of x on y and y on x from the following data and estimate x when y = 26 and estimate y when x = 21.

x_i	10	12	13	17	18	20	24	30
y_i	5	6	7	9	13	15	20	21

Soln. :

Step I : Since, regression line of y on x is

$$\left(y-\bar{y}\right) = \frac{r(x,y)\,\sigma_y}{\sigma_x}\left(x-\bar{x}\right) \quad ...(1)$$

and regression line of x on y is

$$\left(x-\bar{x}\right) = \frac{r(x,y)\,\sigma_x}{\sigma_y}\left(y-\bar{y}\right) \quad ...(2)$$

Step II :

Where, $r(x,y) = \dfrac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$

$$Cov(x,y) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$$

$$\sigma_x = \sqrt{\frac{1}{n}\sum x_i^2 - \left(\bar{x}\right)^2}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum y_i^2 - \left(\bar{y}\right)^2}$$

$$\bar{x} = \frac{\sum x_i}{n};\ \frac{\sum y_i}{n} = \bar{y}$$

Here, $n = 8$

Step III : Construct the table as per formula :

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
10	5	100	25	50
12	6	144	36	72
13	7	169	49	91
17	9	289	81	153
18	13	324	169	234
20	15	400	225	300
24	20	576	400	480
30	21	900	441	630
$\sum x_i = 144$	$\sum y_i = 96$	$\sum x_i^2 = 2902$	$\sum y_i^2 = 1426$	$\sum x_i y_i = 2010$

Step IV :

$$\bar{x} = \frac{144}{8} = 18\ ;\ \bar{y} = \frac{96}{8} = 12$$

$$\sigma_x = \sqrt{\frac{1}{8}(2902) - (18)^2} = 6.22495$$

$$\sigma_y = \sqrt{\frac{1}{8}(1426) - (12)^2} = 5.85235$$

$$Cov(x,y) = \frac{1}{8}(2010) - (18)(12) = 35.25$$

Coefficient of correlation

$$\therefore\ r(x,y) = \frac{(35.25)}{(6.22495)\times(5.85235)} = \frac{35.25}{36.43058}$$

$$\boxed{r(x,y) = 0.96759}$$

Step IV : From Equation (1) lines of regression of y on x is

$$(y-12) = \frac{(0.96759)\times(5.85255)}{(6.22495)}(x-18)$$

$$y-12 = \left(\frac{5.66268}{6.22495}\right)(x-18)$$

$$\boxed{y-12 = 0.90968\,(x-18)}$$

at $x = 21$; $\quad y = (0.90968)\,x - 4.37424$

$$\boxed{y = 14.72904}$$

Step V : From Equation (2), lines of regression of x on y is

$$(x-18) = \frac{(0.96759)\times(6.22495)}{(5.85235)}\times(y-12)$$

$$= (1.0292)(y-12)$$

$$\boxed{x = (1.0292)\,y + 5.6496}$$

at $y = 26$ $\quad \boxed{x = 32.4088}$ ✓ ...Ans.

Ex. 5.7.15 : Calculate the regression equation of y on x from the following data.

x	10	14	18	22	26	30
y	18	12	24	6	30	36

Soln. :

Step I : Since, regression line of y on x is

$$\left(y-\bar{y}\right) = \frac{r(x,y)\,\sigma_y}{\sigma_x}\left(x-\bar{x}\right) \quad ...(1)$$

Where, $r(x,y) = \dfrac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$

Step II :

$$Cov(x,y) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$$

$$\sigma_x = \sqrt{\frac{1}{n}\sum x_i^2 - \left(\bar{x}\right)^2}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum y_i^2 - \left(\bar{y}\right)^2}$$

$$\bar{x} = \frac{\sum x_i}{n};\ \frac{\sum y_i}{n} = \bar{y}$$

Here, $n = 6$

Step III :

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
10	18	100	324	180
14	12	196	144	168
18	24	324	576	432
22	6	484	36	132
26	30	676	900	780
30	36	900	1296	1080
$\sum x_i = 120$	$\sum y_i = 126$	$\sum x_i^2 = 2680$	$\sum y_i^2 = 3276$	$\sum x_i y_i = 2772$

Step IV : $\quad \bar{x} = \dfrac{120}{6} = 20\ ;\ \bar{y} = \dfrac{126}{6} = 21$

$$\sigma_x = \sqrt{\frac{1}{6}(2680) - (20)^2} = 6.8313$$

$$\sigma_y = \sqrt{\frac{1}{6}(3276) - (21)^2} = 10.2470$$

$$Cov(x,y) = \frac{1}{6}(2772) - (20)(21) = 42$$

Step V : Coefficient of correlation

$$\therefore\ r(x,y) = \frac{(42)}{(6.8313)\times(10.2470)} = \frac{42}{70.003}$$

$$r(x,y) = 0.6$$

Step VI : From Equation (1) lines of regression of y on x is

$$(y-21) = \frac{(0.6)\times(10.2470)}{(6.8313)}(x-20)$$

$$= \left(\frac{6.1482}{6.8313}\right)(x-20)$$

$$y-21 = 0.9\,(x-20)\ ✓$$ ...Ans.

Ex. 5.7.16 : Obtain lines of regression for the following data.

x	6	2	10	4	8
y	9	11	5	8	7

Soln. :

Step I : Since regression lines are

$$\left(y-\bar{y}\right) = b_{yx}\left(x-\bar{x}\right) \quad ...(1)$$

and $\left(x-\bar{x}\right) = b_{xy}\left(y-\bar{y}\right) \quad ...(2)$

Step II :

where, $b_{yx} = \dfrac{Cov(x,y)}{\sigma_x^2}$ and $b_{xy} = \dfrac{Cov(x,y)}{\sigma_y^2}$

Step III :

$$\sigma_x^2 = \frac{1}{n}\sum x_i^2 - \left(\bar{x}\right)^2 \sigma_y^2 = \frac{1}{n}\sum y_i^2 - \left(\bar{y}\right)^2$$

$$\bar{x} = \frac{\sum x_i}{n}\ ;\quad \bar{y} = \frac{\sum y_i}{n}\ ;$$

Here $n = 5$; $Cov(x,y) = \dfrac{1}{n}\sum x_i y_i - \left(\bar{x}\bar{y}\right)$

Step IV :

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
$\sum x_i = 30$	$\sum y_i = 40$	$\sum x_i^2 = 220$	$\sum y_i^2 = 340$	$\sum x_i y_i = 214$

Step V : $\quad \bar{x} = \dfrac{30}{5} = 6\ ;\ \bar{y} = \dfrac{40}{5} = 8$

$$\sigma_x^2 = \frac{220}{5} - (6)^2 = 8$$

$$\sigma_y^2 = \frac{340}{5} - (8)^2 = 4$$

Step VI :

$$Cov(x,y) = \frac{214}{5} - (6)(8) = -5.2$$

$$\therefore\ b_{xy} = \frac{(-5.2)}{4} = -1.3,\quad b_{yx} = \frac{(-5.2)}{8} = -0.65$$

Step VII : Regression line of y on x is,

$$\left(y-\bar{y}\right) = b_{yx}\left(x-\bar{x}\right) = (-0.65)(x-6)$$

$$y-8 = -0.65\,x + 3.9$$

at $x = 5$, $y = 8.65$

Step VIII : Regression line of x on y is,

$$\left(x-\bar{x}\right) = b_{xy}\left(y-\bar{y}\right)$$

$$(x-6) = (-1.3)(y-8)$$

$$(x-6) = -(1.3)\,y + 10.4$$

at $y = 10$

$x = 3.4$ ✓ ...Ans.

Ex. 5.7.17 : Find the correlation co-efficient and value of y at $x = 60$.

Given, $\bar{x} = 53.2\ ;\ \bar{y} = 27.9\ ;\ b_{yx} = -1.5,\ b_{xy} = -0.2$

Soln. :

Step I :

Given, $\bar{x} = 53.2$; $\quad \bar{y} = 27.9$

$b_{yx} = -1.5$; $\quad b_{xy} = -0.2$

Step II : Co-efficient of correlation

$$r = \pm\sqrt{b_{xy}\cdot b_{yx}} = \pm\sqrt{(-1.5)\cdot(-0.2)}$$

$$\therefore\ r = \pm 0.548$$

Since, both b_{yx} and b_{xy} are negative, r must be negative

$$\therefore\ r = -0.548$$

Step III : We know, the lines of regression of y on x is

$$(y-\bar{y}) = b_{yx}(x-\bar{x})$$

$$\Rightarrow (y-27.9) = (-1.5)(x-53.2)$$

$$y = (-1.5)\,x + 79.8 + 27.9$$

$$y = (-1.5)\,x + 107.7$$

Step IV :

At $x = 60\ ;\ y = (-1.5)(60) + 107.7$

$= -90 + 107.7 = 17.7$

Exercise

Ex. 1 : Find the line of regressions from the following data:

Students	1	2	3	4	5	6	7	8	9	10
Marks in Economics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Estimate marks in statistics if marks in Economics are 30.

Ex. 2 : Find the lines of regression for the following data :

x	10	14	19	26	30	34	39
y	12	16	18	26	29	35	38

Estimate y for x = 14.5 and x for y = 29.5.

Ex. 3 : From the record of correlation data, the following results are available :

$\sigma_x^2 = 9$, regression lines are : $8x - 10y + 66 = 0$;

$40x - 18y = 214$

Find (i) Mean values of x and y

(ii) S. D. for x

(iii) Correlation co-efficient between x and y

Ex. 4 : If r (x, y) = 0.8 ; cov (x, y) = 20 ; $\sigma_x^2 = 16$.

Find standard deviation of y.

Multiple Choice Questins

Q. 1 $\bar{x}, \bar{y}$ and σ_x, σ_y are mean and standard deviations of two series x and y respectively. If r (x, y) is coefficient of correlation between x and y then equation of line of regression of y on x is given by _____.

(a) $y + \bar{y} = r\frac{\sigma_y}{\sigma_x}(x + \bar{x})$

(b) $y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$

(c) $y - \bar{y} = r\frac{\sigma_x}{\sigma_y}(x + \bar{x})$

(d) $x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$

Ans. : (b)

Explanation : The line of regression of y on x is

$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$.

Q. 2 $\bar{x}, \bar{y}$ are mean of two series x and y respectively. If b_{xy} and b_{yx} are regression coefficients then equation of line of regression of y on x is _____.

(a) $y - \bar{y} = b_{yx}(x - \bar{x})$ (b) $y - \bar{y} = b_{yx}(x + \bar{x})$

(c) $x - \bar{x} = b_{xy}(y - \bar{y})$ (d) $y - \bar{y} = b_{xy}(x + \bar{x})$

Ans. : (a)

Explanation : If b_{xy} and b_{yx} are regression coefficients then the line of regression of y on x is

$y - \bar{y} = b_{yx}(x - \bar{x})$

Q. 3 $\bar{x}, \bar{y}$ and σ_x, σ_y are mean and standard deviations of two series x and y respectively. If r (x, y) is coefficient of correlation between x and y then equation of line of regression of x on y is _____.

(a) $x - \bar{x} = r\frac{\sigma_y}{\sigma_x}(y - \bar{y})$ (b) $y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$

(c) $x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$ (d) $(y + \bar{y}) = r\frac{\sigma_x}{\sigma_y}(x + \bar{x})$

Ans. : (c)

Explanation : The line of regression of y on x is

$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$

Q. 4 $\bar{x}, \bar{y}$ are mean of two series x and y respectively. If b_{xy} and b_{yx} are regression coefficients then equation of line of regression x on y is _____.

(a) $y - \bar{y} = b_{yx}(x - \bar{x})$ (b) $x + \bar{x} = r\, b_{yx}(y - \bar{y})$

(c) $x - \bar{x} = b_{xy}(y + \bar{y})$ (d) $x - \bar{x} = b_{xy}(y - \bar{y})$

Ans. : (d)

Explanation : If b_{xy} and b_{yx} are regression coefficients then the line of regression of y on x is

$x - \bar{x} = b_{xy}(y - \bar{y})$.

Q. 5 If σ_x, σ_y are standard deviations of two series x and y respectively. r (x, y) is coefficient of correlation between x and y then regression coefficient b_{xy} is _____.

(a) $r(\sigma_x + \sigma_y)$ (b) $r\frac{\sigma_x}{\sigma_y}$

(c) $r\frac{\sigma_y}{\sigma_x}$ (d) $\frac{\sigma_x \sigma_y}{r}$

Ans. : (b)

Explanation : If σ_x, σ_y are standard deviations of two series x and y respectively. r (x, y) is coefficient of correlation between x and y then regression coefficient $b_{xy} = r\frac{\sigma_x}{\sigma_y}$

Q. 6 If σ_x, σ_y are Standard Deviations of two series x and y respectively r (x, y) is coefficient of correlation between x and y then regression coefficient b_{yx} is _____.

(a) $r(\sigma_x + \sigma_y)$ (b) $r\frac{\sigma_x}{\sigma_y}$

(c) $\frac{\sigma_x \sigma_y}{r}$ (d) $r\frac{\sigma_y}{\sigma_x}$

Ans. : (d)

Explanation : σ_x, σ_y are Standard Deviations of two series x and y respectively r (x, y) is coefficient of correlation between x and y then regression coefficient $b_{yx} = r\frac{\sigma_y}{\sigma_x}$.

Q. 7 The value of b_{xy} is _____.

(a) $\frac{\text{cov}(x,y)}{\sigma_x^2}$ (b) $\frac{\text{cov}(x,y)}{\sigma_y^2}$

(c) cov (x, y) (σ_x) (d) $\sigma_y \cdot$ cov (x, y)

Ans. : (b)

Explanation : The value of $b_{xy} = \frac{\text{cov}(x,y)}{\sigma_y^2}$.

Q. 8 The value of b_{yx} is _____.

(a) $\frac{\text{cov}(x,y)}{\sigma_y^2}$ (b) σ_x cov (x, y)

(c) σ_y cov (x, y) (d) $\frac{\text{cov}(x,y)}{\sigma_x^2}$

Ans. : (d)

Explanation : The value of $b_{yx} = \frac{\text{cov}(x,y)}{\sigma_x^2}$.

Q. 9 If coefficient of regression b_{xy} and b_{yx} both are positive then value of coefficient of correlation r(x, y) is —

(a) Positive (b) Negative

(c) 0 (d) −1

Ans. : (a)

Explanation : If coefficient of regression b_{xy} and b_{yx} both are positive then value of coefficient of correlation r(x, y) is always positive.

Q. 10 If coefficient of regression b_{xy} and b_{yx} both are negative then value of coefficient of correlation r(x, y) is —

(a) Positive (b) Negative (c) 0 (d) 1

Ans. : (b)

Explanation : If coefficient of regression b_{xy} and b_{yx} both are negative then value of coefficient of correlation r(x, y) is always negative.

Q. 11 If coefficients of correlation are b_{xy} and b_{yx}, then correlation coefficient r(x, y) is —

(a) $\frac{b_{xy}}{b_{yx}}$ (b) $\frac{b_{yx}}{b_{xy}}$ (c) $\pm\sqrt{b_{xy} b_{yx}}$ (d) $b_{xy} b_{yx}$

Ans. : (c)

Explanation : If coefficient of correlation are b_{xy} and b_{yx}, then correlation coefficient $r(x, y) = \pm\sqrt{b_{xy} b_{yx}}$.

Q. 12 The relation between cov (x, y), standard deviation of x (i.e. σ_x) and regression coefficient b_{yx} is —

(a) $\text{cov}(x,y) = \frac{b_{xy}}{\sigma_x^2}$ (b) $\text{cov}(x,y) = \frac{b_{xy}}{\sigma_y^2}$

(c) $b_{yx} = \frac{\text{cov}(x,y)}{\sigma_x^2}$ (d) $b_{xy} = \frac{\text{cov}(x,y)}{\sigma_y^2}$

Ans. : (c)

Explanation : Since, we know $b_{yx} = \frac{\text{cov}(x,y)}{\sigma_x^2}$.

Q. 13 The relation between cov (x, y), standard deviation of y (i.e. σ_y) and regression coefficient b_{xy} is —

(a) $\text{cov}(x,y) = \frac{b_{xy}}{\sigma_x^2}$ (b) $\text{cov}(x,y) = \frac{b_{xy}}{\sigma_y^2}$

(c) $b_{yx} = \frac{\text{cov}(x,y)}{\sigma_x^2}$ (d) $b_{xy} = \frac{\text{cov}(x,y)}{\sigma_y^2}$

Ans. : (d)

Explanation : Since, we know $b_{xy} = \frac{\text{cov}(x,y)}{\sigma_y^2}$.

Q. 14 If the two regression lines are perpendicular to each other, then their coefficient of correlation r is _____.

(a) 1 (b) 0 (c) −1 (d) ±1

Ans. : (b)

Explanation : If r = 0,

i.e. $\tan\theta \to \infty \Rightarrow \theta \to \pi/2$

If the variables are independent, then the two lines of regression are perpendicular to each other. i.e. the variables are more spread means not correlated..

Q. 15 If coefficient of correlation r is zero then two regression lines are _____.

(a) parallel to X - axis

(b) Perpendicular to X – axis

(c) Parallel to each other

(d) Perpendicular to each other

Ans. : (d)

Explanation : If r = 0,

i.e. $\tan\theta \to \infty \Rightarrow \theta \to \pi/2$

If the variables are independent, then the two lines of regression are perpendicular to each other.

Q. 16 If θ is the angle between two regression lines then tan θ = _____.

(a) $\left(\frac{1+r^2}{r}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)$ (b) $\frac{1-r^2}{r}$

(c) $\left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)$ (d) $\frac{1+r^2}{r}$

Ans. : (c)

Explanation : If θ is the acute angle between the two regression lines for two variables x and y,

then, $\tan\theta = \left|\left(\frac{1-r^2}{r}\right)\cdot\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right|$.

