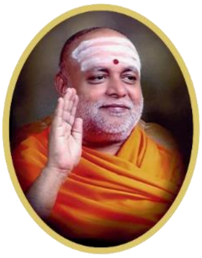


||JAI SRI GURUDEV||
S.J.C. INSTITUTE OF TECHNOLOGY, CHICKBALLAPUR
Department of Mathematics
LECTURE NOTES
MATHEMATICS-3 FOR COMPUTER SCIENCE STREAM (BCS301)
MODULE - 1
RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Prepared by:
Purushotham P
Assistant Professor
SJC Institute of Technology
Email id: purushotham@sjcit.ac.in

Random Experiment:

An activity that yield some results called the random experiment. The random variable means a real number, i.e. X associated with the outcomes of a random experiment.

Definition: Let S be a sample space associated with a random experiment with a real value function defined and taking its values is called a Random variable.

The random variables are two types. They are,

i) Discrete Random Variables (DRV)
ii) Continuous Random Variables (CRV)

Discrete Random Variables: A Discrete random variable is a variable which can only take a countable number of values.

For example, if a coin is tossed three times, the number of heads can be obtained is 0, 1, 2 or 3. The probabilities of each of these probabilities can be tabulated as shown.

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

X	0	1	2	3
P(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Continuous Random variables: A Continuous random variable is a random variable where the data can take infinitely many values. For example, a random variable measuring the time taken for something to be done is continuous since there are an infinite number of possible times that can be taken.

Ex: Temperature of the climate, Age of a person, etc.

Probability Mass Function:

Probability mass function is the probability distribution of a discrete random variable and provides the possible values and their associated probabilities.

1. $P(x_i) \geq 0$
2. $\sum_{i=1}^{n} P(X = x_i) = 1$
3. $0 \leq P(x) \leq 1$
4. $Mean\ \mu = \sum_{i=1}^{n} x_i P(x_i)$
5. $Variance\ \sigma^2 = \sum_{i=1}^{n} x_i^2 P(x_i) - \mu^2$

Probability Density Function:

Probability density function is the probability distribution of a continuous random variable and provides the possible values and their associated probabilities infinitely.

1. $P(x_i) \geq 0 \; or \; f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. Mean $\mu = \int_{-\infty}^{\infty} xf(x)dx$
4. Variance $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$
5. $P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) = P(a < x < b) = \int_a^b f(x)dx$

PROBLEMS

1) Show that the following probabilities one satisfying the properties of discrete random variables, hence find it's mean and variance.

x	10	20	30	40
P(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Sol^n: Let X be the random variable for the random values,

$x_1 = 10, x_2 = 20, x_3 = 30, x_4 = 40$

and given

$$P(X = x_1) = P(x_1) = p_1 = \frac{1}{8}$$
$$P(X = x_2) = P(x_2) = p_2 = \frac{3}{8}$$
$$P(X = x_3) = P(x_3) = p_3 = \frac{3}{8}$$
$$P(X = x_4) = P(x_4) = p_4 = \frac{1}{8}$$

Let $\sum_{i=1}^{4} P(X = x_i) = P(X = x_1) + P(X = x_2) + P(X = x_3) + P(X = x_4)$

$= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8}$

$= \frac{8}{8}$

$= 1$

Hence the given probabilities can satisfy the DRV property.

Mean $\mu = \sum_{i=1}^{4} x_i P(x_i)$

$= 10 \times \frac{1}{8} + 20 \times \frac{3}{8} + 30 \times \frac{3}{8} + 40 \times \frac{1}{8}$

$= \frac{200}{8}$

$= 25$

Variance $\sigma^2 = \sum_{i=1}^{4} x_i^2 P(x_i) - \mu^2$

$= 10^2 \times \frac{1}{8} + 20^2 \times \frac{3}{8} + 30^2 \times \frac{3}{8} + 40^2 \times \frac{1}{8} - 25^2$

$= 700 - 625$

$= 75$

S.D$= \sqrt{Variance} = \sqrt{75} = 8.66$

2) Find the value of k, such that the following distribution represents discrete probability distribution. Hence find Mean, S.D, $P(x \leq 1)$, $P(x > 1)$ and $P(-1 < x \leq 2)$.

x	-3	-2	-1	0	1	2	3
P(x)	k	2k	3k	4k	3k	2k	k

Sol[n]: Let X be the random variable for the random values,
$x_1 = -3, x_2 = -2, x_3 = -1, x_4 = 0, x_5 = 1, x_6 = 2, x_7 = 3$
and the given probabilities are,
$P(X = x_1) = P(-3) = k$
$P(X = x_2) = P(-2) = 2k$
$P(X = x_3) = P(-1) = 3k$
$P(X = x_4) = P(0) = 4k$
$P(X = x_5) = P(1) = 3k$
$P(X = x_6) = P(2) = 2k$
$P(X = x_7) = P(3) = k$

We know that,
$\sum_{i=1}^{7} P(X = x_i) = 1$

$$\Rightarrow k + 2k + 3k + 4k + 3k + 2k + k = 1$$
$$\Rightarrow 16k = 1$$
$$\Rightarrow k = \frac{1}{16}$$

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
-3	K	-3k	9	9k
-2	2k	-4k	4	8k
-1	3k	-3k	1	3k
0	4k	0	0	0
1	3k	3k	1	3k
2	2k	4k	4	8k
3	K	3k	9	9k
$\sum$		0	-	40k

Mean $\mu = \sum_{i=1}^{4} x_i P(x_i) = 0$

Variance $\sigma^2 = \sum_{i=1}^{4} x_i^2 P(x_i) - \mu^2$

$= 40k - 0^2$

$= 40 \times \frac{1}{16}$

$= 2.5$

S.D $= \sqrt{2.5} = 1.5811$

$i)\ P(x \le 1) = P(-3) + P(-2) + P(-1) + P(0) + P(1)$
$\Rightarrow P(x \le 1) = k + 2k + 3k + 4k + 3k$
$\Rightarrow P(x \le 1) = 13k = \frac{13}{16} = 0.8125$

$ii)\ P(x > 1) = P(2) + P(3) = 2k + k = 3k = \frac{3}{16} = 0.1875$

$iii)\ P(-1 < x \le 2) = P(0) + P(1) + P(2) = 4k + 3k + 2k = 9k = \frac{9}{16} = 0.5625$

3) Find the value of k, such that the following distribution represents discrete probability distribution. Hence find Mean, S.D, $P(x \geq 5)$ and $P(3 < x \leq 6)$.

x	0	1	2	3	4	5	6
$P(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

Sol^n: Let X be the random variable for the random values,
$x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 4$, $x_6 = 5$, $x_7 = 6$
and the given probabilities are,

$P(X = x_1) = P(0) = k$
$P(X = x_2) = P(1) = 3k$
$P(X = x_3) = P(2) = 5k$
$P(X = x_4) = P(3) = 7k$
$P(X = x_5) = P(4) = 9k$
$P(X = x_6) = P(5) = 11k$
$P(X = x_7) = P(6) = 13k$

We know that,

$\sum_{i=1}^{7} P(X = x_i) = 1$

$$\Rightarrow k + 3k + 5k + 7k + 9k + 11k + 13k = 1$$
$$\Rightarrow 49k = 1$$
$$\Rightarrow k = \frac{1}{49}$$

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
0	k	0	0	0
1	3k	3k	1	3k
2	5k	10k	4	20k
3	7k	21k	9	63k
4	9k	36k	16	144k
5	11k	55k	25	275k
6	13k	78k	36	468k
$\sum$		203k	-	973k

Mean $\mu = \sum_{i=1}^{4} x_i P(x_i) = 203k = \frac{203}{49} = 4.1428$

Variance $\sigma^2 = \sum_{i=1}^{4} x_i^2 P(x_i) - \mu^2$

$= 973k - 4.1428^2$

$= \frac{973}{49} - 17.1628$

$= 2.6943$

S.D $= \sqrt{2.6943} = 1.6414$

$$i)\ P(x \geq 5) = P(5) + P(6)$$
$$\Rightarrow P(x \geq 5) = 11k + 13k$$
$$\Rightarrow P(x \geq 5) = 24k = \frac{24}{49} = 0.4898$$

$$ii)\ P(3 < x \leq 6) = P(4) + P(5) + P(6) = 9k + 11k + 13k = 33k = \frac{33}{49} = 0.6734$$