Q. 17 If coefficient of correlation $r = \pm 1$ then two regression lines are

(a) Perpendicular (b) coincide

(b) parallel to X - axis (c) parallel to Y-axis

Ans. : (b)

Explanation : If $r = \pm 1$,

i.e. $\theta \to 0 \Rightarrow \theta = 0$ or π

i.e. The given two lines are coincides or parallel. There is a perfect correlation between the two variables.

Q. 18 If one of the regression coefficient is greater then unity then other regression coefficient is

(a) 1 (b) <1 (c) >1 (d) 0

Ans. : (b)

Explanation : If one of the regression co-efficient is greater than unity, other must be less than unity.

i.e. If $b_{xy} > 1$ then $b_{yx} < 1$

Q. 19 Slope of lines of regression of y on x is,

(a) $b_{xy} \cdot b_{yx}$ (b) $r\frac{\sigma_x}{\sigma_y} = b_{xy}$

(c) r(x, y) (d) $r\frac{\sigma_y}{\sigma_x} = b_{yx}$

Ans. : (d)

Explanation :

The line of regression of y on x is :

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

Compare it with $x = my + c$, we get slope $m = r\frac{\sigma_y}{\sigma_x}$

Q. 20 Slope of line of regression of x on y is,

(a) $b_{xy} = r\frac{\sigma_x}{\sigma_y}$ (b) $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

(c) r(x, y) (d) None of the above

Ans. : (a)

Explanation : The line of regression of x on y is :

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

Compare it with $y = mx + c$, we get slope $m = r\frac{\sigma_x}{\sigma_y}$

Q. 21 If the following table gives the information about advertisement expenditure and sales :

	Advertisement expenditure (x) (lakhs)	Sales (y) (lakhs)
Mean	10	90
S.D.	3	12

and correction co-efficient = 0.8. Two lines of regression are

(a) $x = -8 + 0.2y$ and $y = 58 + 3.2x$

(b) $x = 8 - 0.2y$ and $y = 58 - 3.2x$

(c) $x = 53 + 3.2y$ and $y = 8 + 0.2x$

(d) None of the above

Ans. : (a)

Explanation :

Given, $\bar{x} = 10$, $\bar{y} = 90$; $\sigma_x = 3$;

$\sigma_y = 12$; $r = 0.8$

(i) The lines of regression of Y on X is

$$(y - \bar{y}) = \left(r \cdot \frac{\sigma_y}{\sigma_x^2}\right)(x - \bar{x})$$

(ii) The lines of regression or X on Y is :

$$(x - \bar{x}) = \left(r \cdot \frac{\sigma_x}{\sigma_y^2}\right)$$

Substitute the value of r, mean and S.D. in above equations (i) and (ii) ,we get required lines

Q. 22 For a given distribution, if value of cov (x, y) = – 5.2, $\sigma_x = 2.82$ the value of regression coefficient b_{yx} is ——

(a) –0.55 (b) –0.85 (c) –0.75 (d) –0.65

Ans. : (d)

Explanation :

We have, $b_{yx} = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{-5.2}{(2.85)^2} = -0.65$

Q. 23 For a given distribution, cov (x, y) = 35.25, $\sigma_y = 5.82$ the value of regression coefficient b_{xy} is ——

(a) 1.80 (b) 2.8 (c) 1.04 (d) 2

Ans. : (c)

Explanation :

We have, $b_{yx} = \frac{\text{cov}(x, y)}{\sigma_y^2} = \frac{35.25}{(5.82)^2} = 1.04$

Q. 24 For a given distribution, if mean of two series x and y are respectively $\bar{x} = 1$, $\bar{y} = 2$ and regression coefficient $b_{xy} = -1.5$ then by using equation, of line of regression of x on y, the value of x when y = 4 is – ——

(a) –2 (b) 2 (c) 1 (d) –1

Ans. : (a)

Explanation : The equation of line of regression x on y is given by

$x - \bar{x} = b_{xy}(y - \bar{y})$; $x - 1 = -1.5(y - 2)$

for y = 4

$x - 1 = -1.5(4 - 2)$ $\therefore x = -2$

Q. 25 For given distribution, if mean of two series x and y are respectively $\bar{x} = 8.2$, $\bar{y} = 12.4$, $b_{yx} = 2.92$ then by using equation of line of regression of y on x, the value of y when x = 10 is ——

(a) 15.42 (b) 17.65 (c) 18.60 (d) 14.13

Ans. : (b)

Explanation : We have by equation of line of regression y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x});$$

$\therefore y - 12.4 = 2.92(x - 8.2)$

For $x = 10$, $y - 12.4 = 2.92(10 - 8.2)$

$\therefore y = 5.256 + 12.4 = 17.65$

Q. 26 The two regression equation of the variables x and y are $x = 19.3 - 0.87y$, $y = 11.64 - 0.50x$ then values of $\bar{x}$ and $\bar{y}$ are _____.

(a) $\bar{x} = 11.64$; $\bar{y} = 19.3$ (b) $\bar{x} = 0.50$; $\bar{y} = 0.87$

(c) $\bar{x} = 16.23$; $\bar{y} = 3.52$ (d) $\bar{x} = 17.3$; $\bar{y} = 4.50$

Ans. : (c)

Explanation : Two regression lines are

$x = 19.3 - 0.87y$;

$y = 11.64 - 0.50x$

$\therefore x + 0.87y = 19.3$;

$0.50x + y = 11.64$

Since $\bar{x}$ and $\bar{y}$ satisfies these equations

$\bar{x} + 0.87\bar{y} = 19.3$

$0.50\bar{x} + \bar{y} = 11.64$

$\therefore \bar{x} = 16.23, \bar{y} = 3.52$

Q. 27 The two regression equation of the variables x and y are $x = 19.3 - 0.87y$; $y = 11.64 - 0.50x$ then correlation coefficient between x and y is _____.

(a) 0.8 (b) 0.7 (c) – 0.66 (d) + 0.66

Ans. : (c)

Explanation : Two regression lines are

$x = 19.3 - 0.87y$;

$y = 11.64 - 0.50x$

$\therefore b_{xy} = -0.87$; $b_{yx} = -0.50$

$r = \sqrt{b_{xy} b_{yx}} = 0.66$

Since both b_{xy} and b_{yx} are negative $\therefore r = -0.66$

Q. 28 Two regression lines are $5y - 8x + 17 = 0$ and $2y - 5x + 14 = 0$ then mean values of x and y are

(a) $\bar{x} = 3$; $\bar{y} = 4$ (b) $\bar{x} = 8$; $\bar{y} = 2$

(c) $\bar{x} = 5$; $\bar{y} = 5$ (d) $\bar{x} = 4$; $\bar{y} = 3$

Ans. : (d)

Explanation : Given : $\sigma_y^2 = 16 \Rightarrow \sigma_y = \pm 4$

The given regression lines are

$-8x + 5y = -17$ and $-5x + 2y = -14$

i.e. $8x - 5y = 17$ ...(1)

$5x - 2y = 14$ ...(2)

(i) Let $(\bar{x}, \bar{y})$ be the point of intersection of lines (1) and (2),

Therefore, $(\bar{x}, \bar{y})$ satisfies the Equations (1) and (2)

$\therefore$ put $x = \bar{x}$ and $y = \bar{y}$ in Equations (1) and (2)

$\therefore 8\bar{x} - 5\bar{y} = 17$ and $5\bar{x} - 2\bar{y} = 14$

Solving simultaneously, we get

$\bar{x} = 4$; $\bar{y} = 3$

mean $\bar{x} = 4$ and $\bar{y} = 3$

Q. 29 Two regression lines are $5y - 8x + 17 = 0$ and $2y - 5x + 14 = 0$ then the coefficient of correlation between x and y is _____.

(a) 0.8 (b) 0.9 (c) 0.6 (d) 0.7

Ans. : (a)

Explanation : Let us consider the lines of regression of y on x, from Equation (1)

$$y = \frac{8}{5}x - \frac{17}{5} = (1.6)x - \frac{17}{5}$$

and lines of regression of x on y from Equation (2)

$$x = \frac{2}{5}y + \frac{14}{5}$$

$$= (0.4)y + \frac{14}{5}$$

$\therefore b_{yx} = 1.6$ and $b_{xy} = 0.4$

$r(x, y) = 0.8$

Q. 30 Two regression lines are $5y - 8x + 17 = 0$ and $2y - 5x + 14 = 0$ with $b_{xy} = 0.4$, $r(x, y) = 0.8$ and $\sigma_y^2 = 16$ then variance of x is

(a) 8 (b) 2 (c) 16 (d) 4

Ans. : (d)

Explanation : Since, $\sigma_x = \frac{b_{xy} \times \sigma_y}{r(x, y)} = \frac{(0.4) \times (4)}{0.8} = 2$

$\sigma_x^2 = 4$

$\therefore$ variance in x is $\sigma_x^2 = 4$

Q. 31 Two regression lines are $8x - 10y + 66 = 0$; $40x - 18y = 214$ then mean values of x and y are

(a) $\bar{x} = 8$, $\bar{y} = 18$ (b) $\bar{x} = 40$, $\bar{y} = 10$

(c) $\bar{x} = 13$, $\bar{y} = 17$ (d) $\bar{x} = 9$, $\bar{y} = 11$

Ans. : (c)

Explanation : The regression lines are

$8x - 10y + 66 = 0$ and $40x - 18y = 214$

Since mean values of x and y are $\bar{x}, \bar{y}$ satisfies these equations.

$8\bar{x} - 10\bar{y} = -66 \Rightarrow 40\bar{x} - 18\bar{y} = 214$

$\Rightarrow \therefore \bar{x} = 13 ; \quad \bar{y} = 17$

Q. 32 Two regression lines are $8x - 10y + 66 = 0$; $40x - 18y = 214$ then coefficient of correction is

(a) 0.7 (b) 0.8 (c) 0.6 (d) 0.9

Ans. : (c)

Explanation : The lines of regression lines are

$8x - 10y + 66 = 0 ; 40x - 18y = 214$

The regression line of y on x : y

$$= \frac{8}{10}x + \frac{66}{10} \Rightarrow y = 0.8x + 6.6$$

The regression line of x on y :

$$x = \frac{18}{40}y + \frac{214}{40} \Rightarrow x = 0.45y + 5.35$$

$\therefore b_{yx} = 0.8$ and $b_{xy} = 0.45$

$\therefore r = \sqrt{b_{xy} b_{yx}} = 0.6$

Q. 33 Two regression lines are $8x - 10y + 66 = 0$; $40x - 18y = 214$ also variance of x = 9 then variance of y is _____.

(a) 9 (b) 8 (c) 4 (d) 16

Ans. : (d)

Explanation : The regression lines are

$8x - 10y + 66 = 0$ and $40x - 18y = 214$

From Q. 97 we have, $b_{xy} = 0.45$, $r(x, y) = 0.6$

Now, Variance of $x = 9 \Rightarrow \sigma_x^2 = 9 \Rightarrow \sigma_x = 3$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} \Rightarrow \sigma_y = \frac{r\sigma_x}{b_{xy}}$$

$$\sigma_y = 4 \quad \Rightarrow \sigma_y^2 = 16$$

$\therefore$ Variance of y = 16

Q. 34 The equation of two regression lines obtained in a correlation analysis are $4x - 5y + 33 = 0$; $20x - 9y - 107 = 0$ then mean values of x and y are

(a) $\bar{x} = 4 ; \bar{y} = 9$ (b) $\bar{x} = 13 ; \bar{y} = 17$

(c) $\bar{x} = 20 ; \bar{y} = 5$ (c) $\bar{x} = 3 ; \bar{y} = 4$

Ans. : (b)

Explanation : Given, $\sigma_y^2 = 16 \Rightarrow \sigma_y = \pm 4$

Regression lines are

$4x - 5y = -33$ ...(1)

$20x - 9y = 107$ ...(2)

(i) Let $(\bar{x}, \bar{y})$ be the point of intersection of given regression lines

$\therefore$ put, $x = \bar{x}$ and $y = \bar{y}$

$\therefore 4\bar{x} - 5\bar{y} = -33$

and $20\bar{x} - 9\bar{y} = 107$; solving these two equations.

$\bar{x} = 13 ; \quad \bar{y} = 17$

Q. 35 The equation of two regression lines obtained in correlation analysis are $4x - 5y + 33 = 0$; $20x - 9y - 107 = 0$ then the correlation coefficient between x and y is _____.

(a) 0.7 (b) 0.6 (c) −0.8 (d) 0.9

Ans. : (b)

Explanation : Let us consider the lines of regression of y on x is from Equation (1),

$$y = \frac{4}{5}x + \frac{33}{5} = 0.8x + \frac{33}{5}$$

$\therefore \boxed{b_{yx} = 0.8}$

and the lines of regression of x on y is from Equation (2),

$$x = \frac{9}{20}y + \frac{107}{20} = 0.45y + \frac{107}{20}$$

$\boxed{\therefore b_{xy} = 0.45}$

$\therefore$ Correlation co-efficient r(x, y)

$$r(x, y) = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{0.8 \cdot (0.45)}$$

$$r = 0.6$$

Q. 36 If regression coefficient y on x is − 0.50 and x on y is − 0.87 then coefficient of correlation is _____.

(a) 0.8 (b) − 0.66 (c) − 0.3 (d) 0.9

Ans. : (b)

Explanation : $b_{yx} = -0.50$ and $b_{xy} = -0.87$

$$r = \sqrt{b_{xy} b_{yx}} = \sqrt{(-0.50) \times (-0.87)}$$

$$= 0.66$$

Since both b_{yx} and b_{xy} are negative

$\therefore r = -0.66$

Q. 37 If the two regression coefficients are 0.8 and 0.45 then coefficient of correlation r is _____.

(a) 0.3 (b) 0.8 (c) 0.9 (d) 0.6

Ans. : (d)

Explanation : $b_{yx} = 0.8$ and $b_{xy} = 0.45$

$$r = \sqrt{b_{xy} b_{yx}} = \sqrt{(0.8 \times 0.45)} = 0.6$$

Chapter Ends...

□□□

CHAPTER 6

UNIT VI

Applied Statistics

Syllabus

Curve Fitting by the Method of Least Square - Fitting of Straight Lines, Second degree parabolas and More General Curves, Test of Significance : Large Sample Test for single Proportion, Difference of Proportion, Single Mean, Difference of Means and Difference of Standard Deviations.

6.1 Introduction

Generally in Engineering, Economics, Agriculture, Social, problems are expressed as set of values of variables (x and y). But it is required to express a given data in the form of law with functional relationship between the involved variables. We can express the given data (set of values of variables X and Y), in the form of many equations of different types approximately.

Here we are finding the equation of the curve of "best fit" which may be most suitable with the given data and helpful to find unknown values. The process of finding such functional relationship (equation) of "best fit" is called "curve fitting". As we know, if there are n pairs of observed / experimental values, then it may be n arbitrary constants in n simultaneous equations. Now, the functional relation may be in linear form (y = ax + b), parabolic form ($y = ax^2 + bx + c$), exponential form : ($y = e^{ax}$, $y = ax^n$) or logarithmic form (y = log (ax + b)).

There are four methods to fit a curve to the observed data :

(i) Graphical Method,

(ii) Method of Least Squares,

(iii) Method of Group averages

(iv) Method of Moments

6.1.1 General Rules (Laws) Reducible to Linear form (Straight Line)

There are some common laws which can be reduced in the linear form by using some suitable substitution or modification.

When the given data is from observation, then these points may not approximate to straight line at that time we can draw a smooth curve through these points and then from shape of curve we may try to write law which can be converted into linear form.

(i) **For the rule : $y = ax^n + b$**

Here, we can put $x^n = X$ and $y = Y$, the given law reduced to $Y = aX + b$

(ii) **For the rule : $y = ax^n$**

Here, we can take logarithm on both the sides, we get

$$\log_{10} y = \log_{10}(ax^n) = \log_{10} a + \log_{10} x^n$$

$$= \log_{10} a + n \log_{10} x$$

Now, put $\log_{10} x = X$;

$\log_{10} y = Y$;

$\log_{10} a = b$

The reduced law is,

$$Y = nX + b$$

(iii) **For the curve $y = ax^n + b \log x$**

The given law can be written as

$$\frac{y}{\log x} = \frac{ax^n}{\log x} + b$$

Put, $\frac{y}{\log x} = Y$;

$\frac{x^n}{\log x} = X$,

The reduced form is : $Y = aX + b$

(iv) **For the rule $y = m e^{nx}$**

Taking log on sides, we get

$$\log_{10} y = \log_{10}(me^{nx}) = \log_{10} m + \log_{10} e^{nx}$$
$$= \log_{10} m + n(\log_{10} e)\, x$$

Put, $\log_{10} y = Y$;

$\log_{10} m = a$;

$n \log_{10} e = b$; and $x = X$

We get,

$Y = a + bX$, which is linear

(v) For the rule, $xy = ax + by$

In this case, divide by x,

$$y = a + b\frac{y}{x}$$

Put $\frac{y}{x} = X$ and $y = Y$,

The reduced rule is,

$Y = a + bX$

6.1.2 Fitting a Straight Line by Graphical Method

It is a very simple method, when the law which is representing the given data is in linear form ($y = ax + b$), where a is slope and b is y intercept of the straight line. Use following points.

Step I : Write the given law and convert it into linear form.

Step II : Plot the given points on paper by taking a desirable scale

Step III : Draw the straight line such that, maximum points lies on the line and points are evenly distributed about the line.

Step IV : Consider any two points say, A (x_1, y_1) and B (x_2, y_2) on the line.

Step V : Calculate the slope of the line $\left(\text{by using } \frac{y_2 - y_1}{x_2 - x_1}\right)$, we get "a"

Step VI : Find y intercept "b".

Note : The graphical method fails to give the unique and accurate values of unknowns a and b.

Solved Examples

Ex. 6.1.1 : Find the law of the type $R = C + mV^2$ connecting R and V, using the following data :

V (miles/hours)	10	20	30	40	50
R (lb/ton)	8	10	15	21	30

Where, R = is the resistance ; V = be the speed of vehicle.