4) A random variable X has a probability function for various values of x. Find i) k, ii) $P(x<6)$, iii) $P(x\geq 6)$ and $P(3<x\leq 6)$. Also find the probability distribution and distribution function of x.

x	0	1	2	3	4	5	6	7
$P(x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

Sol^n: Let X be the random variable for the random values,

$x_1 = 0,\ x_2 = 1,\ x_3 = 2,\ x_4 = 3,\ x_5 = 4,\ x_6 = 5,\ x_7 = 6,\ x_8 = 7$

and the given probabilities are,

$P(X=x_1)=P(0)=0$
$P(X=x_2)=P(1)=k$
$P(X=x_3)=P(2)=2k$
$P(X=x_4)=P(3)=2k$
$P(X=x_5)=P(4)=3k$
$P(X=x_6)=P(5)=k^2$
$P(X=x_7)=P(6)=2k^2$
$P(X=x_7)=P(7)=7k^2+k$

We know that,

$\sum_{i=1}^{7} P(X=x_i)=1$

$$\Rightarrow 0+k+2k+2k+3k+k^2+2k^2+7k^2+k=1$$

$$\Rightarrow 10k^2+9k=1$$

$$\Rightarrow 10k^2+9k-1=0$$

$$\Rightarrow (10k-1)(k+1)=0$$

$$\Rightarrow 10k-1=0\,,k+1=0$$

$$\Rightarrow k=\frac{1}{10}\,,k\neq -1$$

x	0	1	2	3	4	5	6	7
$P(x)$	O	0.1	0.2	0.2	0.3	0.01	0.02	0.17
CP	0	0.1	0.3	0.5	0.8	0.81	0.83	1

$$i)\ P(x<6)=1-P(x\geq 6)=1-\{P(6)+P(7)\}=1-\{0.02+0.17\}=0.81$$

$$ii)\ P(x\geq 6)=P(6)+P(7)=0.02+0.17=0.19$$

$$iii)\ P(3<x\leq 6)=P(4)+P(5)+P(6)=0.3+0.01+0.02=0.33$$

5) A random variable has the following probability function for the various values of X=x. Find i) Value of k, ii) $P(x < 1)$, iii)$P(x \geq 1)$.

x	-2	-1	0	1	2	3
$P(x)$	0.1	k	0.2	$2k$	0.3	k

Sol^n: Let X be the random variable for the random values,
$x_1 = -2$, $x_2 = -1$, $x_3 = 0$, $x_4 = 1$, $x_5 = 2$, $x_6 = 3$,
and the given probabilities are,
$P(X = x_1) = P(-2) = 0.1$
$P(X = x_2) = P(-1) = k$
$P(X = x_3) = P(0) = 0.2$
$P(X = x_4) = P(1) = 2k$
$P(X = x_5) = P(2) = 0.3$
$P(X = x_6) = P(3) = k$

i) We know that,
$\sum_{i=1}^{6} P(X = x_i) = 1$

$$\Rightarrow 0.1 + k + 0.2 + 2k + 0.3 + k = 1$$

$$\Rightarrow 4k + 0.6 = 1$$

$$\Rightarrow 4k = 0.4$$

$$\Rightarrow k = 0.1$$

$$ii)\ P(x < 1) = P(-2) + P(-1) + P(0) = 0.1 + k + 0.2 = k + 0.3 = 0.1 + 0.3 = 0.4$$

$$iii)\ P(x \geq -1) = P(-1) + P(0) + P(1) + P(2) + P(3) = k + 0.2 + 2k + 0.3 + k = 4k + 0.5 = 0.9$$

6) A random variable has the following probability function for the various values of X=x. Find i)Value of k, ii)$P(x \leq 1)$, iii) $P(0 \leq x < 3)$.

x	0	1	2	3	4	5
$P(x)$	k	$5k$	$10k$	$10k$	$5k$	k

Sol^n: Let X be the random variable for the random values,
$x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 4$, $x_6 = 5$,
and the given probabilities are,
$P(X = x_1) = P(0) = k$
$P(X = x_2) = P(1) = 5k$
$P(X = x_3) = P(2) = 10k$
$P(X = x_4) = P(3) = 10k$
$P(X = x_5) = P(4) = 5k$
$P(X = x_6) = P(5) = k$

i) We know that,
$\sum_{i=1}^{6} P(X = x_i) = 1$

$$\Rightarrow k + +5k + 10k + 10k + 5k + k = 1$$

$$\Rightarrow 32k = 1$$

$$\Rightarrow k = \frac{1}{32}$$

$$ii)\ P(x \leq 1) = P(0) + P(1) = k + 5k = 6k = \frac{6}{32} = 0.1875$$

$$iii)\ P(0 \leq x < 3) = P(0) + P(1) + P(2) = k + 5k + 10k = 16k = \frac{16}{32} = 0.5$$

7) Show that the function $f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ is probability density function. Hence find $P(1.5 < x < 2.5)$.

Sol^n: Given probability function,

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Let $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{\infty} f(x)dx$

$$= \int_{-\infty}^{0} 0\, dx + \int_{0}^{\infty} e^{-x}\, dx$$

$$= 0 + \left[\frac{e^{-x}}{-1}\right]_0^{\infty}$$

$$= -[e^{-\infty} - e^{0}]$$

$$= -[0 - 1]$$

$$= 1$$

Hence the given probability function is p.d.f.

$$P(1.5 < x < 2.5) = \int_{1.5}^{2.5} f(x)dx$$

$$= \int_{1.5}^{2.5} e^{-x} dx$$

$$= -[e^{-x}]_{1.5}^{2.5}$$

$$= -[e^{-2.5} - e^{-1.5}] = \left[\frac{1}{e^{1.5}} - \frac{1}{e^{2.5}}\right]$$

8) A random variable X has probability density function $f(x) = \begin{cases} kx^2, & 0 \leq x \leq 3 \\ 0 & , otherwise \end{cases}$, Evaluate i) k ii) $P(x \leq 1)$, iii) $P(x > 1)$, iv) $P(1 \leq x \leq 2)$, v) $P(x \leq 2)$, vi) $P(x \geq 2)$.

Sol^n: Given probability function,

$$f(x) = \begin{cases} kx^2, & 0 \leq x \leq 3 \\ 0 & , otherwise \end{cases}$$

$$i) \int_{-\infty}^{\infty} f(x)dx = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx = 1$$

$$\Rightarrow 0 + \int_{0}^{3} kx^2 dx + 0 = 1$$

$$\Rightarrow k\left[\frac{x^3}{3}\right]_0^3 = 1$$

$$\Rightarrow 9k = 1$$

$$\Rightarrow k = \frac{1}{9}$$

$$ii)\ P(x \leq 1) = \int_{-\infty}^{1} f(x)dx$$

$$\Rightarrow P(x \leq 1) = \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx$$