Soln. :

Step 1 : The given law is,

$R = C + mV^2$ ...(1)

Now, convert it into linear form,

Put $V^2 = x$

and $R = y$ in Equation (1),

we get, $y = mx + C$ ...(2)

Which is linear equation (straight line)

Step II : To construct table for R, V, x, y

V	R	$x = V^2$	$y = R$
10	8	100	8
20	10	400	10
30	15	900	15
40	21	1600	21
50	30	2500	30

Step III : Plot these new points on paper

Scale : X-axis : 1 cm = 400 miles/h

Y-axis : 1 cm = 10/lb/tan

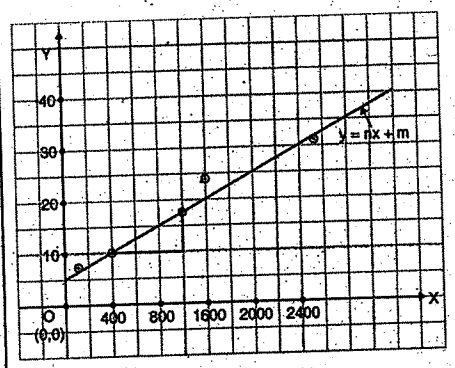


Fig. P. 6.1.1

Step IV : Form Fig. P. 6.1.1,

Let us consider two points for slope (m)

$$\text{slope (m)} = \frac{y_2 - y_1}{x_2 - x_1}$$
$$= \frac{15 - 10}{900 - 400} = \frac{5}{500} = 0.01$$

(since, (400, 10) and (900, 15) lies on the straight line (2))

$\therefore$ $m = 0.01$

Step V : To find C :

From Equation (2)

$y = mx + C$

$10 = (0.01)(400) + C$

$C = 10 - 4 = 6$

Step VI : The required law is,

$y = mx + C = 0.1x + 6$

Ex. 6.1.2 : The resistance R of a carbon filament lamp was measured at various values of the voltage V and the following observations were made.

Voltage V	62	70	78	84	92
Resistance R	73	70.7	69.2	67.8	66.3

Assuming a law of the form $R = \frac{a}{V} + b$, find by graphical method the best value of a and b.

Soln. :

Step I : Given law is : $R = \frac{a}{V} + b$ ...(1)

Now reduce it into linear form by putting $\frac{1}{V} = x$ and $R = y$,

We get,

$Y = ax + b$ ...(2)

Which is linear equation (straight line)

Step II : To construct table for V, R, x, y

V	R	$x = \frac{1}{V}$	$Y = R$
62	73	0.0162	73
70	70.7	0.0143	70.7
78	69.2	0.0129	69.2
84	67.8	0.0120	67.8
92	66.3	0.0110	66.3

Step III : Plot the new points on paper

Scale : X-axis : 1 cm = 0.001 voltage

Y-axis : 1 cm = 1 ohms

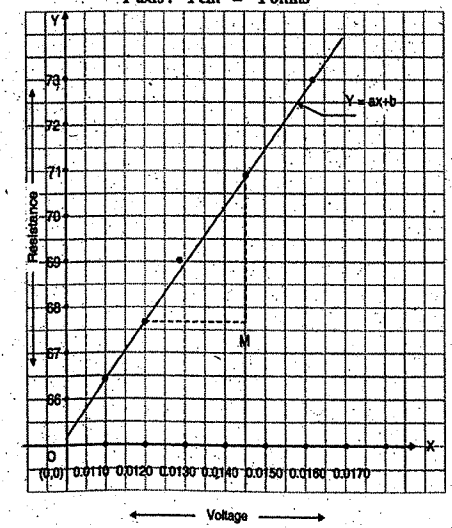


Fig. P. 6.1.2

Step IV : From Fig. P. 6.1.2

Let us consider two points for slope a

$$\text{Slope (a)} = \frac{y_2 - y_1}{x_2 - x_1} = \left(\frac{70.7 - 67.8}{0.0143 - 0.0120}\right)$$
$$= \frac{2.9}{0.0023} \quad a = 1260.87$$

Step V : To find value of b

From Equation (2)

$Y = ax + b \Rightarrow 67.8 = (1260.87)(0.0120) + b$

$67.8 = 15.130 + b \Rightarrow b = 52.67$

(Since, (0.0120, 67.8) lies on the line)

Step VI : The required law is,

$Y = ax + b = (1260.87)x + (52.67)$

Syllabus Topic : Curve Fitting by the Method of Least Square

6.2 Curve Fitting by the Method of Least Square

Since, the graphical method can not give the unique values of constant, so we are unable to find the best fit. Least square method able to determine the best possible values of the constants. So, we can fit a unique curve to a given data. Here, first we will see the Least Square Principal and then we find the best fit of Straight line, parabolas and exponential curves.

6.2.1 Principle of Least Squares

The principle of least square states that the straight line should be fitted through the given points such that the sum of squares of the distances of the points from the straight line (i.e. $\sum e_i^2 = E$) measured along the Y-axis must be minimum.

Hence, the line which is obtained by using least square principle (i.e. e_i's are so small) is called the line of best fit.

Suppose, $P_1(x_1, y_1), P_2(x_2, y_2), P_3(x_3, y_3), \ldots P_i(x_i, y_i), \ldots P_n(x_n, y_n)$ are the given set of observed (or experimental) values. Let, $y = mx + c$ be the equation of the line of best fit. The line P_1L_1 intersects the line

$y = mx + c$ at Q_1

So, the co-ordinates of Q_1 are $(x_1, mx_1 + c)$

Also the line P_2L_2, P_3L_3, P_nL_n intersects the given line

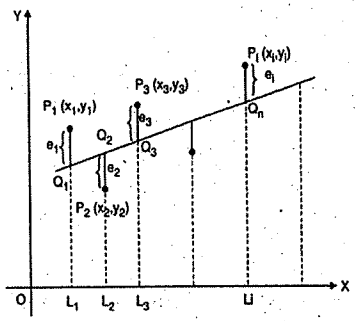


Fig. 6.2.1

At $Q_2, Q_3, \ldots, Q_n$

Therefore, $Q_2\ (x_2,\ mx_2 + c)$, $Q_3\ (x_3,\ mx_3 + c), \ldots,$ $Q_i\ (x_i, mx_i + c)$

The values $Q_1\ L_1$, $Q_2\ L_2$, $Q_3\ L_3$, ..., Q_iL_i are called expected (or calculated) values

(Since, at experimental value x_i ; we get $y_i = mx_i + c \leftarrow$ calculated value)

The difference between the observed and calculated values.

i.e. $L_iP_i - L_iQ_i = e_i$ **is called error or residual at** $x = x_i$

Now the error may be positive or sometimes negative. For equal weightage to each error, we square each error

$$E = (L_1P_1 - L_1Q_1)^2 + (L_2P_2 - L_2Q_2)^2 + (L_3P_3 - L_3Q_3)^2 + \ldots + (L_iP_i - L_iQ_i)^2 + \ldots + (L_nP_n - L_nQ_n)^2$$

$$= [y_1 - (mx_1 + c)]^2 + [y_2 - (mx_2 + c)]^2 + \ldots + [y_i - (mx_i + c)]^2$$

$$= e_1^2 + e_2^2 + e_3^2 + \ldots + e_i^2 + \ldots + e_n^2 = \sum e_i^2$$

is the sum of squares of the error, which must be **minimum.**

Syllabus Topic : Fitting of Straight Line by the Method of Least Squares

6.3 Fitting of Straight Line by the Method of Least Squares

Suppose, $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, $P_3(x_3, y_3)$...., P_n, $P_i(x_y, y_i)$, ... be the set of observed values. (from Fig. 5.8.2)

Let, $y = mx + c$ be the equation of line of best fit. The line $P_i\ L_i$ intersects the line $y = mx + c$ at Q_i, so the co-ordinates of Q_i are $(x_i, mx_i + c)$

In the method of least square, we have to determine m and c such that the $E = \sum e_i^2$ **is minimum.** This must be true for all points. So, we should have $E = \sum (y_i - (mx_i + c))^2$ is **minimum.** (So this method is called least squares)

For least square, from calculus i.e. by the concept of maxima and minima for two variables, the condition is

$$\frac{\partial E}{\partial m} = 0 \quad \text{and} \quad \frac{\partial E}{\partial c} = 0$$

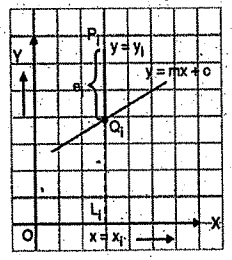


Fig. 6.3.1

Since, $E = \sum (y_i - mx_i - c)^2$ ...(1)

Differentiate Equation (1) partially w.r.t. m and c respectively, we get

$$\frac{\partial E}{\partial m} = 0 = 2\sum (y_i - mx_i - c)(-x_i)$$

i.e. $\sum (y_i - mx_i - c)(x_i) = 0$

i.e. $\sum y_i x_i - m\sum x_i^2 - c\sum x_i = 0$

$$\therefore \sum x_i y_i = c\sum x_i + m\sum x_i^2 \quad \ldots(2)$$

and $\frac{\partial E}{\partial c} = 0 = 2\sum (y_i - mx_i - c)(-1)$

$$\Rightarrow \sum (y_i - mx_i - c) = 0$$

$$\therefore \sum y_i = m\sum x_i + \sum c = m\sum x_i + nc \quad \ldots(3)$$

where, n-be the number of observations. By droping the suffix i, Equation (2) and (3) becomes

$$\sum y = nc + m\sum x \quad \ldots(4)$$

$$\text{and } \sum xy = c\sum x + m\sum x^2 \quad \ldots(5)$$

These equations are called **Normal Equations.** Find the values of constants m and c. Put these values of constants in Equation (1), we get the desired curve of best fit.

6.3.1 There are different Methods for Finding the Constants m and c

(a) Direct Method

- In this method find first : $\sum x, \sum x^2, \sum y, \sum xy$.
- Put these values in normal Equations (4) and (5).
- We get two equations of two unknown m and c.
- Solve these two simultaneous equations, we get two constants m and c. Put these values of m and c in the equation of straight line, we get the desired curve of best fit.

(b) Assumed Mean Method : (Deviation Method)

In this method first we assume the means of variable x and y.

Let a and b are assumed means of x and y respectively.

$X = x - a$ and $Y = y - b$ are the deviations in X and Y.

Then find $\sum X, \sum Y, \sum X^2, \sum XY$

- Find the values of c and m like previous method in section 5.8.5.1(a)

(If $x_i - a$ have common factor say h and $y_i - b$ is k then sometimes, we can divide the deviations by common factors. Then $X = \frac{x-a}{h}$ and $Y = \frac{y-b}{k}$)

Generally, when the values of X and Y are evenly spaced, deviation method is more convenient.

(c) Actual Mean Method

Here, find the mean of x and y, as $\bar{x}$ and $\bar{y}$ respectively.

Write $X = x_i - \bar{x}$ and

$$\bar{Y} = y_i - \bar{y} \left(\text{or } X = \frac{x_i - \bar{x}}{h} \text{ and } Y = \frac{y_i - \bar{y}}{k}\right)$$

Find $\sum X, \sum Y, \sum X^2, \sum XY$

Find the values of m and c by using 5.8.5.1(a)

Solved examples

Ex. 6.3.1 : If P is the pull required to lift a load W by means of pulley block, find a linear law of the form $P = mW + C$, connecting P and W, using the following data,

P	12	15	21	25
W	50	70	100	120

Where P and W are taken in kg weight

Soln. :

Step I : To find the constants m and C by Direct Method : The given law is

$$P = mW + C \quad \ldots(1)$$

The corresponding normal equation are

$$\sum P = nC + m\sum W \quad \ldots(2)$$

$$\text{and } \sum PW = C\sum W + m\sum W^2 \quad \ldots(3)$$

Here $n = 4$

Step II : Construct table for $\sum P, \sum W, \sum W^2, \sum PW$

P	W	W^2	PW
12	50	2500	600
15	70	4900	1050
21	100	10000	2100
25	120	14400	3000
$\sum P = 73$	$\sum W = 340$	$\sum W^2 = 31800$	$\sum PW = 6750$

Step III : From Equation (2),

$$73 = 4C + m(340) \text{ and from Equation (3),}$$

$$6750 = C(34.0) + m(31800)$$

Therefore, we get,

$$4C + 340\,m = 73 \quad \ldots(4)$$

$$\text{and } 34C + 3180\,m = 675 \quad \ldots(5)$$

To eliminate C,

$34 \times$ Equation (4) $- 4 \times$ Equation (5)

$$136C + 11560\,m = 2482$$

$$-136C - 12720\,m = 2700$$

$$-1160\,m = -218$$

$$\Rightarrow m = 0.1879$$

Put this value of m in Equation (4), we get

$$4C = 73 - 340\,m = 73 - (340)\cdot(0.1879)$$

$$C = \frac{9.114}{4}$$

$$\therefore C = 2.2785$$

Step IV :

Hence, the line of best fit is, from Equation (1),

$$P = 2.2785 + 0.1879\,W \checkmark \quad \ldots\text{Ans.}$$

Ex. 6.3.2 : Fit a straight line to the following data :

(x, y) : (1, 1), (2, 5), (3, 11), (4, 8), (5, 14)

by (a) Direct method

(b) Assumed mean method

(c) Actual mean method

Soln. :

Step I : Let the equation of straight line be

$$Y = mx + C \quad \ldots(1)$$

To find constants m and C by

(a) Direct Method

Since the normal equations are

$$\sum y = nC + m\sum x \quad \ldots(2)$$

$$\text{and } \sum xy = C\sum x + m\sum x^2 \quad \ldots(3)$$

Here, $n = 5$

Step II : Construct table for $\sum x, \sum y, \sum x^2$

x	y	x^2	xy
1	1	1	1
2	5	4	10
3	11	9	33
4	8	16	32
5	14	25	70
$\sum x = 15$	$\sum y = 39$	$\sum x^2 = 55$	$\sum xy = 146$

Step III : From Equation (2)

$$39 = 5C + 15m \quad ...(4)$$

and $$146 = 15C + 55m \quad ...(5)$$

To eliminate C first

$\therefore$ 15 × Equation (4) − 5 × Equation (5)

$$585 = 75C + 225m$$
$$-730 = -75C - 275m$$
$$-145 = -50m$$
$$\Rightarrow \quad m = 2.9$$

Put this value of m in Equation (4)

$$39 - 15m = 5C$$
$$39 - 15 \times (2.9) = 5C$$
$$-4.5 = 5C$$
$$C = -0.9$$

Step IV : Equation of line of best fit is :

$$Y = mx + C = (2.9)x - 0.9$$

(b) Assumed mean method

Step V :

Here, we will assume 3 and 11 as means of x and y

$$\therefore X = x - a = x - 3 \quad \text{and} \quad Y = y - b = y - 11$$

The corresponding normal equations are

$$\sum Y = nC + m\sum X \quad ...(6)$$

and $$\sum XY = C\sum X + m\sum X \quad ...(7)$$

Step VI : Construct table for $\sum X, \sum Y, \sum X^2, \sum XY$

x	y	X = x − 3	Y = y − 11	X^2	XY
1	1	−2	−10	4	20
2	5	−1	−6	1	6
3	11	0	0	0	0
4	8	1	−3	1	−3
5	14	2	3	4	6
		$\sum X = 0$	$\sum Y = -16$	$\sum X^2 = 10$	$\sum XY = 29$

Step VII : From Equation (6)

$$-16 = 5C + 0$$

and $$29 = 0 + m(10)$$

$$\Rightarrow C = -3.2$$

and $$m = 2.9$$

Step VIII : Hence, equation of line of best fit is

$$Y = mX + C$$
$$y - 11 = m(x - 3) + C$$
$$y - 11 = (2.9)(x - 3) - 3.2$$
$$= (2.9)x - 8.7 - 3.2$$

Hence, $$y = (2.9)x - 0.9.$$

(c) Actual mean method

Step IX : Here, we have find actual mean $\bar{x}$ and $\bar{y}$ of x and y respectively.

$$\therefore \bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3 \quad \text{and}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{39}{5} = 7.8$$

$$\therefore X = x - \bar{x} \quad ; Y = y - \bar{y}$$
$$X = x - 3 \quad ; Y = y - 7.8$$

The normal equations are Equation (6) and (7)

Step X : Construct table for $x, y, X = x - \bar{x}, Y = y - \bar{y}, X^2, XY$

x	y	$X = x - \bar{x} = x - 3$	$Y = y - \bar{y} = y - 7.8$	X^2	XY
1	1	−2	−6.8	4	13.6
2	5	−1	−2.8	1	2.8
3	11	0	3.2	0	0
4	8	1	0.2	1	0.2
5	14	2	6.2	4	12.4
		$\sum X = 0$	$\sum Y = 0$	$\sum X^2 = 10$	$\sum XY = 29$

Step XI : From Normal Equation (6)

$$0 = 5C + 0 \Rightarrow \boxed{C = 0}$$

and from Equation (7)

$$29 = 0 + m(10) \Rightarrow m = 2.9$$

Step XII : Hence Equation of line of best fit is

$$Y = mX + C$$
$$y - 7.8 = (2.9)(x - 3) + 0$$
$$\Rightarrow y - 7.8 = (2.9)x - 8.7$$
$$\Rightarrow y = (2.9)x - 0.9 \checkmark \quad ...\text{Ans.}$$

Note : We observed from above example, by using actual mean method we get constant C = 0, directly we can find value of m. So the actual mean method is more convenient.

Ex. 6.3.3 : Fit a straight line Y = mx + C to the following set of values by using assumed mean method.

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Soln. :

Step I : Let the equation of straight line be,

$$Y = mx + C \quad ...(1)$$

To find constants m and C by Assumed Mean Method

Here, we consider 10 and 17 are the assumed means of x and y respectively.