$$\Rightarrow P(x \le 1) = 0 + \int_0^1 kx^2 dx$$

$$\Rightarrow P(x \le 1) = k\left[\frac{x^3}{3}\right]_0^1$$

$$\Rightarrow P(x \le 1) = \frac{k}{3} = \frac{1}{27}$$

$$iii)\ P(x > 1) = \int_1^\infty f(x)dx$$

$$\Rightarrow P(x > 1) = \int_1^3 f(x)dx + \int_3^\infty f(x)dx$$

$$\Rightarrow P(x > 1) = 0 + \int_1^3 kx^2 dx$$

$$\Rightarrow P(x < 1) = k\left[\frac{x^3}{3}\right]_1^3$$

$$\Rightarrow P(x < 1) = \frac{26k}{3} = \frac{26}{27}$$

$$iv)\ P(1 \le x \le 2) = \int_1^2 f(x)dx$$

$$\Rightarrow P(1 \le x \le 2) = \int_1^2 kx^2 dx$$

$$\Rightarrow P(1 \le x \le 2) = k\left[\frac{x^3}{3}\right]_1^2$$

$$\Rightarrow P(1 \le x \le 2) = \frac{7k}{3} = \frac{7}{27}$$

$$v)\ P(x \le 2) = \int_{-\infty}^2 f(x)dx$$

$$\Rightarrow P(x \le 2) = \int_{-\infty}^0 f(x)dx + \int_0^2 f(x)dx$$

$$\Rightarrow P(x \le 2) = 0 + \int_0^1 kx^2 dx$$

$$\Rightarrow P(x \le 2) = k\left[\frac{x^3}{3}\right]_0^2$$

$$\Rightarrow P(x \le 2) = \frac{8k}{3} = \frac{8}{27}$$

$$vi)\ P(x \ge 2) = \int_2^\infty f(x)dx$$

$$\Rightarrow P(x \ge 2) = \int_2^3 f(x)dx + \int_3^\infty f(x)dx$$

$$\Rightarrow Px \ge 2) = 0 + \int_2^3 kx^2 dx$$

$$\Rightarrow P(x \ge 2) = k\left[\frac{x^3}{3}\right]_2^3$$

$$\Rightarrow P(x \ge 2) = \frac{19k}{3} = \frac{19}{27}$$

9) A random variable X has the pdf, $f(x)=\begin{cases} kx^2 & ,-3\le x\le 3 \\ 0 & ,\ otherwise\end{cases}$, find i) k, ii) $P(x\le 2)$, iii) $P(x\ge 2)$, iv) $P(x>1)$, v) $P(1\le x\le 2)$.

Sol^n: Given probability function,

$$f(x)=\begin{cases} kx^2, -3\le x\le 3 \\ 0 \qquad ,\ otherwise\end{cases}$$

$$i)\ \int_{-\infty}^{\infty} f(x)dx = 1$$
$$\Rightarrow \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{-3} f(x)dx + \int_{-3}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx = 1$$
$$\Rightarrow 0 + \int_{-3}^{3} kx^2dx + 0 = 1$$
$$\Rightarrow k\left[\frac{x^3}{3}\right]_{-3}^{3} = 1$$
$$\Rightarrow 18k = 1$$
$$\Rightarrow k = \frac{1}{18}$$

$$ii)\ P(x\le 2) = \int_{-\infty}^{2} f(x)dx$$
$$\Rightarrow P(x\le 2) = \int_{-\infty}^{-3} f(x)dx + \int_{-3}^{2} f(x)dx$$
$$\Rightarrow P(x\le 2) = 0 + \int_{-3}^{2} kx^2dx$$
$$\Rightarrow P(x\le 2) = k\left[\frac{x^3}{3}\right]_{-3}^{3}$$
$$\Rightarrow P(x\le 2) = \frac{35k}{3} = \frac{35}{3}\times\frac{1}{18} = \frac{35}{54}$$

$$iii)\ P(x\ge 2) = \int_{2}^{\infty} f(x)dx$$
$$\Rightarrow P(x\ge 2) = \int_{2}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx$$
$$\Rightarrow P(x\ge 2) = \int_{2}^{3} kx^2dx$$
$$\Rightarrow P(x\ge 2) = k\left[\frac{x^3}{3}\right]_{2}^{3}$$
$$\Rightarrow P(x\ge 2) = \frac{19k}{3} = \frac{19}{3}\times\frac{1}{18} = \frac{19}{54}$$

$$iv)\ P(x>1) = \int_{1}^{\infty} f(x)dx$$
$$\Rightarrow P(x>1) = \int_{1}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx$$
$$\Rightarrow P(x>1) = \int_{1}^{3} kx^2dx$$
$$\Rightarrow P(x>1) = k\left[\frac{x^3}{3}\right]_{1}^{3}$$
$$\Rightarrow P(x>1) = \frac{26k}{3} = \frac{26}{3}\times\frac{1}{18} = \frac{26}{54}$$

$$v)\ P(1 \le x \le 2) = \int_1^2 f(x)dx$$
$$\Rightarrow P(1 \le x \le 2) = \int_1^2 kx^2 dx$$
$$\Rightarrow P(1 \le x \le 2) = k\left[\frac{x^3}{3}\right]_1^2$$
$$\Rightarrow P(1 \le x \le 2) = \frac{7k}{3} = \frac{7}{3} \times \frac{1}{18} = \frac{7}{54}$$

10) The diameter of an electric cable is assumed to be a CRV with pdf $f(x) = \begin{cases} kx(1-x) & ,0 \le x \le 1 \\ 0 & otherwise \end{cases}$, find i) value of k, ii) Mean & Variance.

Sol^n: Given probability function,

$$WKT, \int_{-\infty}^{\infty} f(x)dx = 1$$
$$\Rightarrow i) \int_{-\infty}^{0} f(x)dx + \int_0^1 f(x)dx + \int_1^{\infty} f(x)dx = 1$$
$$\Rightarrow 0 + \int_0^1 kx(1-x)dx + 0 = 1$$
$$\Rightarrow k\int_0^1 (x - x^2)dx = 1$$
$$\Rightarrow k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 1$$
$$\Rightarrow \frac{k}{6} = 1$$
$$\Rightarrow k = 6$$

$$ii)\ Mean\ \mu = \int_{-\infty}^{\infty} xf(x)dx$$
$$= \int_0^1 xkx(1-x)dx$$
$$= k\int_0^1 x^2(1-x)dx$$
$$= k\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1$$
$$= k\left[\frac{1}{3} - \frac{1}{4}\right] = \frac{k}{12} = \frac{6}{12} = \frac{1}{2}$$

$$iii)\ Variance\ \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$
$$= \int_0^1 kx^3(1-x)dx - \mu^2$$
$$= k\int_0^1 (x^3 - x^4)dx - \left[\frac{1}{2}\right]^2$$
$$= k\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 - \frac{1}{4}$$
$$= \frac{k}{20} - \frac{1}{4} = \frac{6}{20} \times \frac{1}{4} = \frac{1}{20}$$

11) Find the constant k such that $f(x) = \begin{cases} kxe^{-x} & ,0 < x < 1 \\ 0 & otherwise \end{cases}$, is pdf. Find the mean.

Sol^n: Given probability function,

$$f(x) = \begin{cases} kxe^{-x} & ,0 < x < 1 \\ 0 & otherwise \end{cases}$$

also given f(x) represents pdf for the CRV 'X'.

$$WKT, \int_{-\infty}^{\infty} f(x)dx = 1$$

$$\Rightarrow \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{\infty} f(x)dx = 1$$

$$\Rightarrow \int_{0}^{1} kxe^{-x}dx = 1$$

$$\Rightarrow k\left\{\int_{0}^{1} e^{-x}dx - \int_{0}^{1}\left(1 \times \int e^{-x}dx\right)dx\right\} = 1$$

$$\Rightarrow k(-e^{-1} - e^{-1} + 1 = 1$$

$$\Rightarrow k\left(1 - \frac{2}{e}\right) = 1$$

$$\Rightarrow k = \frac{e}{e-2}$$

$$Mean\ \mu = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \int_{0}^{1} xkxe^{-x}dx$$

$$= k\int_{0}^{1} x^2e^{-x}dx$$

$$= k\left\{x^2\int_{0}^{1} e^{-x}dx - \int_{0}^{1}\left(2x \times \int e^{-x}dx\right)dx\right\}$$

$$= k\left\{-(x^2e^{-x})_0^1 + 2\int_{0}^{1} xe^{-x}dx\right\}$$

$$= k(2 - 5e^{-1})$$

$$= k\left(2 - \frac{5}{e}\right)$$

$$= \frac{e}{e-2} \times \frac{2e-5}{e}$$

$$\mu = \frac{2e-5}{e-2}$$

Binomial Distribution:

Let X be a discrete random variable, 'p' be the probability of success and let 'q' be the probability of failure, then the probability mass function of the binomial distribution can be defined as,

$$P(X=x)=b(n,p,x)=\begin{cases} n_{c_x}p^x q^{n-x} & ,x\geq 0 \\ 0 & ,Otherwise \end{cases}$$

where, n is the number of trials and n & p are the parameters which follows,

i) $P(X=x)=b(n,p,x)\geq 0$

ii) p + q = 1

iii) $\sum_{x=0}^{n} n_{c_x}p^x q^{n-x}=1$

iv) The mean of B.D $\mu=np$, Vairance $\sigma^2=npq$ and S.D is $\sigma=\sqrt{npq}$.