Therefore, $X = x - a$ and $Y = y - b$

$$X = x - 10 \text{ and } Y = y - 17$$

Step II : Since, the corresponding normal equations are

$$\sum Y = nC + m\sum X \quad \text{and} \quad ...(2)$$
$$\sum XY = C\sum X + m\sum X^2 \quad ...(3)$$

Step III : Construct table for $\sum X, \sum Y, \sum XY, \sum X^2$

x	y	X = x − 10	Y = y − 17	X^2	XY
0	12	−10	−5	100	50
5	15	−5	−2	25	10
10	17	0	0	0	0
15	22	5	5	25	25
20	24	10	7	100	70
25	30	15	13	225	195
		$\sum X = 15$	$\sum Y = 18$	$\sum X^2 = 475$	$\sum XY = 350$

From Equation (2),

$$18 = 6C + 15m \quad ...(4)$$

and From (3), $350 = 15C + 475m$ ...(5)

5 × Equation (4) − 2 × Equation (5), we get

$$90 = 30C + 75m$$
$$700 = 30C + 950m$$
$$- \quad - \quad -$$
$$-610 = 0 - 875m$$
$$\Rightarrow m = 0.69714$$
$$\approx m = 0.7$$

Put this value of m in Equation (4),

$$18 - 15m = 6C$$
$$\Rightarrow 18 - 15 \times (0.7) = 6C$$
$$\Rightarrow \frac{7.5}{6} = C$$
$$\Rightarrow C \approx 1.25$$

Step IV : Hence, equation of line of best fit is,

$$y = mx + C$$
$$\Rightarrow y - 17 = (0.7)(x - 10) + 1.25$$
$$\Rightarrow y - 17 = (0.7)x - 7 + 1.25$$
$$y = (0.7)x + 11.25 \checkmark \quad ...\text{Ans.}$$

Ex. 6.3.4 : Fit a straight line y = ax + b to the following data, by using actual mean method.

x	1	2	3	4	5	6
y	49	54	60	73	80	86

Also calculate y for x = 5.6

Soln. :

Step I : Given equation of straight line

$$y = ax + b \quad ...(1)$$

To find constants a and b by Actual Mean Method :

Here, we have to find the means $\bar{x}$ and $\bar{y}$ of x and y respectively

$$\therefore \bar{x} = \frac{\sum x_i}{n} \quad \text{and} \quad \bar{y} = \frac{\sum y}{n}$$

$$\bar{x} = \frac{21}{6} = 3.5 \quad \text{and} \quad \bar{y} = \frac{402}{6} = 67$$

Therefore, $$X = x - \bar{x} = x - 3.5$$
$$Y = y - \bar{y} = y - 67$$

Step II : Since, the corresponding normal equations are

$$\sum Y = nb + a\sum X \quad ...(2)$$

and $$\sum XY = b\sum X + a\sum X^2 \quad ...(3)$$

Construct table for $\sum X, \sum Y, \sum X^2, \sum XY$

x	y	X = x − 3.5	Y = y − 67	X^2	XY
1	49	−2.5	−18	6.25	45
2	54	−1.5	−13	2.25	19.5
3	60	−0.5	−7	0.25	3.5
4	73	0.5	6	0.25	3.0
5	80	1.5	13	2.25	19.5
6	86	2.5	19	6.25	47.5
$\sum x = 21$	$\sum y = 404$	$\sum X = 0$	$\sum Y = 0$	$\sum X^2 = 17.5$	$\sum XY = 138$

Step III : From Equation (2)

$$0 = 6b + 0 \Rightarrow b = 0$$

and $$138 = 0 + (17.5)a \Rightarrow a \approx 7.89$$

Step IV : Hence equation of line of best fit is,

$$Y = ax + b$$
$$y - 67 = (7.89)(x - 3.5) + 0$$
$$y = (7.89)x + 39.38$$

Step V : Now, for x = 5.6 :

$$y = (7.89)(5.6) + 39.38 = 83.56 \checkmark \quad ...\text{Ans.}$$

Ex. 6.3.5 : Fit a straight line to the following data by least square method :

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Soln. :

Step I : Let the equation of straight line be,

$$Y = mx + C \quad ...(1)$$

To find constants m and c by Assumed Mean Method

Here, we consider 10 and 17 are the assumed means of x and y respectively.

Therefore, $X = x - a$ and $Y = y - b$

$X = x - 10$ and $Y = y - 17$

Step II : Since, the corresponding normal equations are

$\sum Y = nC + m\sum X$ and ...(2)

$\sum XY = C\sum X + m\sum X^2$ ...(3)

Step III : Construct table for $\sum X, \sum Y, \sum XY, \sum X^2$

x	y	X = x − 10	Y = y − 17	X^2	XY
0	12	−10	−5	100	50
5	15	−5	−2	25	10
10	17	0	0	0	0
15	22	5	5	25	25
20	24	10	7	100	70
25	30	15	13	225	195
		$\sum X = 15$	$\sum Y = 18$	$\sum X^2 = 475$	$\sum XY = 350$

From Equation (2),

$18 = 6C + 15m$ ...(4)

and From (3),

$350 = 15C + 475m$ ...(5)

5 × Equation (4) − 2 × Equation (5), we get

$90 = 30C + 75m$

$700 = 30C + 950m$

$-\quad -\quad -$

$-610 = 0 - 875m$

$\Rightarrow m = 0.69714$

$\approx m = 0.7$

Put this value of m in Equation (4),

$18 - 15m = 6C$

$\Rightarrow 18 - 15 \times (0.7) = 6C$

$\Rightarrow \frac{7.5}{6} = C$

$\Rightarrow C \approx 1.25$

Step IV : Hence, equation of line of best fit is,

$y = mx + C$

$\Rightarrow y - 17 = (0.7)(x - 10) + 1.25$

$\Rightarrow y - 17 = (0.7)x - 7 + 1.25$

$y = (0.7)x + 11.25$ ✓ ...Ans.

Ex. 6.3.6 : By the method of least squares, find the straight line that best fits the following data :

x	1	2	3	4	5
y	14	27	40	55	68

Soln. :

Step I : Let the equation of straight line is $y = ax + b$ ...(1)

The normal equations of straight line (1) are

$\sum y = a\sum x + nb$ ...(2)

and $\sum xy = a\sum x^2 + b\sum x$ ...(3)

Here, n = 5

Step II : Now, construct table for $\sum x, \sum y, \sum x^2, \sum xy$

x	y	xy	x^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
$\sum x = 15$	$\sum y = 204$	$\sum xy = 748$	$\sum x^2 = 55$

Step III : The normal Equations (2) and (3) becomes

$204 = 15a + 5b$ ...(4)

and $748 = 55a + 15b$ ...(5)

By solving Equations (4) and (5) simultaneously, we get

$a = \frac{68}{5} = 13.6$

and $b = 0$

Hence the best fit equation of straight line (1) becomes

$y = (13.6)x$ ✓ ...Ans.

Exercise

Ex. 1 : If P is pull required to lift the weight by means of a pulley block, find a linear law of the form $P = a + bW$, connecting P and W, using the following data : by graphical method.

W (lb)	50	70	100	120
P (lb)	12	15	21	25

Compute P, when w = 150 lb

Ex. 2 : The values of x and y obtained in an experiment, obey the law $y = a + b\sqrt{x}$. Find graphically the values of a and b.

x	500	1000	2000	4000	6000
y	0.20	0.33	0.38	0.45	0.51

Ex. 3 : Fit a straight line $y = a + bx$ to the following data

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Ex. 4 : Fit a straight line $y = mx + C$ to the following data :

x (years)	1951	1961	1971	1981	1991
y (Production in tons)	10	12	8	10	5

Also calculate the production in 1988.

Ex. 5 : Fit a straight line $y = ax + b$ to the following data

x	1	2	3	4	5	6
y	83	92	71	90	160	191

Use actual mean method give. Also calculate y for x = 3.5

Ex. 6 : Fit a straight line to the following data.

$(x, y) \equiv (-1, -5), (1, 1), (2, 4), (3, 7), (4, 10)$

Estimate y for x = 5

Syllabus Topic : Fitting of Second degree parabolas by the Method of Least Squares

6.4 Fitting of Parabola by the Method of Least Square

Let $P_1(x_1, y_1), P_2(x_2, y_2), P_3(x_3, y_3), \ldots P_i(x_i, y_i)$ be the set of observed values and we have to fit Parabola Curve is of the form $y = a + bx + cx^2$ to these given values. By the least square principle the parabolic curve should be fitted through the given points such that the sum of squares of the distances of the point from parabola (i.e. $\sum e_i^2 = E$) measured along Y-axis must be minimum. So, the curve which is obtained by using least square principle is called the parabolic curve of best fit.

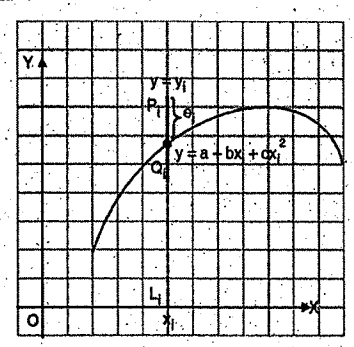


Fig. 6.4.1

The line $P_i L_i$ intersects the parabolic curve $y = a + bx + cx^2$ at point Q_i. So the co-ordinates of Q_i are $(x_i, a + bx_i + cx_i^2)$. We have to find a, b, c such that $E = \sum e_i^2$ is minimum. It must true for all points.

We have, $E = \sum e_i^2 = \sum (y_i - (a + bx_i + cx_i^2))^2$

$E = \sum (y_i - a - bx_i - cx_i^2)^2$ ...(1)

For least square, from calculus (i.e. by maxima and minima), condition is

$\frac{\partial E}{\partial a} = 0$;

$\frac{\partial E}{\partial b} = 0$;

$\frac{\partial E}{\partial c} = 0$

Therefore, differentiate Equation (1) partially w.r.t a, b and c we get

$\frac{\partial E}{\partial a} = 0 = 2\sum(y_i - a - bx_i - cx_i^2)(-1)$

$\Rightarrow \sum(y_i - a - bx_i - cx_i^2) = 0$

$\Rightarrow \sum y_i = an + b\sum x_i + c\sum x_i^2$ ...(2)

$\frac{\partial E}{\partial b} = 0 = 2\sum(y_i - a - bx_i - cx_i^2)(-x_i)$

$\Rightarrow \sum(y_i - a - bx_i - cx_i)(x_i) = 0$

$\Rightarrow \sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3$ ...(3)

and

$\frac{\partial E}{\partial c} = 0 = 2\sum(y_i - a - bx_i - cx_i^2)(-x_i^2)$

$\Rightarrow \sum(y_i - a - bx_i - cx_i^2)(x_i^2) = 0$

$\Rightarrow \sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4$ ...(4)

where, n-be the number of observations. By dropping suffix i, we get

$\sum y = na + b\sum x + c\sum x^2$ ...(5)

$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$ ...(6)

And $\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$ ...(7)

These equations are called **normal equations for parabola curve.**

As like methods of finding the constants m and C in case of straight line (see section 5.8.5.1), for fitting of parabolic curve we may use

(i) Direct Method and

(ii) Assumed Mean Method or Deviation method,

For finding the values of arbitrary constants a, b, c

Ex. 6.4.1 : By the method of least squares, find the best values of a, b, c in the parabola $y = a + bx + cx^2$ to fit the following data :

x	0	1	2	3	4
y	1	1.8	1.3	2.5	6.3

Soln. :

Step I : The given law is

$y = a + bx + cx^2$ ...(1)

(a) By Direct Method

Since, the normal equations are

$\sum y = na + b\sum x + c\sum x^2$ ...(2)

$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$ ...(3)

and $\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$ ...(4)

Step II : Construct table for $\sum x, \sum y, \sum x^2, \sum xy, \sum x^3, \sum x^4, \sum x^2 y$

Here, $n = 5$

x	y	xy	x^2	x^2y	x^3	x^4
0	1	0	0	0	0	0
1	1.8	1.8	1	1.8	1	1
2	1.3	2.6	4	5.2	8	16
3	2.5	7.5	9	22.5	27	81
4	6.3	25.2	16	100.8	64	256
$\Sigma x = 10$	$\Sigma x = 12.9$	$\Sigma xy = 37.1$	$\Sigma x^2 = 30$	$\Sigma x^2y = 130.3$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$

Step III : By using normal equations

From Equation (2),

$12.9 = 5a + 10b + 30c$ ...(5)

From Equation (3),

$37.1 = 10a + 30b + 100c$ ...(6)

and from Equation (4),

$130.3 = 30a + 100b + 354c$

Step IV : Now eliminate a

Therefore, 2 × Equation (5) – Equation (6)

$25.8 = 10a + 20b + 60c$

$37.1 = 10a + 30b + 100c$

$-\quad -\quad -\quad -$

$-11.3 = -10b - 40c$

$11.3 = 10b + 40c$ ...(7)

and Equation (6) – 6 × Equation (5)

$130.3 = 30a + 100b + 354c$

$77.4 = 30a + 60b + 180c$

$-\quad -\quad -\quad -$

$52.9 = 40b + 174c$ ...(8)

Now, eliminate b from Equation (7) and (8)

∴ Equation (8) – 4 × Equation (7)

$-52.9 = 40b + 174c$

$45.2 = 40b + 160c$

$7.7 = 14c$

$\therefore\ c = 0.55$

Put this value in Equation (7), we get

$11.3 = 10b + 40(0.55)$

$-10.7 = 10b$

$\Rightarrow\ b = -1.07$

Now putting the values of b and c in Equation (5), we get

$12.9 = 5a + 10(-1.07) + 30(0.55)$

$12.9 = 5a - 10.7 + 16.5$

$\Rightarrow\ a = 1.42$

Step V :

Hence, the parabolic curve of best fit is

$y = a + bx + cx^2$

$= (1.42) + (-1.07)x + (0.55)x^2$

(b) By Deviation Method (or Assumed Mean Method)

Step VI

Here, we have to assume means of x and y

Let, a = 2 and b = 1.3 be the assumed means of x and y respectively.

Let $X = x - a = x - 2$ and $Y = y - b = y - 1.3$

The parabolic curve is $y = a + bx + cx^2$

The normal equations becomes

$\sum Y = na + b\sum X + c\sum X^2$ ...(9)

$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$ ...(10)

and $\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$ ...(11)

Step VII

Construct table for $\sum X, \sum X^2$ etc

x	y	X = x–2	Y = y–1.3	X^2	X^3	X^4	XY	X^2Y
0	1	–2	–0.3	4	–8	16	0.6	–1.2
1	1.8	–1	0.5	1	–1	1	–0.5	0.5
2	1.3	0	0	0	0	0	0	0
3	2.5	1	1.2	1	1	1	1.2	1.2
4	6.3	2	5	4	8	16	10	20
		$\Sigma X = 0$	$\Sigma Y = 6.4$	$\Sigma X^2 = 10$	$\Sigma X^3 = 0$	$\Sigma X^4 = 34$	$\Sigma XY = 11.3$	$\Sigma X^2Y = 20.5$

Step VIII :

Putting the above table values in normal Equations (9), (10), (11)

From Equation (9),

$6.4 = 5a + 0 + 10c$

From Equation (10),

$11.3 = 0 + 10b + 0c$

and from Equation (11),

$20.5 = 10a + 0 + 34c$

Hence, $6.4 = 5a + 10c$ ...(12)

$11.3 = 10b$ ...(13)

and $20.5 = 10a + 34c$ ...(14)

From Equation (13)

$b = 1.13$

Equation (14) – 2 × Equation (12)

$\therefore\ 20.5 = 10a + 34c$

$12.8 = 10a + 20c$

$7.7 = 14c$

$\Rightarrow\ \boxed{c = 0.55}$

and from Equation (12)

$6.4 = 5a + 10(0.55) \Rightarrow$ $\boxed{a = 0.18}$

Step IX :

Hence, the parabolic curve of best fit is,

$Y = a + bX + cX^2$

$y - 1.3 = (0.18) + (1.13)(x-2) + (0.55)(x-2)^2$

$y - 1.3 = (0.18) + (1.13)x - 2.26 + (0.55)(x^2 - 4x + 4)$

$= (0.18) + (1.13)x - 2.26 + (0.55)x^2 - 2.2x + 2.2$

$\therefore\ y = (1.42) + (-1.07)x + (0.55)x^2$

Note :

1. By observation, Deviation Method is more convenient because, in this method, we get value of b directly, and three normal equations are more simple. **(See above Example 6.3.1(b))**
2. Observe the last simplified equations by both methods. The coefficients of x, x^2 and constant terms are same of that particular example.
3. If the central value of x is 0, then use Direct Method, because here we get $\sum x, \sum x^3$ is 0 (zero) **(See Example 6.4.2)**
4. Sometimes, for simple calculations, we can shift the origin to the mid-point of x only i.e. use derivation in x only and then use derivation method. Here also, we get $\sum x, \sum x^3$ values as zero. **(See Example 6.4.3)**

Ex 6.4.2 : By the method of least squares, find the best values of a, b, c in the law $y = a + bx + cx^2$ to fit the following data :

x	–2	–1	0	1	2
y	–3.1	–1.3	0.6	2.8	5.3

Soln. :

Step I : The given parabolic curve is,

$y = a + bx + cx^2$ ...(1)

Here, the central value of x is zero, so we can use **Direct Method,**

The normal equations are,

$\sum y = na + b\sum x + c\sum x^2$ ...(2)

$\sum xy = a\sum x + b\sum x + c\sum x^3$ ...(3)

$\sum x^2y = a\sum x^2 + b\sum x^3 + c\sum x^4$ ...(4)

Here, $n = 5$

Step II : Construct table for $\sum x, \sum y, \sum xy, \sum x^2, \sum x^3$ and $\sum x^4, \sum x^2y$

x	y	x^2	x^3	x^4	xy	x^2y
–2	–3.1	4	–8	16	6.2	–12.4
–1	–1.3	1	–1	1	1.3	–1.3
0	0.6	0	0	0	0	0
1	2.8	1	1	1.	2.8	2.8
2	5.3	4	8	16	10.6	21.2
$\Sigma x = 0$	$\Sigma y = 4.3$	$\Sigma x^2 = 10$	$\Sigma x^3 = 0$	$\Sigma x^4 = 34$	$\Sigma xy = 20.9$	$\Sigma x^2y = 10.3$

Step III : From Equation (2), (3), (4)

$4.3 = 5a + 0 + 10c$ ...(5)

$20.9 = 0 + 10b + 0$ ...(6)

and $10.3 = 10a + 0 + 34c$ ...(7)

From Equation (6), $b = 2.09$

Equation (7) – 2 × Equation (5), we get

$10.3 = 10a + 34c$

$(-)(8.6 = 10a + 20c)$

$1.7 = 0 + 14c \Rightarrow c = 0.12$

and from Equation (5),

$4.3 = 5a + 10(0.12) \Rightarrow a = 0.62$

Step IV : Hence, the required parabolic curve of best fit is

$y = a + bx + cx^2$

$y = (0.62) + (2.09)x + (0.12)x^2$ ✓ ...Ans.