MEAN & VARIANCE OF A BINOMIAL DISTRIBUTION:

WKT, the probability mass function of the binomial distribution is,

$$P(X=x)=f(x)=\begin{cases} n_{c_x}p^x q^{n-x}, & x\geq 0 \\ 0 \quad , & otherwise \end{cases}$$

i) Mean:

$$\mu=E(x)=\sum_{x=0}^{n} xP(X=x)$$

$$=\sum_{x=0}^{n} x\, n_{c_x}p^x q^{n-x}$$

$$=\sum_{x=0}^{n} x\frac{n!}{x!\,(n-x)!}p^{x-1}\,p^1 q^{n-x}$$

$$=\sum_{x=0}^{n} x\frac{n\,(n-1)!}{x\,(x-1)!\,(n-x)!}p^{x-1}\,p^1 q^{n-x}$$

$$=np\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!\,(n-x)!}p^{x-1}\,q^{n-x}$$

$$=np\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!\,((n-1)-(x-1))!}p^{x-1}\,q^{(n-1)-(x-1)}$$

$$=np\sum_{x=1}^{n}(n-1)_{c_{(x-1)}}p^{x-1}q^{(n-1)-(x-1)}$$

$$=np(1)$$

$$\mu=E(x)=np$$

ii) Variance:

$$\sigma^2=E(x^2)-[E(x)]^2 \;----(1)$$

$$\Rightarrow E(x^2)=E(x(x-1)+x)$$

$$\Rightarrow E(x^2)=E\big(x(x-1)\big)+E(x)----(2)$$

$$\therefore E(x(x-1))=\sum_{x=0}^{n} x(x-1)p(x)$$

$$=\sum_{x=0}^{n} x(x-1)\,n_{c_x}p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\frac{n!}{x!\,(n-x)!}p^x\, q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\frac{n\,(n-1)(n-2)!}{x\,(x-1)(x-2)!\,(n-x)!}p^{x-2}\, p^2 q^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)!\,(n-x)!}p^{x-2}\, p^2 q^{n-x}$$

$$= n(n-1)p^2\sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!\,(n-x)!}p^{x-2}\, q^{n-x}$$

$$= n(n-1)p^2\sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!\,((n-2)-(x-2))!}p^{x-2}\, q^{(n-2)-(x-2)}$$

$$= n(n-1)p^2\sum_{x=2}^{n} (n-2)_{c_{(x-2)}} p^{x-2} q^{(n-2)-(x-2)}$$

$$= n(n-1)p^2(1)$$

$$\therefore E(x(x-1)) = n(n-1)p^2$$

$$(2) \Rightarrow E(x^2) = E\big(x(x-1)\big) + E(x)$$

$$\Rightarrow E(x^2) = n(n-1)p^2 + np$$

$$(1) \Rightarrow \sigma^2 = E(x^2) - [E(x)]^2$$

$$\Rightarrow \sigma^2 = n(n-1)p^2 + np - [np]^2$$

$$\Rightarrow \sigma^2 = n^2p^2 - np^2 + np - n^2p^2$$

$$\Rightarrow \sigma^2 = np - np^2$$

$$\Rightarrow \sigma^2 = np(1-p)$$

$$but \quad 1-p = q$$

$$\therefore \sigma^2 = npq$$

PROBLEMS

1) Let X be a binomially distributed random variable based on 6 repetitions of an experiment. If p=0.3, evaluate the following probabilities i) $P(x \le 3)$, ii) $P(X > 4)$.

Sol^n: Given p=0.3 and n=6, hence q = 1-p = 1-0.3 = 0.7

and $P(X = x) = b(6{,}0.3, x) = 6_{C_x}(0.3)^x(0.7)^{6-x}$

i) $P(x \le 3) = P(0) + P(1) + P(2) + P(3)$

$$= 6_{C_0}(0.3)^0(0.7)^6 + 6_{C_1}(0.3)^1(0.7)^5 + 6_{C_2}(0.3)^2(0.7)^4 + 6_{C_3}(0.3)^3(0.7)^3$$
$$= 0.1176 + 0.3025 + 0.3241 + 0.1852$$
$$= 0.9294$$

ii) $P(x \le 3) = P(5) + P(6)$

$$= 6_{C_5}(0.3)^5(0.7)^1 + 6_{C_6}(0.3)^6(0.7)^0$$
$$= 0.0102 + 0.0007$$
$$= 0.0109$$

2) The probability that a pen manufactured by a company will be defective is 0.1. If 12 such pens are selected at random, find the probability that
i)Exactly 2 pens will be defective
ii)Atmost 2 pens will be defective
iii)None will be defective

Sol^n: Let the probability that a pen manufactured is defective, p=0.1
then, q = 1-p = 1-0.1 = 0.9 and given n=12
Hence $P(X = x) = b(6,0.3, x) = 12_{C_x}(0.1)^x(0.9)^{12-x}$

i) The probability that exactly 2 pens will be defective, $P(2) = 12_{C_2}(0.1)^2(0.9)^{12-2}$
$= (66)(0.01)(0.3487)$
$= 0.2301$

ii) The probability that atmost 2 pens will be defective, $P(x \leq 2) = P(0) + P(1) + P(2)$
$= 12_{C_0}(0.1)^0(0.9)^{12} + 12_{C_1}(0.1)^1(0.9)^{11} + 12_{C_2}(0.1)^2(0.9)^{10}$
$= 0.2824 + 0.3766 + 0.2301$
$= 0.8891$

iii) The probability that none will be defective, $P(0) = 12_{C_0}(0.1)^0(0.9)^{12}$
$= (1)(1)(0.2824)$
$= 0.2824$

3) The number of telephonic lines busy at an instant is a binomial variant with a probability 0.1. If 10 lines are chosen at random, what is the probability that,
i)No line is busy
ii)All lines are busy
iii)Atleast one line is busy
iv)Atmost two lines are busy

Sol^n: Let the probability that a telephonic line is busy p=0.1
then q = 1-p = 1-0.1 = 0.9 and number of lines chosen is n = 10
Hence $P(X = x) = b(10,0.1, x) = 10_{C_x}(0.1)^x(0.9)^{10-x}$

i) The probability that no line is busy, $P(0) = 10_{C_0}(0.1)^0(0.9)^{10}$
$= (1)(1)(0.3487)$
$= 0.3487$

ii) The probability that all lines are busy, $P(10) = 10_{C_{10}}(0.1)^{10}(0.9)^0$
$= (1)(10^{-10})(1)$
$= 10^{-10}$

iii) The probability that atleast one line is busy, $P(x \geq 1) = 1 - P(0)$
$= 1 - 10_{C_0}(0.1)^0(0.9)^{10}$
$= 1 - 0.3487$
$= 0.6513$

iv) The probability that atmost two lines are busy, $P(x \leq 2) = P(0) + P(1) + P(2)$
$= 10_{C_0}(0.1)^0(0.9)^{10} + 10_{C_1}(0.1)^1(0.9)^9 + 10_{C_2}(0.1)^2(0.9)^8$
$= 0.3487 + 0.3874 + 0.1937$
$= 0.92968$

4) When a coin is tossed 4 times, find the probability of getting i)Exactly one head, ii)Atmost three heads, iii)Atleast two heads.

Sol^n: The number o times a coin is tossed, n=4

Let x be the binomial variant getting head, p=0.5

then q = 1-p = 1-0.5 = 0.5

Hence $P(X=x) = b(4,0.5,x) = 4_{C_x}(0.5)^x(0.5)^{4-x}$

$= 4_{C_x}(0.5)^4 = 4_{C_x}(0.0625)$

i) The probability of getting exactly one head, $P(1) = 4_{C_1}(0.0625) = 4 \times 0.0625 = 0.25$

ii) The probability of getting atmost three heads, $P(x \leq 3) = 1 - P(4)$

$= 1 - 4_{C_4}(0.0625)$

$= 1 - 0.0625$

$= 0.9375$

iii) The probability of getting atleast two heads, , $P(x \geq 2) = P(2) + P(3) + P(4)$

$= 4_{C_2}(0.0625) + 4_{C_3}(0.0625) + 4_{C_4}(0.0625)$

$= 0.375 + 0.25 + 0.0625$

$= 0.6875$

5) The probability of germination of a seed in a packet of seeds is found to be 0.7. If 10 seeds are taken for experimenting on germination in a laboratory, find the probability that
i)8 seeds germinate
ii)Atleast 8 seeds germinate
iii)Atmost 8 seeds germinate

Sol^n: Let X be the binomial variant of seed germination.