Ex. 6.4.3 : Fit a second degree parabola to the following data

x	1.0	1.5	2	2.5	3.0	3.5	4.0
y	1.1	1.3	1.6	2.0	2.7	3.4	4.1

Soln. :

Step I : The parabolic equation is,

$y = a + bx + cx^2$ ...(1)

The value of x evenly spaced, so use **Assumed Mean Method,**

Let $a = 2.5$,

$\therefore\ X = x - a = x - 2.5$ be the deviation in x and $y = Y$.

∴ The normal equations are,

$\sum Y = na + b\sum X + c\sum X^2$ ...(2)

$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$ ...(3)

$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$ ...(4)

where, $n = 7$

Step II : Construct table for $\sum X, \sum Y$ etc.

x	Y = y	X = x – 2.5	X^2	X^3	X^4	XY	X^2Y
1.0	1.1	–1.5	2.25	–3.375	5.063	–1.65	2.475
1.5	1.3	–1	1	–1	1	–1.3	1.3
2.0	1.6	–0.5	0.25	–0.125	0.063	–0.8	0.4
2.5	2.0	0	0	0	0	0	0
3.0	2.7	0.5	0.25	0.125	0.063	1.35	0.675

x	Y=y	X=x−2.5	X^2	X^3	X^4	XY	X^2Y
3.5	3.4	1	1	1	1	3.4	3.4
4.0	4.1	1.5	2.25	3.375	5.063	6.15	9.225
	$\Sigma Y=16.2$	$\Sigma X=0$	$\Sigma X^2=6.75$	$\Sigma X^3=0$	$\Sigma X^4=12.252$	$\Sigma XY=7.15$	$\Sigma X^2Y=17.475$

Step III : From Normal equations :

$$16.2 = 7a + 6.75c \quad ...(5)$$
$$7.15 = 6.75b \quad ...(6)$$
$$17.475 = 6.75a + 12.252c \quad ...(7)$$

From Equation (6), $b = 1.059$

To eliminate a from Equation (5) and (7)

(6.75) × Equation (5) − 7 × Equation (7)

$$109.35 = 47.25a + 45.562c$$
$$-\ 122.33 = 47.25a + 82.701c$$
$$-12.98 = 0 - 37.139c$$
$$\Rightarrow c = 0.35$$

and from Equation (5),

$$7a = (16.2) - (6.75)(c) = (16.2) - (6.75)(0.35)$$
$$\Rightarrow a = 1.98$$

Step IV : Hence, the required parabolic curve of best fit is

$$y = a + bx + cx^2$$
$$y = (1.98) + (1.059)x + (0.35)x^2 \checkmark \quad ...\text{Ans.}$$

Ex 6.4.4 : Fit a non-linear trend of the form $y = a + bx + cx^2$ to the following data :

x	0	1	2	3	4
y	1.0	1.5	1.5	2.5	3.5

Soln. :

Step I : The parabolic curve is,

$$y = a + bx + cx^2 \quad ...(1)$$

Here, we can use Deviation Method

Let, $a = 2$ and $b = 1.5$

$\therefore X = x - a = x - 2$ and

$Y = y - b = 1.5$

$\therefore$ The corresponding normal equations are,

$$\left.\begin{aligned} \Sigma Y &= an + b\Sigma X + c\Sigma X^2 \\ \Sigma XY &= a\Sigma X + b\Sigma X^2 + c\Sigma X^3 \\ \Sigma X^2Y &= a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 \end{aligned}\right\} \quad ...(2)$$

where, $n = 5$

Step II : Construct table for ΣX, ΣY etc.

x	y	X=x−2	Y=y−1.5	X^2	X^3	X^4	XY	X^2Y
0	1.0	−2	−0.5	4	−8	16	10	−2
1	1.5	−1	0	1	−1	1	0	0
2	1.5	0	0	0	0	0	0	0
3	2.5	1	1	1	1	1	1	1
4	3.5	2	2	4	8	16	4	8
		$\Sigma X=0$	$\Sigma Y=2.5$	$\Sigma X^2=10$	$\Sigma X^3=0$	$\Sigma X^4=34$	$\Sigma XY=15$	$\Sigma X^2Y=7$

Step III : From normal Equation (2), we get

$$2.5 = 5a + 10c \quad ...(3)$$
$$15 = 10b \quad ...(4)$$
$$7 = 10a + 34c \quad ...(5)$$

From Equation (4), $b = 1.5$

Equation (5) − 2 × Equation (3)

$$7 = 10a + 34C$$
$$-\ 5 = 10a + 20C$$
$$2 = 14C \quad \Rightarrow \quad c = 0.14$$

From Equation (3),

$$2.5 = 5a + 10(0.14)$$
$$\Rightarrow a = 0.22$$

Step IV : Hence, the required parabolic curve or best fit is

$$Y = a + bX + cX^2$$
$$y - 1.5 = a + b(x-2) + c(x-2)^2$$
$$y - 1.5 = a + bx - 2b + c(x^2 - 4x + 4)$$
$$y - 1.5 = a + bx - 2b + cx^2 - 4cx + 4c$$
$$y = (0.22) + (1.5)x - 2(1.5) + (0.14)x^2 - 4(0.14)x + 4(0.14) + 1.5$$
$$y = (-0.72) + (0.94)x + (0.14)x^2 \checkmark \quad ...\text{Ans.}$$

Ex. 6.4.5 : The velocity v of a liquid is known to vary with temperature according to law $v = a + bT + CT^2$. Find the best values of a, b, c for the following table.

T :	1	2	3	4	5	6	7
V :	2.31	2.01	3.80	1.66	1.55	1.47	1.41

Soln. :

Step I :

Given, $v = a + bT + CT^2 \quad ...(1)$

The normal equations of parabolic curve are

$$\Sigma v = na + b\Sigma T + c\Sigma T^2 \quad ...(2)$$
$$\Sigma Tv = a\Sigma T + b\Sigma T^2 + c\Sigma T^3 \quad ...(3)$$
$$\Sigma T^2v = a\Sigma T^2 + b\Sigma T^3 + c\Sigma T^4 \quad ...(4)$$

Step II :

Construct table for $\Sigma v, \Sigma T, \Sigma T^2, \Sigma T^3$ ...etc.

T	V	T^2	T^3	T^4	TV	T^2V
1	2.31	1	1	1	2.31	2.31
2	2.01	4	8	16	4.02	8.04
3	3.80	9	27	81	11.4	34.2
4	1.66	16	64	256	6.64	26.56

T	V	T^2	T^3	T^4	TV	T^2V
5	1.55	25	125	625	7.75	38.75
6	1.47	36	216	1296	8.82	52.92
7	1.41	49	343	2401	9.87	69.09
$\Sigma T = 28$	$\Sigma v = 14.21$	$\Sigma T^2 = 140$	$\Sigma T^3 = 584$	$\Sigma T^4 = 4676$	$\Sigma Tv = 50.81$	$\Sigma T^2v = 231.87$

Here $n = 7$

Step III :

Normal equations (2), (3) and (4) becomes

$$14.21 = 7a + 28b + 140c$$
$$50.81 = 28a + 140b + 584c$$
$$231.87 = 140a + 584b + 4676c$$

By solving the above three equations simultaneously we get.

$$a = 3.313;\ b = -0.1965;\ c = -0.0254$$

Step IV :

The best fit curve of parabola is

$$V = a + bT + cT^2$$
$$\therefore V = (3.313) + (-0.1935)T + (-0.0254)t^2 \checkmark \quad ...\text{Ans.}$$

Exercise

Ex. 1 : Fit a second degree parabolic curve to the following data and estimate production in 1982.

Year (x)	1974	1975	1976	1977	1978	1979	1980
Production (y)	12	14	26	42	40	50	52

Ex. 2 : Fit a second degree parabolic curve to the following data

x	1	2	3	4	5	6	7	8	9
y	2	6	7	8	10	11	11	10	9

Ex. 3 : Fit a parabola $y = a + bx + cx^2$ to the following data :

x	2	4	6	8	10
y	3.07	12.85	31.47	57.38	91.29

Ex. 4 : The velocity V of a liquid is known to vary with temperature according to quadratic law $V = a + bT + CT^2$. Find the best values of a, b and c for the following data.

T	1	2	3	4	5	6	7
V	2.31	2.01	2.80	1.66	1.55	1.47	1.41

Ex. 5 : Fit a parabolic curve to the following data and estimate the value of y for x = 6

x	1	2	3	4	5
y	25	28	33	39	46

Ex. 6 : Fit a second degree curve to the following data and estimate the value of y when x = 80.

x	10	20	30	40	50	60	70
y	20	60	70	80	90	100	100

Ex. 7 : Fit a second degree curve to the following data :

Year	1965	1966	1967	1968	1969	1970	1971
Profit (in crores)	125	140	165	195	200	215	220

Also estimate the profit in 1972.

Syllabus Topic : Fitting More General Curves by the Method of Least Squares

6.5 Fitting of Other Curves (Or Fitting Exponential Curve)

Now, we have to find the procedure to best fit of exponential curves. The exponential curves like

(i) $y = a x^b$ (ii) $y = a e^{bx}$ (iii) $y = ab^x$ (iv) $xy^n = b$

have to best fit, by taking logarithm of both sides.

(As discussed in section 5.8.2) we get linear equation of the form $y = mx + c$. We have to find the values of arbitrary constants m and c by using normal equation of straight line (by using principle of least squares)

$$\Sigma y = n.c + m\Sigma x$$

and $\Sigma xy = c\Sigma x + m\Sigma x^2$

use any method (direct, deviation or direct mean).

Ex. 6.5.1 : Fit a curve $y = ab^x$ to the following data, using the method of least squares.

x	2	3	4	5	6
y	144	172.8	207.4	248.8	298.5

Soln. :

Step I : Given curve is

$$y = a b^x \quad ...(1)$$

Taking log on both sides

$$\log_{10} y = \log(ab^x) = \log_{10} a + x \log_{10} b$$

Let, $\log_{10} y = Y$; $\log_{10} a = C$ and $\log_{10} b = m$

We get

$$Y = C + mx, \text{ which is straight line.} \quad ...(2)$$

By the principle of least square, the normal equations of straight line are

$$\Sigma Y = nc + m\Sigma x \quad ...(3)$$
$$\Sigma XY = c\Sigma X + m\Sigma x^2 \quad ...(4)$$

Here n = (Here, we use Direct Method)

Step II : Construct table for $\sum X, \sum Y, \sum X^2, \sum XY$.

Sr. No.	y	X = x	$\log_{10} y = Y$	X^2	XY
1	144.0	2	2.1584	4	4.3168
2	172.8	3	2.2375	9	6.7125
3	207.4	4	2.3168	16	9.2672
4	248.8	5	2.3959	25	11.9745
5	298.5	6	2.4749	36	14.8494
		$\sum X = 20$	$\sum Y = 11.5835$	$\sum X^2 = 90$	$\sum XY = 47.1254$

Step III : By using normal equations

$$11.5835 = 5c + 20m \quad ...(5)$$
$$47.1254 = 20c + 90m \quad ...(6)$$

Equation (4) – 4 × Equation (3)

$$47.1254 = 20c + 90m$$
$$(-)\ 46.334 = 20c + 80m$$
$$0.7914 = 10m \Rightarrow \quad m = 0.07914$$

From Equation (3),

$$5c = (11.5835) - 20 \times (0.07914)$$
$$c = 2.00014$$

Step IV : Since, $e = \log_{10} a$

$$a = \text{antilog}(2.00014) = (100.03)$$
$$\text{and } m = \log_{10} b$$
$$\therefore\ b = \text{antilog}(m) = \text{antilog}(0.07914)$$
$$= 1.19 \approx 1.2$$

Hence, the required law for the best fit is

$$y = ab^x = (100.03)(1.2)^x \checkmark \quad ...\text{Ans.}$$

Ex. 6.5.2 : Find the least squares fit of the form $V = at^b$. Find the best possible values of a and b.

V	350	400	500	600
t	61	26	7	26

Soln. :

Step I : The given law is

$$V = at^b \quad ...(1)$$

Taking log on both sides,

$$\log_{10} V = \log_{10} a + b \log_{10} t$$

put $\log_{10} V = Y$; $\log_{10} a = A$; $\log_{10} t = X$; we get

$$Y = A + bX \quad ...(2)$$

which is linear equation

By least square method, the normal equations

$$\sum Y = nA + b\sum X \quad ...(3)$$
$$\sum XY = A\sum X + b\sum X^2 \quad ...(4)$$

Here $n = 4$

Step II : Construct the table for $\sum X, \sum Y$ etc.

V	t	$X = \log_{10} t$	$Y = \log_{10} V$	XY	X^2
350	61	1.7853	2.5441	4.542	3.187
400	26	1.4150	2.6021	3.682	2.002
500	7	0.8451	2.6990	2.281	0.714
600	2.6	0.4150	2.7782	1.153	0.172
		$\sum X = 4.4604$	$\sum Y = 10.6234$	$\sum XY = 11.658$	$\sum X^2 = 6.075$

Step III : From Equation (3) and (4),

$$10.623 = 4A + 4.46b \quad ...(5)$$
$$\text{and } 11.658 = 4.46A + 6.075b \quad ...(6)$$

(4.46) × Equation (5) – 4 × Equation (6)

$$47.38 = 17.84A + 19.90b$$
$$(-)\ 46.64 = 17.84A + 24.3b$$
$$0.74 = -4.4b$$
$$\Rightarrow b = -0.168$$

and from Equation (5)

$$10.623 = 4A + (4.46)(-0.168)$$
$$\Rightarrow A = 2.843$$

Hence, $A = \log_{10} a \Rightarrow a = \text{antilog}(2.843)$

$$\therefore a = 699.8$$

Step IV : Hence, the required curve of best fit is,

$$V = (699.8)\ t^{-0.168} \checkmark \quad ...\text{Ans.}$$

Ex. 6.5.3 : The pressure and volume of a gas are related by the equation $PV^a = k$, where a and k are constants. Fit the equation to the following set of observations.

P(kg/cm²)	0.5	1.0	1.5	2.0	2.5	3.0
V (liters)	1.62	1.00	0.75	0.62	0.52	0.46

Soln. :

Step I : The given curve is

$$PV^a = k \quad ...(1)$$

Taking log on both the sides, we have

$$\log_{10} P + a \log_{10} V = \log_{10} k$$
$$\log_{10} V = \frac{1}{a}\log_{10} k - \frac{1}{a}\log_{10} P$$

Let, $\log_{10} V = Y$; $\frac{1}{a}\log_{10} k = C$

and $-\frac{1}{a} = m$; $\log_{10} P = X$

we get, $Y = C + mX \quad ...(2)$

By least square principle, the normal equations are

$$\sum Y = nC + m\sum X \quad ...(3)$$
$$\text{and } \sum XY = n\sum X + m\sum X^2 \quad ...(4)$$

Here, $n = 6$

Step II : Construct table for $\sum X, \sum X^2$ etc.

P	V	$X = \log_{10} V$	$Y = \log_{10} V$	XY	X^2
0.5	1.62	–0.30	0.21	–0.063	0.090
1.0	1.00	0.00	0.00	0.000	0.00
1.5	0.75	0.18	–0.13	–0.023	0.032
2.0	0.62	0.30	–0.21	–0.063	0.090
2.5	0.52	0.40	–0.29	–0.116	0.160
3.0	0.46	0.48	–0.34	–0.163	0.230
		$\sum X = 1.06$	$\sum Y = -0.76$	$\sum XY = -0.428$	$\sum X^2 = 0.602$

Step III : The normal equations becomes

$$-0.76 = 6C + (1.06)m \quad ...(5)$$
$$\text{and } -0.428 = (1.06)C + (0.602)m \quad ...(6)$$

(1.06) × Equation (5) – 6 × Equation (6)

$$-0.8056 = (6.36)C + (1.1236)m$$
$$(-)$$
$$-2.568 = (6.36)C + (3.612)m$$
$$1.7624 = -(2.4884)m$$
$$\Rightarrow m = -0.7282$$

Putting the value of m in Equation (5),

$$-0.76 = 6C + (1.06) \times (-0.7282)$$
$$C = 0.00198$$
$$\therefore C \approx 0.01$$

Step IV : Since, $m = -\frac{1}{a} \Rightarrow a = -\frac{1}{m}$

$$\therefore a = -\left(\frac{1}{-0.7282}\right) = 1.3732$$

$$\text{and } c = \frac{1}{a}\log_{10} k$$
$$\log_{10} k = Ca = (0.01)(1.3732) = 0.0137$$
$$\therefore k = \text{antilog}(0.0137) = 1.032$$

Step V : The required curve of best fit is

$$PV^a = k \Rightarrow PV^{1.3732} = 1.032 \checkmark \quad ...\text{Ans.}$$

Exercise

Ex. 1 : Find the law of the form $y = ab^x$ to the following data :

x	1	2	3	4	5	6	7
y	1	1.2	1.8	3.6	4.7	6.6	9.1

Ex. 2 : Fit a curve of the form $y = ax^b$ to the following data :

x	1	2	3	4	5	6
y	120	90	60	20	11	5

Ex. 3 : Fit a curve of the form $y = ax^b$ to the following data by the method of least squares.

x	1	2	3	4
y	76	210	390	605

Ex. 4 : Fit a curve of the form $y = ab^x$ to the following data, using least squares method

x	2	3	4	5	6
y	34.38	79.08	181.90	418.36	962.23

Ex. 5 : The voltage V across a capacitor at time t seconds is given by the following data :

t	0	2	4	6	8
v	150	63	28	12	5.6

Use the method of least squares, to fit a curve of the form $V = a\,e^{bt}$.