Given the number of seeds taken for experimenting in laboratory, n=10

The probability of germination of a seed in a packet of seeds is, p=0.7

then q = 1-p = 1-0.7 = 0.3

Hence, $P(X=x) = 10_{C_x}(0.7)^x(0.3)^{10-x}$

i) The probability that exactly 8 seeds germinate, $P(8) = 10_{C_8}(0.7)^8(0.3)^2$

$= (45)(0.0576)(0.09)$

$= 0.2334$

ii) The probability that atleast 8 seeds germinate, $P(x \geq 8) = P(8) + P(9) + P(10)$

$= 10_{C_8}(0.7)^8(0.3)^2 + 10_{C_9}(0.7)^9(0.3)^1 + 10_{C_{10}}(0.7)^{10}(0.3)^0$

$= 0.2334 + 0.1210 + 0.0282$

$= 0.3826$

iii) The probability that atmost 8 seeds germinate, $P(x \leq 8) = 1 - \{P(9) + P(10)\}$

$= 1 - \{10_{C_9}(0.7)^9(0.3)^1 + 10_{C_{10}}(0.7)^{10}(0.3)^0\}$

$= 1 - \{0.1210 + 0.0282\}$

$= 0.8508$

6) A communication channel receives independent pulses at the rate of 12 pulses per micro second. The probability of transmission error is 0.001 for each micro second. Compute the probability of,
i)No error during a micro second
ii)1 error
iii)Atleast 1 error
iv)2 error
v)Atmost 2 error

Sol^n: Let X be the binomial variant of Transmission error.

Given the number of pulses per micro second, n=12

Let p be the probability of transmission error, p=0.001

then q = 1-p = 1-0.001 = 0.999

The pmf of binomial distribution is, $P(X=x)=P(x)=n_{C_x}p^x q^{n-x}$
$= 12_{C_x}(0.001)^x(0.999)^{12-x}$

i) The probability of no error during a micro second, $P(0)=12_{C_0}(0.001)^0(0.999)^{12}$
$=(1)(1)(0.9880)$
$=0.9880$

ii) The probability of only one error during a micro second, $P(1)=12_{C_1}(0.001)^1(0.999)^{11}$
$=(12)(0.001)(0.9890)$
$=0.01186$

iii) The probability of atleast one error during a micro second, $P(x\geq 1)=1-P(0)$
$=1-12_{C_0}(0.001)^0(0.999)^{12}$
$=1-0.9880$
$=0.0120$

iv) The probability of two error during a micro second, $P(2)=12_{C_2}(0.001)^2(0.999)^{10}$
$=(66)(0.000001)(0.9900)$
$=0.00006534$

v) The probability of atmost two error during a micro second, $P(x\leq 2)=P(0)+P(1)+P(2)$
$=12_{C_0}(0.001)^0(0.999)^{12}+12_{C_1}(0.001)^1(0.999)^{11}+12_{C_2}(0.001)^2(0.999)^{10}$
$=0.9880+0.01186+0.00006534$
$=0.999925$

7) In 800 families with 5 children each, how many family would be expected to have,
i)3 boys
ii)5 girls
iii)Atmost 2 girls
iv)Either 2 or 3 boys
by assuming probability for boys and girls to be equal.

Sol^n: The total number of families given is 800 and number of children per family is, n=5
Given the probability of boy or girl to born, p=0.5
then q = 1-p = 1-0.5 = 0.5

The pmf of binomial distribution is, $P(X=x)=P(x)=n_{C_x}p^x q^{n-x}=5_{C_x}(0.5)^x(0.5)^{5-x}$
$=5_{C_x}(0.5)^5=5_{C_x}(0.03125)$

i) The probability to have exactly 3 boys, $P(3)=5_{C_3}(0.03125)$
$=(10)(0.03125)$
$=0.3125$
$\therefore$ The total number of families may have exactly 3 boys $=800\times 0.3125=250.$

ii) The probability to have exactly 5 girls, $P(5)=5_{C_5}(0.03125)$
$=(1)(0.03125)$
$=0.03125$
$\therefore$ The total number of families may have exactly 5 girls, $=800\times 0.03125=25.$

iii) The probability to have atmost two girls, $P(x\leq 2)=P(0)+P(1)+P(2)$
$=5_{C_0}(0.03125)+5_{C_1}(0.03125)+5_{C_2}(0.03125)$
$=0.03125+0.15625+0.3125$
$=0.5$
$\therefore$ The total number of families may have atmost two girls, $=800\times 0.5=400.$

iv) The probability to have either 2 or 3 boys, $P(2 \le x \le 3) = P(2) + P(3)$

$$= 5_{C_2}(0.03125) + 5_{C_3}(0.03125)$$
$$= 0.03125 + 0.03125$$
$$= 0.0625$$

$\therefore$ The total number of families may have either 2 or 3 boys, $= 800 \times 0.0625 = 500$.

Poisson Distribution:

Let X be the discrete random variable for any real value λ , such that the probability mass function of poisson distribution can be defined as,

$$P(X = x) = P(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & , x \ge 0 \\ 0, & otherwise \end{cases}$$

where, λ is called the parameter and,

i) $P(X = x) = P(x) \ge 0$

ii) $\sum_{x=0}^{n} P(x) = \sum_{x-0}^{n} \frac{e^{-\lambda}\lambda^x}{x!} = 1$

iii) Mean $\mu = np = \lambda$

iv) Vairance $\sigma^2 = \lambda$, S.D = $\sqrt{\lambda}$

The poisson distribution can be used to find the probabaility that an event might happen a definite number of times based on how often it usually occurs and the companies can utilize the poisson distribution to examine how they may be able to take steps to improve their operational effieciency.

MEAN & VARIANCE OF A POISSON DISTRIBUTION:

WKT, the probability mass function of the poisson distribution is,

$$P(X = x) = P(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & , \quad x \ge 0 \\ 0 & , \quad otherwise \end{cases}$$

i) Mean:

$$\mu = E(x) = \sum_{x=0}^{\infty} xP(x)$$
$$= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$$
$$= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^{x-1}\lambda}{x(x-1)!}$$
$$= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}$$
$$= \lambda(1)$$
$$\mu = \lambda$$

ii) Variance:

$$\sigma^2 = E(x^2) - \mu^2 \; ----(1)$$

$$= E(x(x-1)+x) - \mu^2$$

$$= E\big(x(x-1)\big) + E(x) - \mu^2 ----(2)$$

$$\therefore E(x(x-1)) = \sum_{x=0}^{\infty} x(x-1)P(x)$$

$$= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} x(x-1)\,\frac{e^{-\lambda}\lambda^{x-2}\lambda^2}{x(x-1)(x-2)!}$$

$$= \lambda^2 \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!}$$

$$= \lambda^2(1)$$

$$E(x(x-1)) = \lambda^2(1)$$

$$(2) \Rightarrow \sigma^2 = \lambda^2 + \lambda - \lambda^2$$

$$\Rightarrow \sigma^2 = \lambda$$

$$S.D = \sigma = \sqrt{\lambda}$$

$$Mean = \lambda = np$$

PROBLEMS

1) The number of accidents in a year to taxi drivers in a city follows a poisson distribution with mean 3. Out of 1000 taxi drivers find approximately the number of drivers with,
i)No accident in a year
ii)More than 3 accidents in a year.

Sol^n: Let X be the poisson variant follows accident in the year of the poisson distribution.

The probability mass function of the poisson distribution is $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Given the mean of poisson distribution is $\mu = \lambda = 3$

$P(X=x) = \frac{e^{-3}3^x}{x!}$

i) No accident in a year out of 1000 taxi drivers $= 1000 \times P(0)$

$$= 1000 \times \frac{e^{-3}3^0}{0!}$$
$$= 1000 \times 0.05$$
$$= 50$$

Hence 50 drivers out of 1000 having no accidents in a year.

ii) More than 3 accidents in a year out of 1000 taxi drivers $= 1000 \times P(x > 3)$

$$= 1000 \times [1 - P(x \le 3)]$$
$$= 1000 \times [1 - P(0) - P(1) - P(2]) - P(3)]$$
$$= 1000 \times \left[1 - \frac{e^{-3}3^0}{0!} - \frac{e^{-3}3^1}{1!} - \frac{e^{-3}3^2}{2!} - \frac{e^{-3}3^3}{3!}\right]$$
$$= 1000 \times \left(1 - \left(e^{-3} + e^{-3}(3) + e^{-3}\left(\frac{9}{2}\right) + e^{-3}\left(\frac{27}{6}\right)\right)\right)$$

$= 1000 \times (1 - 0.06472) = 352.8$
$= 353$

Therefore 353 drivers out of 1000 have done more than 3 accidents in the year.

2) In a certain factory turning out razor blades there is a small probability of $\frac{1}{500}$ for any blade to be defective. The blades are supplied in a packets of 10. Use poisson distribution to calculate approximate number of packets containing,
i)No defective
ii)2 defective
iii)3 defective
in the consignment of 10000 packets.

Sol^n: Let X be the poisson variant follows the blades to be defective of the poisson distribution.

The probability mass function of the poisson distribution is $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Given, p=$\frac{1}{500} = 0.002$, n=10, $\mu = np = 0.002 \times 10 = 0.02 = \lambda$

$$\therefore P(X = x) = \frac{e^{-0.02}(0.02)^x}{x!}$$

i) No blades are defective out 10000 packets $= 10000 \times P(x = 0)$

$$= 10000 \times \frac{e^{-0.02}(0.02)^0}{0!}$$
$$= 10000 \times 0.9802$$
$$= 9802$$

$\therefore$9802 packet blades are not defective out of 10000 packets.

ii) 2 defective blades out of 10000 packets $= 10000 \times P(x = 2)$

$$= 10000 \times \frac{e^{-0.02}(0.02)^2}{2!}$$
$$= 10000 \times 0.0002$$
$$= 2$$

$\therefore$2 packets blades are 2 defective out of 10000 packets.

iii) 3 defective blades out of 10000 packets $= 10000 \times P(x = 3)$

$$= 10000 \times \frac{e^{-0.02}(0.02)^3}{3!}$$
$$= 10000 \times 0.0000$$
$$= 0$$

$\therefore$No packets blades are 3 defective out of 10000 packets.

3) If the probability of bad reaction from a certain injection is 0.001, determine the probability that out of 2000 individuals more than 2 will get a bad reaction.

Sol^n: Let X be the poisson variant follows the bad reaction of the injection.

WKT, The probability mass function of the poisson distribution is $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Given, n=2000, p=0.001 and $\mu = np = 2000 \times 0.001 = 2 = \lambda$

$\therefore P(X = x) = \frac{e^{-2}(2)^x}{x!}$

The probability that of more than two individuals get bad reaction $= P(x > 2)$

$$= 1 - P(x \leq 2)$$
$$= 1 - [P(0) + P(1) + P(2)]$$
$$= 1 - \frac{e^{-2}(2)^0}{0!} - \frac{e^{-2}(2)^1}{1!} - \frac{e^{-2}(2)^2}{2!}$$
$$= 1 - \frac{5}{e^2} = 0.3233$$

4) The probability that a news reader commits no mistakes in reading the news is $\frac{1}{e^3}$. Find a probability on a particular news broadcast he commits,
i)Only 2 mistakes

ii)More than 3 mistakes
iii)Atmost 3 mistakes

Sol^n: Let X be the poisson variant follows the news reader do mistakes of the poisson distribution.

The probability mass function of the poisson distribution is $P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}$

Given, $P(X=0)=\frac{1}{e^3}$

$\frac{e^{-\lambda}\lambda^0}{0!}=\frac{1}{e^3}\Rightarrow\frac{1}{e^\lambda}=\frac{1}{e^3}\Rightarrow\lambda=3$

$$\therefore P(X=x)=\frac{e^{-3}(3)^x}{x!}$$

i) The probability that news reader can do 2 mistakes $=P(2)$

$$=\frac{e^{-3}(3)^2}{2!}$$
$$=0.2240$$

ii) The probability that the news reader can do more than 3 mistakes $P(x>3)=1-P(x\le 3)$

$$\Rightarrow P(x>3)=1-[P(0)+P(1)+P(2)+P(3)]$$
$$\Rightarrow P(x>3)=1-e^{-3}\left[\frac{3^0}{0!}+\frac{3^1}{1!}+\frac{3^2}{2!}+\frac{3^3}{3!}\right]$$
$$\Rightarrow P(x>3)=1-0.05(1+3+4.5+4.5)$$
$$\Rightarrow P(x>3)=1-0.65$$
$$\Rightarrow P(x>3)=0.3500$$

iii) The probability that the news reader can do atmost 3 mistakes $=P(x\le 3)$

$$\Rightarrow P(x\le 3)=P(x=0)+P(x=1)+P(x=2)+P(x=3)$$
$$\Rightarrow P(x\le 3)=e^{-3}\left[\frac{3^0}{0!}+\frac{3^1}{1!}+\frac{3^2}{2!}+\frac{3^3}{3!}\right]$$
$$\Rightarrow P(x\le 3)=(0.05)(1+3+4.5+4.5)$$
$$\Rightarrow P(x\le 3)=0.6500$$

5) Suppose 300 misprints are randomly distributed throughout a book of 500 pages, find the probability that a given page contains,
i)Exactly 3 misprints
ii)Less than 3 misprints
iii)4 or more misprints

Sol^n: Let X be the poisson variant of misprints throughout a book of 500 pages.

The probability mass function of the poisson distribution is $P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}$

Given, suppose 300 misprints are randomly distributed throughout a book of 500 pages.

$\therefore$Mean $\lambda=\frac{300}{500}=0.6$

WKT, the pmf of poisson distribution is $P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}$

i) The probability that exactly 3 misprints $=P(3)$

$$=\frac{e^{-0.6}(0.6)^3}{3!}$$
$$=0.01975$$

ii) The probability that there are less than three misprints $P(x<3)=P(0)+P(1)+P(2)$

$$\Rightarrow P(x<3)=\frac{e^{-0.6}(0.6)^0}{0!}+\frac{e^{-0.6}(0.6)^1}{1!}+\frac{e^{-0.6}(0.6)^2}{2!}$$
$$\Rightarrow P(x<3)=e^{-0.6}[1+0.6+0.18]$$
$$\Rightarrow P(x<3)=0.5488\times 1.78$$

$$\Rightarrow P(x < 3) = 0.9768$$

iii) The probability that there are 4 or more misprints, $P(x \geq 4) = 1 - P(x < 4)$

$$\Rightarrow P(x \geq 4) = 1 - P(0) - P(1) - P(2) - P(3)$$

$$\Rightarrow P(x \geq 4) = 1 - \frac{e^{-0.6}(0.6)^0}{0!} - \frac{e^{-0.6}(0.6)^1}{1!} - \frac{e^{-0.6}(0.6)^2}{2!} - \frac{e^{-0.6}(0.6)^3}{3!}$$

$$\Rightarrow P(x \geq 4) = 1 - e^{-0.6}(1 + 0.6 + 0.18 + 0.036)$$

$$= 1 - 0.5488 \times 1.816 = 0.00338$$

6) A certain screw making machine produces an average 2 defective out of 100 and packs of them in boxes of 500. Find the probability that the box contains,
i)3 defective
ii)Atleast 1 defective
iii)Between 2 & 4 defective

Sol^n: Given the machine producing an average defective screw is $p = \frac{2}{100} = 0.02$
also given, n=500, $\mu = np = 500 \times 0.02 = 10 = \lambda$

The probability mass function of the poisson distribution is $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$$\therefore P(X = x) = \frac{e^{-10}(10)^x}{x!}$$

i) The probability that exactly 3 defective $= P(3)$

$$= \frac{e^{-10}(10)^3}{3!}$$
$$= 0.007566$$

ii) The probability that atleast 1 screw is defective $= 1 - P(0)$

$$= 1 - \frac{e^{-10}(10)^0}{0!}$$
$$= 1 - 0.0000454$$
$$= 0.9999546$$

iii) The probability that between 2 & 4 screw will be defective $= P(2 \leq x \leq 4)$

$$= P(2) + P(3) + P(4)$$

$$= \frac{e^{-10}(10)^2}{2!} + \frac{e^{-10}(10)^3}{3!} + \frac{e^{-10}(10)^4}{4!}$$

$$= e^{-10}\left(\frac{100}{2} + \frac{1000}{6} + \frac{10000}{24}\right)$$

$$= e^{-10} \times 633.32$$

$$= 0.02875$$

Exponential Distribution:

Let X be a continuous random variable for any real value $\alpha > 0$, then the probability density function of an exponential distribution can be defined as, $P(X = x) = f(x) = \begin{cases} \alpha e^{-\alpha x}, & x > 0 \\ 0 \quad , & otherwise \end{cases}$, it follows:

i) $f(x) \geq 0$

ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

iii) $Mean\ \mu = \frac{1}{\alpha}$

iv) $Variance\ \sigma^2 = \frac{1}{\alpha^2}$

v) $S.D\ of\ Exponential\ Distribution, \sigma = \frac{1}{\alpha}$

PROBLEMS

1) If X is an Exponential variant with mean 3, then find $P(x > 1)\ \&\ P(x < 3)$.