Ex. 6 : Find the values of a and b, by least squares method of the curve $y = ax + \frac{b}{x}$ for the following data

x	1	2	3	4
y	–1.5	0.99	3.88	7.66

Ex. 7 : Predict the mean radiation close at an altitude by fitting an exponential curve to the given data :

x	50	450	780	1200	4400
y	28	30	32	36	51

6.6 Testing of Hypothesis

Hypothesis :

- Some assumption or guess or statement about the population or some parameters of population on the basis of sample information. Such assumption or guess which is true or false, is called **Hypothesis.**
- To check, upto what degree of confidence, the known information about the characteristics of the population is accepted or rejected, by using random sample drawn from the population is called **testing of hypothesis.**
- This testing procedure always involves two types of hypothesis. (i) Null hypothesis (ii) Alternative hypothesis. A procedure for deciding whether to accept (or reject) null hypothesis and to reject (or accept) alternative hypothesis is called the test of hypothesis.
- **Level of Significance (LOS) :** As we know, when the hypothesis is wrong, then we reject it. This is called Type I error. If the hypothesis is true, but if we reject it, then it is called Type II error.
- For testing a given hypothesis, we have to make some rules and these rules to be good, they must be designed so as to minimize errors of decision, that limit is called **level of significance** of test. In practice, a level of significance of 0.06 or 0.01 (or 6% or 1%) is made. For example a 0.05 or 5% level of significance is chosen in designing a test of a hypothesis, that means, there are 5 chances in 100 that we would reject the hypothesis, and it should be accepted i.e. whenever the null hypothesis

is true, we 95% confident about that, we would make a right decision. In such cases, we say that, hypothesis is rejected in 5% level of significant i.e. we may wrong with probability 0.05.

[In normal distribution, suppose we decide to reject the hypothesis, if the mean μ is so small or too large. $Z = \frac{x-\mu}{\sigma}$ is a normal statistics with $\mu = 0$ and $\sigma = 1$, the extreme values of Z would lead to the rejection of the hypothesis. So for Z = X with $\mu = 0$ and $\sigma = 1$, the 95% level of confident will lie between – 1.96 and 1.96 (Since under this range is 0.95).

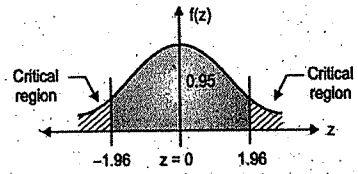


Fig. 6.6.1

The total shaded area is 0.05 which is the level of significance of the test. Reject the hypothesis at 5% level of significance if the Z value lies out the range – 1.96 to 1.96 (i.e. either Z > 1.96 or Z < – 1.96)]

Note :

(i) 5% or 0.05 level of significance : Z > 1.96 and Z < – 1.96

(ii) 95% or 0.95 level of confidence : – 1.96 < Z < 1.96

(iii) Sum of level of significance and level of confidence is 100%.

(iv) $Z_\alpha = 1.96$

Syllabus Topic : Tests of Significance

6.7 Tests of Significance

- Let θ – be the parameter of the population. θ_0 – be some specific constant of parameter (also, θ_0 – is called sample statistic). There will be some difference between θ and θ_0. Since, θ_0 – is coming from sample observations and which is different for different parameters (samples). The difference due to variation in sampling is called **insignificant differences**. If the sample is not drawn from given population, such difference is called **significant difference**.
- Hence the procedure of testing whether the difference between θ and θ_0 is significant or not is called the **test of significance**.
- (Simply, the methods of inference used to accept or reject the assumption based on sample data.)

Syllabus Topic : Large Sample Test for Single Proportion, Difference of Proportion, Single Mean, Difference of Means and Difference of Standard Deviations

6.7.1 Tests of Significance For Large Samples

Generally, if the sample size is greater than 30, such samples are considered as large sample. Tests of significance used for both small and large samples.Test of significance for large samples.are applicable only if.

(i) Sample distribution of statistic is normal, irrespective of population distribution.

(ii) Sample statistic is close to the corresponding population parameters which can be used to calculate the standard error of the sampling distribution.

☞ **There are different test of significance for large samples.**

Test I : Large sample test for single proportion

Let, H_0 – be the null hypothesis

H_1 – be the alternative hypothesis

X – be the number of success from n independent trials.

(or number of trials within the sample having some particular characteristic)

P – Proportion of population (also, it represent the proportion of item or individual possessing a certain characteristics. Generally called probability of success.)

When sample size is large, (i.e. n is large),

X – follows normal probability distribution $N(nP, \sqrt{nPQ})$ (i.e. normal distribution with mean (np) and standard deviation $\sqrt{nPQ}$.)

[Here, nP – mean (μ);

nPQ – variable (σ^2);

$\sqrt{nPQ}$ – Standard deviation (S. D.) (σ);

Q = 1 – P.]

Let, $p = \frac{x}{n}$ – be the sample proportion.

(i.e. Proportion of successes in 'n' trials)

Therefore, sample proportion p, follows normal distribution $N\left(\frac{nP}{n}, \sqrt{\frac{nPQ}{n^2}}\right)$

i.e p follows $N\left(P, \sqrt{\frac{PQ}{n}}\right)$

Let, $Z = \frac{X-\mu}{\sigma}$ be the normal statistic

$$= \frac{X-np}{\sqrt{nPQ}} = \frac{n\left(\frac{X}{n}-P\right)}{\sqrt{n}\sqrt{PQ}} = \frac{\sqrt{n}\,(p-P)}{\sqrt{PQ}}$$

$$= \frac{p-P}{\left(\frac{\sqrt{PQ}}{\sqrt{n}}\right)} = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

$\therefore Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$ be the normal statistic for test.

(i) If $|Z| \le Z_\alpha$ then difference between the sample proportion p and population proportion P is not significant at 5% level of significance.

(ii) If $|Z| > Z_\alpha$; p – P is significant at 5% level of significance.

(iii) for normal distribution, 95% confidence level for P is

$\frac{|p-P|}{\sqrt{\frac{PQ}{n}}} \le 1.96$ i.e. $p - P \le (1.96)\sqrt{\frac{PQ}{n}}$

i.e. $p - (1.96)\sqrt{\frac{PQ}{n}}, p + (1.96)\sqrt{\frac{PQ}{n}}$

Test 2 : Large sample test for difference of proportions.

Let P – be the proportion of pollution

p_1, p_2 – be the sample proportions of large size n_1 and n_2 respectively.

Consider the null hypothesis (H_0) that the difference between the populations P_1 and P_2 is zero.

i.e. $P_1 - P_2 = 0 \Rightarrow P_1 = P_2$

i.e. the samples are drawn from the same population.

Let, σ_1, σ_2 are the standard deviations of sample s_1 and s_2.

$\therefore \sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$

$= \sqrt{n_1P_1Q_1 + n_2P_2Q_2}$

$$\therefore Z = \frac{\sqrt{n}\left[(p_1-p_2)-(P_1-P_2)\right]}{\sigma_{n_1-n_2}} = \frac{(p_1-p_2)-0}{\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}}$$

($\because P_1 = P_2$ and hence $Q_1 = Q_2$)

$$Z = \frac{p_1-p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{p_1-p_2}{\sigma_{p_1-p_2}}$$

If P – is not known, then the proportion of combined samples s_1 and s_2 is

$P = \frac{P_1n_1 + P_2n_2}{n_1+n_2}$ is used

$$\left(E\left(\frac{n_1P_1+n_2P_2}{n_1+n_2}\right) = \frac{1}{n_1+n_2}[n_1E(P_1) + n_2E(P_2)] = \frac{1}{n_1+n_2}[n_1P + n_2P] = P \quad \therefore P = \frac{n_1P_1+n_2P_2}{n_1+n_2} \right)$$

Test 3 : Large sample test for single mean

- Let $X_1, X_2, X_3, \ldots X_n$ – be the independent observations from sample of size n drawn from population
- Each X_i follows normal distribution $N(\mu, \sigma)$
- If $X_1, X_2, X_3, \ldots X_n$ are in independent normal variates with mean μ_i and variance σ_i^2 then $\sum C_i X_i$ is also normal variate with mean $\mu = \sum C_i\mu_i$ and variance $\sigma^2 = \sum C_i^2\sigma_i^2$

Let, $C_i = \frac{1}{n}$; $\mu_i = \mu$; $\sigma_i = \sigma$

$\therefore \sum C_i X_i = \sum \frac{1}{n}\mu = \frac{1}{n}\mu(n) = \mu$

And $\sum C_i^2\sigma_i^2 = \sum \frac{1}{n^2}\sigma^2 = \frac{1}{n^2}\sigma^2(n) = \frac{1}{n}\sigma^2$

the sample mean $(\bar{X})$ follows normal distribution, $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Hence the normal static, $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

95% limits of confidence for μ is $\left|\frac{\mu-\bar{X}}{\frac{\sigma}{\sqrt{n}}}\right| \le 1.96$.

$\bar{X} - (1.96)\frac{\sigma}{\sqrt{n}}, \bar{x} + (1.96)\frac{\sigma}{\sqrt{n}}$

Test 4 : Large sample test of significance for difference of means

- Let, $\bar{X}_1$ and $\bar{X}_2$ – be the means of two large sample sizes n_1 and n_2 which are drawn from population with same mean μ and variance σ_1^2 and σ_2^2 respectively.

Here,

μ – is mean of population ;

σ_1^2, σ_2^2 – variance of sample means. $\bar{X}_1$ and $\bar{X}_2$ respectively

s_1^2, s_2^2 – are variance of sample s_1 and s_2.

This means $\bar{X}_1$ follows normal distribution $N\left(\mu, \frac{\sigma_1}{\sqrt{n_1}}\right)$ and $\bar{X}_2$ also follows normal distribution $N\left(\mu, \frac{\sigma_2}{\sqrt{n_2}}\right)$

- Consider the null hypothesis that these is no difference between the population means i.e. $\mu_1 = \mu_2 = \mu$.

$E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu - \mu = 0$

$Var(\bar{x}_1 - \bar{x}_2) = V(\bar{x}_1) + V(\bar{x}_2)$

($\because$ $\bar{x}_1$, $\bar{x}_2$ are independent means of independent samples)

$$\sigma(\bar{x}_1-\bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

($\because$ variance of the sampling distribution of means denoted by $\sigma_{\bar{X}}^2$ is $\sigma_{\bar{X}}^2 = E(\bar{X}-\mu)^2 = \frac{\sigma^2}{n}$ where, σ^2 – is the variance of population.)

1. If the samples are taken from same population whose S.D. are:

   $\sigma_1 = \sigma_2 = \sigma$ and $\mu_1 = \mu_2 = \mu$ then the normal variable

$$Z = \frac{(\bar{x}_1-\bar{x}_2)-0}{\sigma_{(\bar{x}_1-\bar{x}_2)}}$$

$$\therefore Z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

$$= \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}} = \frac{\bar{x}_1-\bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

2. If $\sigma_1 \neq \sigma_2$ then $Z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ and if σ_1 and σ_2 are approximated sample S. D's. s_1 and s_2 then

$$Z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

If $\sigma_1 = \sigma$; $\sigma_2 = \sigma$ but not known and

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2}$$

$$Z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2}\right)\left(\frac{n_1+n_2}{n_1 n_2}\right)}}$$

$$= \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_2}+\frac{s_2^2}{n_1}\right)}}$$

Test 5 : Large sample test for difference of standard deviations.

If

s - be the sample S. D. of size n

σ - be the population S. D. then

S follows a normal distribution $N\left(\sigma, \frac{\sigma}{\sqrt{2n}}\right)$ approximately

$$\therefore Z = \frac{s-\sigma}{\left(\frac{\sigma}{\sqrt{2n}}\right)}$$ be the normal statistic.

– Now, if s_1 and s_2 – be the S. D's of two large samples of size n_1 and n_2 drawn from population σ be the S.D. of population.

S_1 and S_2 followes normal distribution. $N\left(\sigma, \frac{\sigma}{\sqrt{2n_1}}\right)$ and $N\left(\sigma, \frac{\sigma}{\sqrt{2n_2}}\right)$ respectively.

– Therefore, the difference between S_1 and S_2 follows normal distribution i.e. $(S_1 - S_2)$ follows

$$N\left(0, \sigma\sqrt{\frac{1}{2n_1}+\frac{1}{2n_2}}\right)$$

$(\because \sigma_1 = \sigma_2 = \sigma$ for population. $\Rightarrow \sigma_1 - \sigma_2 = 0)$

– Hence, the normal statistic for test is

$$z = \frac{S_1-S_2}{\sigma\sqrt{\frac{1}{2n_1}+\frac{1}{2n_2}}}$$

If σ is unknown, then $\sigma = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1+n_2}}$

Therefore,

$$z = \frac{(S_1-S_2)}{\sqrt{\left(\frac{n_1 S_1^2+n_2 S_1^2}{n_1+n_2}\right)\left(\frac{1}{2n_1}+\frac{1}{2n_1}\right)}}$$

$$= \frac{(S_1-S_2)}{\sqrt{\left(\frac{n_1 S_1^2+n_2 S_2^2}{n_1+n_2}\right)\left(\frac{2(n_1+n_2)}{4n_1 \cdot n_2}\right)}}$$

$$z = \frac{(S_1-S_2)}{\sqrt{\frac{S_1^2}{2n_2}+\frac{S_2^2}{2n_1}}}$$

Ex. 6.7.1 : Survey Says that, in the engineering admissions, only 20 % of the students are a top quality students. In 2018 year admissions of 00 students, only 50 are of top quality students. Show that the sample of this chosen, is not representative sample or the assumption, 20 % top quality student is wrong. Also find the 95 % level of confidence of top quality students.

Soln. :

Step I :

Let,

H_0 = 20 % top quality students;

$H_0 = \frac{20}{100} = \frac{1}{5} = P$(– population proportion)

$H_1 \neq \frac{1}{5}$ (Alternative statement)

$p = \frac{50}{600} = \frac{1}{12}$ (→ be the proportion of top quality students in the sample)

$Q = 1 - P = 1 - \frac{1}{5} = \frac{4}{5}$; $n = 600$

Step II :

Now, the normal statistic z for proportion is,

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{\left(\frac{1}{12}-\frac{1}{5}\right)}{\sqrt{\frac{1}{5}\times\frac{4}{5}\times\frac{1}{600}}}$$

$$= \frac{-\frac{7}{60}}{0.01632} = -\frac{7}{0.9792}$$

$$\therefore z = -7.1487$$

Step III : $|z| = 7.1487 > 1.96$

($\because$ Here, the boundary point of 95 % level of confidence is 1.96)

– This shows that, the null hypothesis is rejected.

i.e. the difference between P and p is significant at 5 % level of significance.

Here, H_0 = 20 % top quality student is wrong

i.e. the selected sample of a particular year is not representative sample.

Step IV : 95 % level of confidence for sample proportion p is

$$\frac{|P-p|}{\sqrt{\frac{PQ}{n}}} \leq 1.96$$

$$P - (1.96)\sqrt{\frac{PQ}{n}} \leq P \leq P + (1.96)\sqrt{\frac{PQ}{n}}$$

i.e. $\frac{1}{12}\left[(1.96)\times\sqrt{\frac{1}{12}\times\frac{11}{12}\times\frac{1}{600}}\right] \leq P \leq \frac{1}{12}$

$$+\left(\sqrt{\frac{1}{12}\times\frac{11}{12}\times\frac{1}{600}}\right)\times 1.96$$

$$\frac{1}{12} - [(1.96)\times(0.01123)] \leq P \leq \frac{1}{12} + (1.96)\times(0.01123)$$

$$\Rightarrow (0.08333) - (0.022) \leq P \leq (0.0833) + (0.022)$$

$$0.0613 \leq P \leq 0.1053$$

This shows that 95 % confidence limits for the percentage of top quality student are 6.13 and 10.53

Ex. 6.7.2 : The fatality rate of typhoid patients is believed to be 17.26 %. In a certain years, 640 patients suffering from typhsical were treated in a metropolitern hospital and only 63 patients died. Can we say, hospital is efficient ? Level of significente be 1 % is – 2.33.

Soln. :

Step I : Let H_0 = P – Hospital is not efficient.

$H_1 : \neq P$

Given, $Z_\alpha = -2.33$

$p = \frac{63}{640} = 0.0984$; $n = 640$

$P = 17.26\ \% = \frac{17.26}{100} = 0.1726$

$\therefore Q = 1 - P = 1 - 0.1726 = 0.8274$

Step II : Now, the normal statistic Z for proportion is,

$$Z = \frac{p-P}{\sqrt{\frac{QP}{n}}} = \frac{0.0984-0.1726}{\sqrt{\frac{(0.1726)\times(0.8274)}{640}}}$$

$$= \frac{-0.0742}{\sqrt{0.000223}} = \frac{-0.0742}{0.014933}$$

$$Z = -4.96886$$

Step III : It is observed that

$|Z| > |Z_\alpha|$

Hence, hypothesis H_0 is rejected and alternative hypothesis H_1 is accepted.

i.e. Hospital is efficient.

Ex. 6.7.3 : In a college admission cell, cleark claims that at most 60 % of the students entering the college, leave without taking admission. A random sample of 50 students, 35 of them left without taking admission. Is the given sample size is good with the claim of the cleark. Use 5% level of significance = 1.645.

Soln. :

Step I : Let, H_0 = Given sample size is consistent.