Sol^n: Given X be a continuous random variable of an Exponential distribution is,

$$P(X = x) = f(x) = \begin{cases} \alpha e^{-\alpha x} & , for x \geq 0 \\ 0 & , Otherwise \end{cases}$$

and given the mean of exponential distribution is 3.

$\Rightarrow \mu = 3 \Rightarrow \frac{1}{\alpha} = 3 \Rightarrow \alpha = \frac{1}{3}$

$$\therefore f(x) = \begin{cases} \frac{1}{3} e^{-\frac{x}{3}} & for x \geq 0 \\ 0 & Otherwise \end{cases}$$

i) $P(x > 1) = \int_1^{\infty} f(x)dx = \frac{1}{3}\int_1^{\infty} e^{\frac{-x}{3}}\, dx = -\left[e^{\frac{-x}{3}}\right]_1^{\infty} = -\left[e^{-\infty} - e^{\frac{-1}{3}}\right] = -\left[0 - e^{\frac{-1}{3}}\right] = e^{\frac{-1}{3}}$

ii) $P(x < 3) = \int_{-\infty}^{3} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_0^3 f(x)dx = 0 + \frac{1}{3}\int_0^3 e^{\frac{-x}{3}}\, dx = -\left[e^{\frac{-x}{3}}\right]_0^3 = -[e^{-1} - e^0] = 1 - \frac{1}{e}$

2) If X is an exponential variant with mean 4, then find $P(0 < x < 1), P(x > 2)\ \&\ P(-\infty < x < 10)$.

Sol^n: Given X be a continuous random variable of an exponential distribution is,

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & , for x \geq 0 \\ 0 & , Otherwise \end{cases}$$

and given the mean of exponential distribution is 4.

$\Rightarrow \mu = 4 \Rightarrow \frac{1}{\alpha} = 4 \Rightarrow \alpha = \frac{1}{4}$

$$\therefore f(x) = \begin{cases} \frac{1}{4} e^{-\frac{x}{4}} & , x \geq 0 \\ 0 & , otherwise \end{cases}$$

$$P(0 < x < 1) = \int_0^1 fr(x)dx = \frac{1}{4}\int_0^1 e^{\frac{-x}{4}}\, dx = -\left[e^{\frac{-x}{4}}\right]_0^1 = -\left[e^{\frac{-1}{4}} - e^0\right] = 1 - \frac{1}{e^{\frac{1}{4}}}$$

$$P(x > 2) = \int_2^{\infty} fr(x)dx = \frac{1}{4}\int_2^{\infty} e^{\frac{-x}{4}}\, dx = -\left[e^{\frac{-x}{4}}\right]_2^{\infty} = -\left[e^{-\infty} - e^{\frac{-2}{4}}\right] = e^{\frac{-1}{2}}$$

$$P(-\infty < x < 10) = \int_{-\infty}^{10} fr(x)dx = \int_{-\infty}^{0} fr(x)dx + \int_0^{10} fr(x)dx = 0 + \frac{1}{4}\int_0^{10} e^{\frac{-x}{4}}\, dx$$

$$\Rightarrow P(-\infty < x < 10) = -\left[e^{\frac{-x}{4}}\right]_0^{10} = -\left[e^{\frac{-10}{4}} - e^0\right] = 1 - \frac{1}{e^{\frac{5}{2}}}$$

3) In a certain town the duration of shower has mean 5 minutes, what is the probability that shower will last for,

i)10 minutes and more

ii)Less than 10 minutes

iii)Between 10 & 12 minutes.

Sol^n: Given X be a continuous random variable of an exponential distribution is,

$$f(x)=\begin{cases}\alpha e^{-\alpha x} & ,x\ge 0\\ 0 & ,otherwise\end{cases}$$

and given the mean of exponential distribution is 5.

$\Rightarrow \mu=5\Rightarrow \frac{1}{\alpha}=5\Rightarrow \alpha=\frac{1}{5}$

$$\therefore f(x)=\begin{cases}\frac{1}{5}e^{-\frac{x}{5}} & ,x\ge 0\\ 0 & ,otherwise\end{cases}$$

i) The probability that the shower will last 10 minutes and more is,

$$P(x\ge 10)=\int_{10}^{\infty}f(x)dx=\int_{10}^{\infty}\frac{1}{5}e^{-\frac{x}{5}}dx=\frac{1}{5}\int_{10}^{\infty}e^{-\frac{x}{5}}dx=-\left[e^{-\frac{x}{5}}\right]_{10}^{\infty}=-[0-e^{-2}]=\frac{1}{e^2}$$

ii) The probability that the shower will last less than 10 minutes is,

$$P(x<10)=\int_{-\infty}^{10}f(x)dx=\int_{-\infty}^{0}f(x)dx+\int_{0}^{10}f(x)dx$$

$$\Rightarrow P(x<10)=0+\int_{0}^{10}\frac{1}{5}e^{-\frac{x}{5}}dx=\frac{1}{5}\int_{0}^{10}e^{-\frac{x}{5}}dx=-\left[e^{-\frac{x}{5}}\right]_{0}^{10}=-[e^{-2}-1]=1-\frac{1}{e^2}$$

iii) The probability that the shower will last between 10 & 12 minutes is,

$$P(10<x<12)=\int_{10}^{12}f(x)dx=\int_{10}^{12}\frac{1}{5}e^{-\frac{x}{5}}dx=\frac{1}{5}\int_{10}^{12}e^{-\frac{x}{5}}dx$$

$$\Rightarrow P(10<x<12)=-\left[e^{-\frac{x}{5}}\right]_{10}^{12}=-\left[e^{\frac{-12}{5}}-e^{-2}\right]=\frac{1}{e^{\frac{12}{5}}}-\frac{1}{e^2}$$

4) The life of a TV tube manufactured by a company is known to have mean 200 months. Assuming that the life of tube has an exponential distribution, find the probability that the life of a tube manufactured by a company is,
i)Less than 200 months
ii)Between 100 & 300 months
iii)More than 200 months

Sol^n: Given X be a continuous random variable of an exponential distribution is,

$$f(x)=\begin{cases}\alpha e^{-\alpha x} & ,x\ge 0\\ 0 & ,otherwise\end{cases}$$

and given the mean of exponential distribution is 200 months.