$H_0 = P = 60\ \% = \frac{60}{100} = 0.6$;

$Q = 1 - P = 1 - 0.6 = 0.4$;

$p = \frac{35}{50} = 0.7$; $H_1 : \neq P$

Given, $n = 50$, $Z_\alpha = 1.645$

Step II : Normal statistics

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.7-0.6}{\sqrt{\frac{(0.6)(0.4)}{50}}}$$

$$Z = \frac{0.1}{\sqrt{0.0048}} = 1.44337$$

$\therefore |Z| < |Z_\alpha|$

Step III :

$\therefore |Z| < |Z_\alpha|$

i.e. H_0 is accepted,

i.e. The sample size is consistent.

Ex. 6.7.4 : A cubical die is thrown 9000 times. The numbers 3 or 4 appears 3240 times show that the die shows 3 or 4 numbers are not as an unbiased one. Use 5 % level of significance = 1.96.

Soln. :

Step I : Let, H_0 – die is unbaised which shows numbers 3 or 4

$$\therefore \; H_0 = P = \frac{2}{6} = \frac{1}{3}$$

$$Q = 1 - P = 1 - \frac{1}{3} = \frac{2}{3}$$

$$H_1 : \neq P$$

$$Z_\alpha = 1.96$$

Step II : As n = 9000, is very large, it follows normal, normal distribution.

$$N\left(nP, \sqrt{nPQ}\right)$$

$\therefore$ Normal statistics is

$$Z = \frac{X - \mu}{\sigma} = \frac{X - nP}{\sqrt{nPQ}}$$

Step III : Here, $X = 3240$

$$\therefore \; Z = \frac{3240 - \left(9000 \times \frac{1}{3}\right)}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}}$$

$$Z = \frac{240}{\sqrt{2000}} = \frac{240}{44.7214} = 5.3666$$

As $|Z| > |Z_\alpha|$

i.e. H_0 is rejected.

i.e. H_1 is accepted.

i.e. The dice is not unbiased.

Ex. 6.7.5 : In a particular city, there are 20 % of a random sample of 800 school boys had a physical problem, where as in Kolkatta, there are 18.5 % of a random sample of 1500 school boys had the same defect. Is the different between the proportions significant ? Use 5 % level of significance 1.96.

Soln. :

Step I : Let, H_0 – The difference between the proportions is significant.

$$H_0 : P_1 = P_2$$

$$H_1 : P_1 \neq P_2.$$

Here $P_1 = 20\% = \frac{20}{100} = 0.2$;

$P_2 = 18.5\% = \frac{18.5}{100} = 0.185$;

$n_1 = 800$; $n_2 = 1500$;

$Z_\alpha = 1.96$ ...Given

Step II : The normal statistics

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Here, $$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{(800)(0.2) + (1500)(0.185)}{800 + 1500}$$

$$= \frac{160 + 277.5}{2300}$$

$$P = \frac{437.5}{2300} = 0.1902$$

$$Q = 1 - P = 1 - 0.1902 = 0.8098$$

$$\therefore \; Z = \frac{0.2 - 0.185}{\sqrt{(0.1902)(0.8098)\left[\frac{1}{800} + \frac{1}{1500}\right]}}$$

$$= \frac{0.015}{\sqrt{(0.1540)(0.00192)}}$$

$$= \frac{0.015}{(0.0172)} = 0.8720$$

Step III : $|Z| < |Z_\alpha|$

This shows that H_0 is rejected.

i.e. The difference between P_1 and P_2 is not sufficient.

Ex. 6.7.6 : 700 ladies, out of a sample of 1000, were consumers of Lipsticks before GST. After GST, 600 ladies; were consumers of lipsticks in a sample of 900 persons. Find out, whether there is a significant decrease in the consumption of lipsticks after G.S.T.

Given : 1 % level significance $Z_\alpha = 2.33$.

Soln. :

Step I : Let, H_0 – there is no significant decreases in lipsticks consumption of ladies after G.S.T.

$$H_1 : \neq H_0$$

$P_1 = \frac{700}{1000} = \frac{7}{10}$; $P_2 = \frac{600}{900} = \frac{1}{3}$

$n_1 = 1000$; $n_2 = 900$

$$P = \frac{n_1 P_1 + n_2 P_2}{n1 + n_2} = \frac{1000\left(\frac{7}{10}\right) + (900)\left(\frac{1}{3}\right)}{1000 + 900}$$

$$P = \frac{700 + 300}{1900} = \frac{1000}{1900} = 0.5263$$

$$\therefore \; Q = 1 - P = 1 - 0.5263 = 0.4737$$

Step II : Now, the normal statistics

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{\left(\frac{7}{10}\right) - \left(\frac{1}{3}\right)}{\sqrt{(0.5263)(0.4737)\left[\frac{1}{1000} + \frac{1}{900}\right]}}$$

$$= \frac{0.3666}{\sqrt{(2.4931)(0.0021)}} = \frac{0.3666}{\sqrt{0.00524}} = \frac{0.3666}{0.0724}$$

$$Z = 5.0635$$

Step III : $|Z| > |Z_\alpha|$, means,

$\therefore$ (5.0635) > (2.33)

Heance, hypothesis H_0 is rejected.

i.e. H_1 is accepted.

Hence, there is a significant decreasers in consumption of lipsticks after G.S.T.

Ex. 6.7.7 : It is observed that, 14 % of a random sample of 1500 doctors were smokers in a particular region. In the second servey 20 % of a random sample of 900 IT engineers were smokers, in a company. Can we say that, IT engineer's are more number of smokers than doctors. Use 5 % level of significance. $Z_\alpha = -1.645$.

Soln. :

Step I :

Let H_0 : Engineers have more number of smokers than doctors.

$$H_1 : \neq H_0$$

$$P_1 = 14\% = \frac{14}{100} = 0.14$$

$$P_2 = 20\% = \frac{20}{100} = 0.2$$

$$n_1 = 1500; \quad n_2 = 900$$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{(1500)(0.14) + (900)(0.2)}{1500 + 900}$$

$$P = \frac{210 + 180}{2400} = \frac{390}{2400} = 0.1625$$

$$Q = 1 - P = 1 - 0.1625 = 0.8375$$

Given, $Z_\alpha = -1.645$

Step II : Now, for normal statistics

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(0.14) - (0.2)}{\sqrt{(0.1625)(0.8375)\left(\frac{1}{1500} + \frac{1}{900}\right)}}$$

$$Z = -0.06$$

Step III : $|Z| < |Z_\alpha|$

Hence, H_0 is accepted.

i.e. Engineers have more number of smokers than doctors.

Ex. 6.7.8 : A sample of 100 screws are taken from a large population. The mean of sample is 160. Can we say that the mean of population is 165 and S.D. is 10. Consider 1 % of level of significance is $Z_\alpha = 2.58$.

Soln. :

Step I : Let, H_0 – The difference between mean of sample and mean of population is not significant.

$$H_1 : \neq H_0$$

Given, $\bar{x} = 160$; (mean of samples)

$\mu = 165$ (mean of population)

$\sigma = 10$; $n = 100$

Step II : By normal statistics is

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{160 - 165}{\frac{10}{\sqrt{100}}} = \frac{-5}{\frac{10}{10}} = -5$$

$$\therefore \; |Z| > |Z_\alpha|$$

This shows that, H_0 is rejected.

$\therefore$ The difference between $\bar{x}$ and μ is significant.

Ex. 6.7.9 : The mean of razer blades supplied by a manufacturer is 1800 with an S.D. 100. He claims that the strength of sharpness of blades has increased. To test his claim, a sample of 50 blades are tested and it is found that the mean of blades is 1850, with 1 % level of significance.

$Z_\alpha = 2.33$, can we accept the claim.

Soln. :

Step I : Let, H_0 - The difference, between two means is not significant.

$$H_1 : \neq H_0$$

$\bar{x} = 1850$; $\mu = 1800$;

$n = 50$; $\sigma = 100$; $Z_\alpha = 2.33$.

Step II : For normal statistic :

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1850 - 1800}{\left(\frac{100}{\sqrt{50}}\right)} = \frac{50}{14.142}$$

$$Z = 3.5356$$

$$\therefore \; |Z| > |Z_\alpha|$$

Hence, H_0 is regjected.

i.e. There is significant difference between $\bar{x}$ and μ.

i.e. We support the claim that the strength of sharpness of bladas is increased.

Ex. 6.7.10 : In a random sample of size 500, with mean 20. In another independent sample of size 400, mean is 15. Can we accept that the samples are drawn from the same sample with S.D. 4. Also Z_α for 1 % level of significance is 2.58.

Soln. :

Step I :

Let, H_0 : Both samples are taken from same population.

$$H_1 : \neq H_0$$

$\bar{x}_1 = 20;$ $n_1 = 500;$

$\bar{x}_2 = 15;$ $n_2 = 400;$

$\sigma = 4;$ $Z_\alpha = 2.58.$

Step II : For normal, statistics :

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{20-15}{4\sqrt{\frac{1}{500}+\frac{1}{400}}}$$

$$= \frac{5}{4\times(0.0045)} = \frac{5}{0.018}$$

$$Z = 277.778$$

Step III : This shows that $|Z| > |Z_\alpha|$

i.e. H_0 is rejected.

i.e. Difference between $\bar{x}_1$ and $\bar{x}_2$ is significant.

$\therefore$ H_1 is acceptable.

Ex. 6.7.11 : The sample heights of 6400 football players have mean of 170 cm and S.D. is 6.4 cm., while the sample of heights of 1600 soldiers has a mean of 172 cm and S.D. is 6.3 cm. Is the data shows that football players are taller than the soldiers, with 1 % level of significance $Z_\alpha = -2.33$.

Soln. :

Step I : Let, H_0 – There is no significant difference between the means.

$H_1 : \neq H_0$

$\bar{x}_1 = 170;$ $\bar{x}_2 = 172;$ $n_1 = 6400;$

$n_2 = 1600;$ $\sigma_1 = 6.4;$ $\sigma_2 = 6.3$

$Z_\alpha = -2.33$

Step II : For normal statistics

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{170-172}{\sqrt{\frac{(6.4)^2}{6400}+\frac{(6.3)^2}{1600}}}$$

$$= \frac{-2}{\sqrt{(0.0064)+(0.0248)}} = \frac{-2}{0.1766}$$

$$Z = -11.3250$$

Step III : This shows that $|Z| > |Z_\alpha|$

i.e. H_0 is rejected.

i.e. There is significant difference between $\bar{x}_1$ and $\bar{x}_2$.

i.e. Football players are taller than soldiers.

Ex. 6.7.12 : Test the significance of the difference between the means of the samples, drawn from two normal populations with the same S.D., using the following data :

	Size	Mean	S.D.
Sample - 1	100	61	4
Sample - 2	200	63	6

With 5 % level of significance $Z_\alpha = 1.96$.

Soln. :

Step I : Let H_0 – the difference between the means are significant.

$H_1 : \neq H_0$

$\therefore$ $\bar{x}_1 = 61;$ $\bar{x}_2 = 63;$ $\sigma_1 = 4;$

$\sigma_2 = 63;$ $n_1 = 100;$ $n_2 = 200.$

Given, $Z_\alpha = 1.96$

Step II : For normal statistics.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{\sigma_1^2}{n_2}+\frac{\sigma_2^2}{n_1}\right)}}$$

($\because$ If S.D. is same of two different population)

$$= \frac{61-63}{\sqrt{\left(\frac{4^2}{200}+\frac{6^2}{100}\right)}}$$

$$Z = \frac{-2}{\sqrt{0.44}} = -3.0151$$

Step III : This shows that :

$|Z| > |Z_\alpha|$

i.e. H_0 is rejected.

i.e. The difference between two means of sample is significant.

i.e. The means of two populations are not same.

Ex. 6.7.13 : Test the significance at 1 % level of significance whether the husband performs better than wives in examination from the following data :

	Average Marks	S.D.	No. of
Husband	72	8	32
Wives	70	6	36

Given, $Z_\alpha = 2.33$

Soln. :

Step I : H_0 – The difference between means of performance in exam is not significant.

$H_1 : \neq H_0$

$\bar{x}_1 = 72;$ $\bar{x}_2 = 70;$ $\sigma_1 = 8;$

$\sigma_2 = 6;$ $n_1 = 32;$ $n_2 = 36;$

$Z_\alpha = 2.33.$

Step II : For normal statistics :

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

($\because$ Here, it is assumed that the sample standard deviations are equal to populations standard deviations.)

$$\therefore Z = \frac{72-70}{\sqrt{\left(\frac{8^2}{32}+\frac{6^2}{36}\right)}} = \frac{2}{\sqrt{3}} = 1.1547$$

Step III : This shows that : $|Z| < |Z_\alpha|$

i.e. H_0 is accepted.

i.e. The difference between means is not significant at 1 % level of significance.

i.e. We cannot conclude that who perform better, with the help of given statistics.

Ex. 6.7.14 : The heights of men and women are distributed normally.

	Mean	S.D.
Men	171	7
Women	165	6

If man and women are selected randomly from the population. Find the probability that the women is taller than the man.

A (z = 0.650) = 0.2578

Soln. :

Step I :

Let $\bar{x}_1$ - be the mean height of men

$\bar{x}_2$ - be the mean height of women

σ_1 - be the S. D. of men height

σ_2 - be the S. D. of women height

V - is the variance

$\bar{x}_1$ follows normal distribution $N(\bar{x}_1, \sigma_1)$ and $\bar{x}_2$ also follows normal distribution $N(\bar{x}_2, \sigma_2)$

i.e. $\bar{x}_1$ follows N (171, 7) and $\bar{x}_2$ follows N (165, 6)

Therefore, $\bar{x}_1 - \bar{x}_2$ also follows normal distribution

$E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = 171 - 165 = 6$

and

$V(\bar{x}_1 - \bar{x}) = V(\bar{x}_1) + V(\bar{x}_2) = 7^2 + 6^2 = 49 + 36 = 85$

($\because$ Here $\bar{x}_1$ and $\bar{x}_2$ are independent means of different samples)

$\therefore$ S. D. of $(\bar{x}_1 - \bar{x}) = \sqrt{V(\bar{x}_1 - \bar{x}_2)} = \sqrt{85}$

$\therefore \sigma = 9.22$

$\therefore (\bar{x}_1 - \bar{x}_2)$ follows normal distribution N (6, 9.22)

Step II : We have to find he probability that the women are taller than men

i.e. $P(\bar{x}_2 - \bar{x}_1) = P(\bar{x}_2 - \bar{x}_1 > 0)$

$$= P(\bar{x}_1 - \bar{x}_2 < 0) = P\left(\frac{\bar{x}_1 - \bar{x}_2}{\sigma}\right)$$

$$= P\left(\frac{-6}{9.22}\right) = P(-0.650) = P(Z < -0.650)$$

$$= P(Z > 0.650)$$

($\because$ due to symmetry of normal curve about mean)

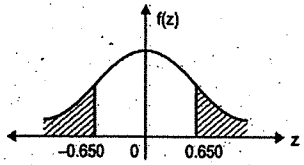


Fig. P. 6.7.14

$= 0.5 - P(0 < z < 0.650)$

$= 0.5 - 0.2578$

$= 0.2422$ ✓ ...Ans.

Ex. 6.7.15 : The population have the same mean, but the S.D. of one is twice that of the other. Show that in sample, each of size 500, drawn under simple random conditions, the difference of the means will not exceed 0.36, where σ is the S.D. of population.

A (z = 3) = 0.4974

Soln. : Step I :

$\bar{x}$ - be the mean of sample;

μ - be the mean of population;

σ - S. D.

V - variabnce

n - 500

σ and 2σ are S. D.s of two samples.

$\bar{x}_1$ follows normal distribution $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(\mu, \frac{\sigma}{\sqrt{500}}\right)$

$\bar{x}_2$ follows normal distribution $N\left(\mu, \frac{2\sigma}{\sqrt{n}}\right) = N\left(\mu, \frac{2\sigma}{\sqrt{500}}\right)$

($\because$ Here means are same for both samples)

Also, $\bar{x}_1 - \bar{x}_2$ follows normal distribution.

$E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu - \mu = 0$

$V(\bar{x}_1 - \bar{x}_2) = V(\bar{x}_1) + V(\bar{x}_2)$

($\because$ Samples are independent and means are also independent)

$$= \left(\frac{\sigma}{\sqrt{500}}\right)^2 + \left(\frac{2\sigma}{\sqrt{500}}\right)^2$$

$$V(\bar{x}_1 - \bar{x}_2) = \frac{\sigma^2}{500} + \frac{4\sigma^2}{500} = \frac{5\sigma^2}{500} = \frac{\sigma^2}{100}$$

$\therefore$ s. d. of $(\bar{x}_1 - \bar{x}_2) = \frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10}$

Step II : $(\bar{x}_1 - \bar{x}_2)$ follows normal distribution $N\left(0, \frac{\sigma}{10}\right)$

($\because$ Means are same)

To find the probability not exceeds 0.36.

$$\therefore P(|\overline{x_1} - \overline{x_2}| \le 0.36)$$

$$= P\left(z \le \frac{0.3\sigma}{\frac{\sigma}{10}}\right) = P(z \le 3)$$

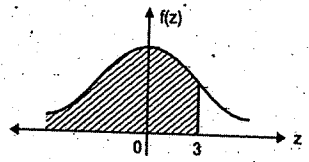


Fig. P. 6.7.15

$$= 0.5 + A(z \le 3) = 0.5 + 0.4974 = 0.9974 \approx 1$$

This shows that the difference of means will not exceed 0.36. ✓

...Ans.

Ex. 6.7.16 : A manufacturer of electric bulbs finds the population S. D. of the life of lamps to be 100 h. By adopting a new process in manufacturing, a sample of 150 bulbs having S. D. is of 95 h. If the manufacturer justified in chaning the process.

Given : 5% level of signficance $Z_\alpha = -1.645$

Soln. ; Step I :

Let H_0 – the difference between the S. D. of population and sample is not significant.

$\sigma = 100$ – be the S. D. of population

$S = 95$ – be the S. D. of sample

$n = 150$; $Z_\alpha = -1.645$

Step II : By the test of significance of the difference between sample S. D. and population S. D.

S follows normal distribution with normal statistics.

$$Z = \frac{S - \sigma}{\sigma/\sqrt{2n}} = \frac{95 - 100}{100/\sqrt{2 \times 150}} = \frac{-5}{(100/\sqrt{300})}$$

$$= \frac{-\sqrt{300}}{20} = -0.866$$

$$|Z| < |Z_\alpha|$$

This shows that H_0 is accepted.

i.e. The differenence beteeen S. D's of sample and population is not significant.

i.e. The manufacturer is not justified to the customers.

Ex. 6.7.17 : The S. D. of a random sample of 1000 is found to be 2.6 and the S. D. of another random sample of 2000 is 2.7. Assume that the samples are independent. Find out, whether the two samples comes from population with same S. D..

Given : 5% level of significance is $Z_\alpha = 1.96$.

Soln. :

Step I :

Let, $S_1 = 2.6$ be the S. D. of sample one

$S_2 = 2.7$ be the S. D. of second sample

$n_1 = 1000$, $n_2 = 2000$

H_0 = Difference between S. D. of sample and population.

$H_1 : \ne H_0$

Given : $Z_\alpha = 1.96$

Step II : Here ∴ Two S.D. of two different samples are given then by "Test of significance of the difference between S. D's of two large samples."

For normal statistics

$$Z = \frac{S_1 - S_2}{\sqrt{\frac{S_1^2}{2n_2} + \frac{S_2^2}{2n_1}}} = \frac{(2.6) - (2.7)}{\sqrt{\frac{(2.6)^2}{1000} + \frac{(2.7)^2}{2000}}}$$

$$= \frac{-0.1}{\sqrt{0.00676 + 0.00365}} = \frac{-0.1}{0.1020}$$

$$= -0.9804$$

Step III :

This shows that

$|Z| < |Z_\alpha|$ i.e. H_α is accepted

i.e. Difference between the S. D's. is not significant.

i.e. The two samples are come from the population with same S. D. ✓

...Ans.

Exercise

Ex. 1 : A manufacturer of electric bulbs finds the population S. D. of the life of lamps to be 70 h. By adopting a new process in manufacturing, a sample of 100 bulbs having S. D. is of 85 h. If the manufacturer justified in chaning the process.

Given : 5% level of signficance $Z\alpha = -1.645$

Ex. 2 : The heights of men and women are distributed normally.

	Mean	S. D.
Men	151	5
Women	125	4

If man and women are selected randomly from the population. Find the probability that the women is taller than the man.

$A(z = 0.650) = 0.2578$

Ex. 3 : Test the significance of the difference between the means of the samples, drawn from two normal populations with the same S.D., using the following data :

	Size	Mean	S. D.
Sample - 1	105	60	3
Sample - 2	275	73	7

With 5 % level of significance $Z_\alpha = 1.96$.

Ex. 4 : Test the significance at 1 % level of significance whether the husband performs better than wives in examination from the following data :

	Average Marks	S. D.	No. of
Husband	52	8	32
Wives	70	6	36

Given, $Z\alpha = 2.33$

Ex. 5 : A cubical die is thrown 900 times. The numbers 3 or 4 appears 340 times show that the die shows 3 or 4 numbers are not as an unbiased one. Use 5 % level of significance = 1.96.

Ex. 6 : 700 ladies, out of a sample of 1000, were consumers of Lipsticks before GST. After GST, 601 ladies; were consumers of lipsticks in a sample of 1100 persons. Find out, whether there is a significant decrease in the consumption of lipsticks after G.S.T.

Given : 1 % level significance $Z\alpha = 2.33$.

Multiple Choice Questins

Q. 1 By using least square method, for best fitting of a straight line y = mx + C, with n observations, the normal equations are :

(a) $\Sigma y = m\,\Sigma x + nC$; $\Sigma xy = m\,\Sigma x^2 + C\,\Sigma x$

(b) $\Sigma x = m\Sigma y + nC$; $\Sigma xy = m\,\Sigma y^2 + C\,\Sigma y$

(c) $\Sigma x = m\Sigma y - nC$; $\Sigma xy = m\,\Sigma y^2 - C\,\Sigma y$

(d) $\Sigma x = m\,\Sigma y + C$; $\Sigma xy = m\Sigma x^2 - C\,\Sigma x$

Ans : (a)

Explanation : Please refer section 6.3.

Q. 2 By using least square method, for best fitting of a straight line x = ay + b, with n observations, the normal equations are :

(a) $\Sigma y = a\,\Sigma x + bC$; $\Sigma xy = a\,\Sigma x^2 + b\,\Sigma x$

(b) $\Sigma x = a\,\Sigma y + nb$; $\Sigma xy = a\,\Sigma y^2 + b\,\Sigma y$

(c) $\Sigma x = m\Sigma y - nC$; $\Sigma xy = m\,\Sigma y^2 - C\,\Sigma y$

(d) $\Sigma x = m\,\Sigma y + C$; $\Sigma xy = m\Sigma x^2 - C\,\Sigma x$

Explanation : Please refer section 6.3.

Q. 3 By using least square method, for best fitting of a straight line ax + by = c , with n observations, the normal equations are (consider $y = f(x)$)

(a) $\Sigma x = \frac{b}{a}\Sigma y - n\frac{c}{a}$; $\Sigma xy = \frac{b}{a}\Sigma y^2 + \frac{c}{a}\Sigma y$

(b) $\Sigma x = -\frac{b}{a}\Sigma y + n\frac{c}{a}$; $\Sigma xy = -\frac{b}{a}\Sigma y^2 + \frac{c}{a}\Sigma y$

(c) $a\,\Sigma x + b\,\Sigma y = C$; $c\,\Sigma xy = a\,\Sigma y^2 - b\,\Sigma y$

(d) $\Sigma x = -\frac{c}{a}\Sigma y + n\frac{b}{a}$; $\Sigma xy = -\frac{c}{a}\Sigma y^2 + \frac{b}{a}\Sigma y$

Ans : (b)

Explanation : Please refer section 6.3.

Q. 4 By using least square method, for best fitting of a straight line ax + by = c , with n observations, the normal equations are (consider $x = f(y)$)

(a) $\Sigma x = \frac{b}{a}\Sigma y - n\frac{c}{a}$; $\Sigma xy = \frac{b}{a}\Sigma y^2 + \frac{c}{a}\Sigma y$

(b) $\Sigma y = -\frac{a}{b}\Sigma x + n\frac{c}{a}$; $\Sigma xy = -\frac{a}{b}\Sigma x^2 + \frac{c}{a}\Sigma x$

(c) $a\,\Sigma x + b\,\Sigma y = C$; $c\,\Sigma xy = a\,\Sigma y^2 - b\,\Sigma y$

(d) $\Sigma x = -\frac{c}{a}\Sigma y + n\frac{b}{a}$; $\Sigma xy = -\frac{c}{a}\Sigma y^2 + \frac{b}{a}\Sigma y$

Ans : (b)

Explanation : Please refer section 6.3.

Q. 5 By using least square method, for fitting a parabolic equation; $y = a + bx + cx^2$, with n observations , the normal equations are :

(a) $\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2$;
$\Sigma xy = a\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3$;
$\Sigma x^2 y = a\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4$

(b) $\Sigma x = na + b\,\Sigma y + c\,\Sigma y^2$
$\Sigma xy = a\Sigma y + b\,\Sigma y^2 + c\,\Sigma y^3$
$\Sigma xy^2 = a\Sigma y^2 + b\,\Sigma y^3 + c\,\Sigma y^4$

(c) $\Sigma y = na - b\,\Sigma x - c\,\Sigma x^2$
$\Sigma xy = a\Sigma x - b\,\Sigma x^2 - c\,\Sigma x^3$;
$\Sigma x^2 y = a\Sigma x^2 - b\,\Sigma x^3 - c\,\Sigma x^4$

(d) $\Sigma x = na - b\,\Sigma y - c\,\Sigma y^2$;
$\Sigma xy = a\Sigma y - b\,\Sigma y^2 - c\,\Sigma y^3$
$\Sigma xy^2 = a\Sigma y^2 - b\,\Sigma y^3 - c\,\Sigma y^4$

Ans : (a)

Explanation : Please refer section 6.4

Q. 6 By using least square method, for fitting a parabolic equation; $x = a + by + cy^2$, with n observations , the normal equations are :

(a) $\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2$;
$\Sigma xy = a\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3$;
$\Sigma x^2 y = a\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4$

(b) $\Sigma x = na + b\,\Sigma y + c\,\Sigma y^2$
$\Sigma xy = a\Sigma y + b\,\Sigma y^2 + c\,\Sigma y^3$
$\Sigma xy^2 = a\Sigma y^2 + b\,\Sigma y^3 + c\,\Sigma y^4$

(c) $\Sigma y = na - b\,\Sigma x - c\,\Sigma x^2$;
$\Sigma xy = a\Sigma x - b\,\Sigma x^2 - c\,\Sigma x^3$;
$\Sigma x^2 y = a\Sigma x^2 - b\,\Sigma x^3 - c\,\Sigma x^4$

(d) $\Sigma x = na - b\,\Sigma y - c\,\Sigma y^2$
$\Sigma xy = a\Sigma y - b\,\Sigma y^2 - c\,\Sigma y^3$
$\Sigma xy^2 = a\Sigma y^2 - b\,\Sigma y^3 - c\,\Sigma y^4$

Ans : (b)

Explanation : Please refer section 6.4.

Q. 7 For a linear load of the form : P = mW + C, where P and W are variables, using following data :

P	12	15	21	25
W	50	70	100	120

Then one of the normal equation is

(a) $12 = 4c + 50\,m$

(b) $73 = 4c + 340\,m$

(c) $25 = 4c + 120\,m$

(d) None of the above

Ans. : (b)

Explanation :

To find the constants m and C by Direct Method :

The given law is

$$P = mW + C \quad ...(1)$$

The corresponding normal equation are

$$\sum P = nC + m\sum W \quad ...(2)$$

and $$\sum PW = C\sum W + m\sum W^2 \quad ...(3)$$

Here $n = 4$

Step II : Construct table for $\sum P, \sum W, \sum W^2, \sum PW$

P	W	W^2	PW
12	50	2500	600
15	70	4900	1050
21	100	10000	2100
25	120	14400	3000
$\sum P = 73$	$\sum W = 340$	$\sum W^2 = 31800$	$\sum PW = 6750$

From Equation (2),

$$73 = 4C + m\,(340)$$

Q. 8 If the linear load of the form P = mW + C, where P and W are variables and $\sum P = 73$, $\sum W = 340$, $\sum W^2 = 31800$ $\sum PW = 6750$, n = 4, then the corresponding normal equations are :

(a) $73 = 4c + 340\,m$; $6750 = 340\,c + 31800\,m$

(b) $340 = 4c + 73\,m$; $31800 = 340c + 6750\,m$

(c) $73 + 4c = 340\,m$; $6750 + 340c = 31800\,m$

(d) $31800 = 4c + 73\,m$; $73 = 4c + 340\,m$

Ans. : (a)

Explanation :

To find the constants m and C by Direct Method :

The given law is

$$P = mW + C \quad ...(1)$$

The corresponding normal equation are

$$\sum P = nC + m\sum W \quad ...(2)$$

and $$\sum PW = C\sum W + m\sum W^2 \quad ...(3)$$

Here $n = 4$

Step II : Construct table for $\sum P, \sum W, \sum W^2, \sum PW$

P	W	W^2	PW
12	50	2500	600
15	70	4900	1050
21	100	10000	2100
25	120	14400	3000
$\sum P = 73$	$\sum W = 340$	$\sum W^2 = 31800$	$\sum PW = 6750$

Step III : From Equation (2),

$$73 = 4C + m\,(340) \text{ and from Equation (3),}$$

$$6750 = C(34.0) + m\,(31800)$$

Therefore, we get,

$$4C + 340\,m = 73 \quad ...(4)$$

and $$34C + 3180\,m = 675 \quad ...(5)$$

Q. 9 For fitting a straight line to the following data : (x, y) = (1, 1), (2, 5), (3, 11), (4, 8), (5, 14), the corresponding normal equations are :

(a) $146 = 15\,c - 55m$; $39 = 5c - 15\,m$

(b) $39 = 5c + 15\,m$; $146 = 15c + 55m$

(c) $5c + 15\,m = 0$; $15c + 55\,m = 0$

(d) $15 = 5c + 39\,m$; $55 = 15c + 146\,m$

Ans. : (b)

Explanation : Since the normal equations are.

$$\sum y = n\,C + m\sum x \quad ...(2)$$

and $$\sum xy = C\sum x + m\sum x^2 \quad ...(3)$$

Here, $n = 5$

Step I : Construct table for $\sum x, \sum y, \sum x^2$

x	y	x^2	xy
1	1	1	1
2	5	4	10
3	11	9	33
4	8	16	32
5	14	25	70
$\sum x = 15$	$\sum y = 39$	$\sum x^2 = 55$	$\sum xy = 146$

Step II : From Equation (2)

$$39 = 5C + 15\,m \quad ...(4)$$

and $$146 = 15C + 55m \quad ...(5).$$

Q. 10 For fitting a straight line y = mx + C, for the following data ; $\sum x = 0$; $\sum y = -16$; $\sum x^2 = 10$, $\sum xy = 29$; n = 5, the corresponding normal

equations are :

(a) $-16 = 5c + 10\,m$; $29 = 10\,m$

(b) $10 = 29c + 5$; $-16 = 10\,c$

(c) $-16 = 5c$; $29 = 10\,m$

(d) $-16 = 5c + 29\,m$; $10 = -16c + m$

Ans. : (c)

Explanation :

The corresponding normal equations are

$$\sum Y = nC + m\sum X \quad ...(1)$$

and $$\sum XY = C\sum X + m\sum X \quad ...(2)$$

From Equations (1) and (2)

$$-16 = 5C + 0$$

and $$29 = 0 + m\,(10)$$

Q. 11 For fitting the data :

x	0	1	2	3	4
y	1	1.8	1.3	2.5	6.3

To the parabola $y = a + bx + cx^2$, the first normal equation is,

(a) $12.9 = 5a + 10b + 30c$

(b) $50 = 12.9a + 30b + 10c$

(c) $12.9 = 5a - 10b - 30c$

(d) $30 = 5a + 12.9b + 10c$

Ans. : (a)

Explanation : Since, the normal equations are

$$\sum y = na + b\sum x + c\sum x^2 \quad ...(1)$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 \quad ...(2)$$

and $$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 \quad ...(3)$$

Construct table for

$$\sum x, \sum y, \sum x^2, \sum xy, \sum x^3, \sum x^4, \sum x^2 y$$

Here, $n = 5$

x	y	xy	x^2	x^2y	x^3	x^4
0	1	0	0	0	0	0
1	1.8	1.8	1	1.8	1	1
2	1.3	2.6	4	5.2	8	16
3	2.5	7.5	9	22.5	27	81
4	6.3	25.2	16	100.8	64	256
$\sum x = 10$	$\sum x = 12.9$	$\sum xy = 37.1$	$\sum x^2 = 30$	$\sum x^2y = 130.3$	$\sum x^3 = 100$	$\sum x^4 = 354$

By using normal equations

From Equation (1),

$$12.9 = 5a + 10b + 30c \quad ...(4)$$

12.10

Q. 12 Fit a non-linear trend of the parabolic form $y = a + bx + cx^2$ to the data n = 5, $\sum x = 0$, $\sum y = 2$ $\sum x^2 = 10$, $\sum x^3 = 0$, $\sum x^4 = 34$.

First two normal equations are :

(a) $5 = 2.5a + 10c$; $15 = 10b$

(b) $2.5 = 5a$; $10 = 15b + 10c$

(c) $2.5 = 5a - 10c$; $10 = 15b$

(d) $2.5 = 5a + 10c$; $15 = 10b$

Ans. : (d)

Explanation :

The parabolic curve is $y = a + bx + c\,x^2$

The normal equations becomes

$$\sum Y = na + b\sum X + c\sum X^2 \quad ...$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3 \quad ...$$

and $$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4 \quad ...$$

By Putting the values from given data in abo equations, we get the required normal equatio

$2.5 = 5a + 10c$; $15 = 10b$

x	1	−1	2
y	3	−1	5

Q. 13 By using least square method, the best fit of straight line y = mx + c with 3 points, for data , is

(a) $y = 2x + 1$ (b) $y = 2x - 1$

(c) $y = x + 1$ (d) $y = 3x + 2$

Ans : (a)

Explanation : Please refer above example 10

y	0	−1	3
x	−1	−6	14

Q. 14 By using least square method, the best fit of straight line x = ay + b with 3 points, for data , is

(a) $x = 5y + 1$ (b) $x = 5y - 1$

(c) $x = y + 11$ (d) $x = 3y - 3$

Ans : (b)

Explanation : Please refer above example 10

Q. 15 The least square fit of the straight line x = my with 3 points, for the data , is

y	1	2	−1
x	2	5	−4

(a) $x = 3y - 1$ (b) $x = 3y + 1$

(c) $x = y - 1$ (d) $x = 5y - 1$

Ans : (a)

Explanation : Please refer above example 10.

Q. 16 The least square fit of the non linear equation of the form $y = a x^b$ with 3 points, for the data, is

x	1	−1	2
y	3	−3	24

(a) $y = 3x^2$ (b) $y = 3x^3$

(c) $y = x + 2$ (d) $y = -3x^2$

Ans : (b)

Explanation : Please refer example 6.5.1.

Q. 17 By using least square method, the best fit of non linear equation of the form $y = ax^b$ with 3 points, for the data, is

x	0	2	3
y	0	8	18

(a) $y = 2x^2$ (b) $y = x^3$

(c) $y = 2x + 4$ (d) $y = -3x^3$

Ans. : (a)

Explanation : Please refer example 6.5.1

Chapter Ends...

□□□

Note