$\Rightarrow \mu=200\Rightarrow \frac{1}{\alpha}=200\Rightarrow \alpha=\frac{1}{200}$

$$\therefore f(x)=\begin{cases}\frac{1}{200}e^{-\frac{x}{200}} & ,x\ge 0\\ 0 & ,otherwise\end{cases}$$

i) The probability that the life of a tube is less than 200 months is,

$$P(x<200)=\int_{-\infty}^{200}f(x)dx=\int_{0}^{200}\frac{1}{200}e^{\frac{-x}{200}}dx=\frac{1}{200}\int_{0}^{200}e^{\frac{-x}{200}}dx=-\left[e^{\frac{-x}{200}}\right]_{0}^{200}=-[e^{-1}-e^{0}]=1-\frac{1}{e}$$

ii) The probability that the life of a tube is between 100 & 300 months is,

$$P(100\le x\le 300)=\int_{100}^{300}f(x)dx=\int_{100}^{300}\frac{1}{200}e^{-\frac{x}{200}}dx=\frac{1}{200}\int_{100}^{300}e^{-\frac{x}{200}}dx$$

$$\Rightarrow P(100 \le x \le 300) = -\left[e^{-\frac{x}{200}}\right]_{100}^{300} = -\left[e^{\frac{-3}{2}} - e^{\frac{-1}{2}}\right] = \frac{1}{e^{\frac{1}{2}}} - \frac{1}{e^{\frac{3}{2}}}$$

iii) The probability that the life of a tube is more than 200 months is,

$$P(x > 200) = \int_{200}^{\infty} f(x)dx = \int_{200}^{\infty} \frac{1}{200} e^{\frac{-x}{200}} dx$$

$$\Rightarrow P(x > 200) = \frac{1}{200}\int_{200}^{\infty} e^{\frac{-x}{200}} dx = -\left[e^{\frac{-x}{200}}\right]_{200}^{\infty} = -\left[e^{\frac{-\infty}{200}} - e^{-1}\right] = e^{-1} = \frac{1}{e}$$

5) The length of a telephone conversation is an exponential variant with mean 3 minutes. Find the probability that a call,
i)ends in less than 3 minutes
ii)ends between 3 & 5 minutes
iii)ends in more than 4 minutes

Sol^n: Given X be a continuous random variable of an exponential distribution is,

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & , x \ge 0 \\ 0 & , otherwise \end{cases}$$

and given the mean of exponential distribution is 3 minutes.

$$\Rightarrow \mu = 3 \Rightarrow \frac{1}{\alpha} = 3 \Rightarrow \alpha = \frac{1}{3}$$

$$\therefore f(x) = \begin{cases} \frac{1}{3} e^{-\frac{x}{3}} & , x \ge 0 \\ 0 & , otherwise \end{cases}$$

i) The probability that the conversation ends in less than 3 minutes is,

$$P(x < 3) = \int_{-\infty}^{3} f(x)dx = \int_{0}^{3} \frac{1}{3} e^{\frac{-x}{3}} dx = \frac{1}{3}\int_{0}^{3} e^{\frac{-x}{3}} dx = -\left[e^{\frac{-x}{3}}\right]_{0}^{3} = -[e^{-1} - e^{0}] = 1 - \frac{1}{e}$$

ii) The probability that the conversation ends in between 3 & 5 minutes is,

$$P(3 \le x \le 5) = \int_{3}^{5} f(x)dx = \int_{3}^{5} \frac{1}{3} e^{\frac{-x}{3}} dx = \frac{1}{3}\int_{3}^{5} e^{\frac{-x}{3}} dx$$

$$\Rightarrow P(100 \le x \le 300) = -\left[e^{-\frac{x}{3}}\right]_{3}^{5} = -\left[e^{\frac{-5}{3}} - e^{-1}\right] = \frac{1}{e} - \frac{1}{e^{\frac{5}{3}}}$$

iii) The probability that the conversation ends in more than 4 minutes is,

$$P(x > 4) = \int_{4}^{\infty} f(x)dx = \int_{4}^{\infty} \frac{1}{3} e^{\frac{-x}{3}} dx$$

$$\Rightarrow P(x > 4) = \frac{1}{3}\int_{4}^{\infty} e^{\frac{-x}{3}} dx = -\left[e^{\frac{-x}{3}}\right]_{4}^{\infty} = -\left[e^{\frac{-\infty}{3}} - e^{\frac{-4}{3}}\right] = e^{\frac{-4}{3}} = \frac{1}{e^{\frac{4}{3}}}$$

Normal Distribution:

Let X be a continuous random variable for any real μ *and* σ^2 , the normal distribution can be defined as,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} ,$$

where $-\infty \leq x \leq \infty$, $-\infty \leq \mu \leq \infty$ and here $\mu, \sigma^2(>0)$ are called the mean and variance of the normal distribution i.e., widely used in statistical inference, hypothesis testing, data analysis, i.e., to analysis the data when there is an equal chance for the data to be above and below the average value of the continuous data. The normal is also known as Gaussian distribution (or) Probability Bell Curve. The normal distribution is a probability distribution i.e., symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

The normal distribution follows as,

$\therefore P(a \leq x \leq b) = \int_a^b f(x)dx$

$$\Rightarrow P(a \leq x \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

$$\Rightarrow P(a \leq x \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left(\frac{x-\mu}{\sigma}\right)^2}{2}} dx$$

$$Let\ Z = \frac{x-\mu}{\sigma}$$

Where $z = \frac{x-\mu}{\sigma}$, $z_1 = \frac{a-\mu}{\sigma}$, $z_2 = \frac{b-\mu}{\sigma}$

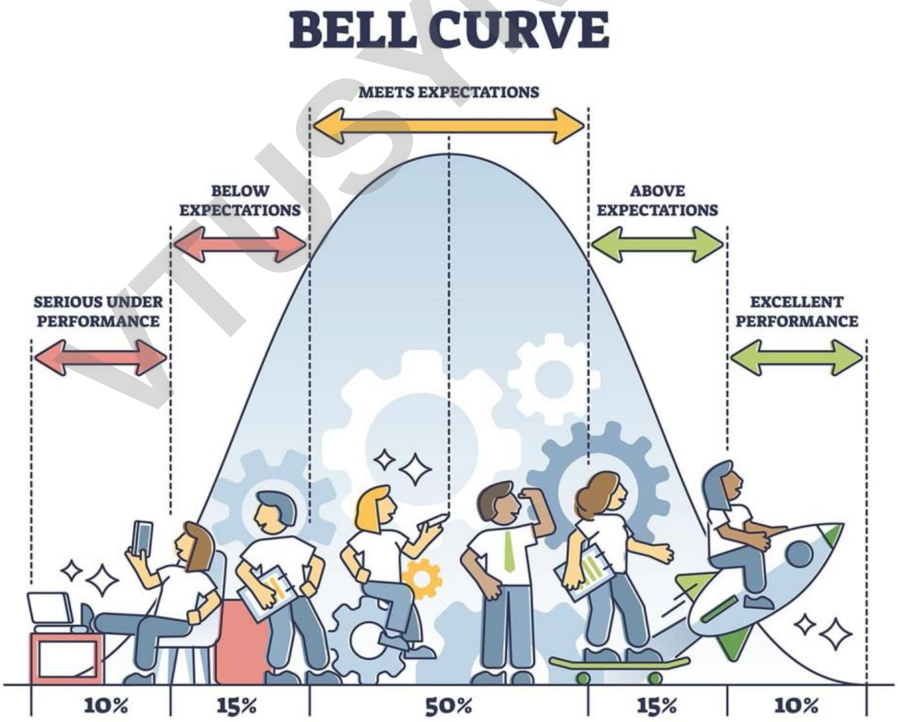
and $F(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ is called the standard normal function

and $z = \frac{x-\mu}{\sigma}$ is called the standard normal variate.

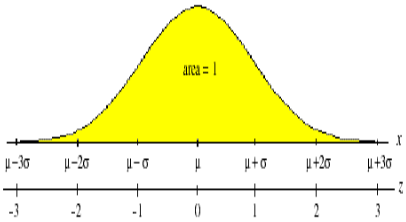
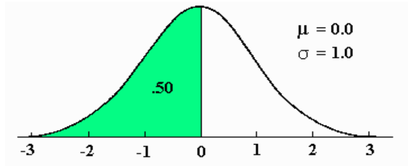
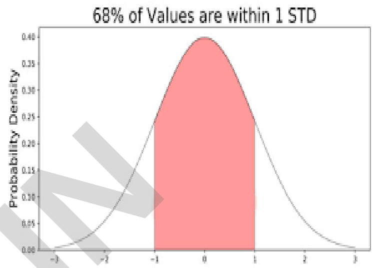
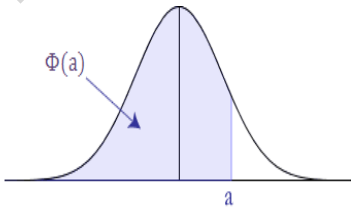
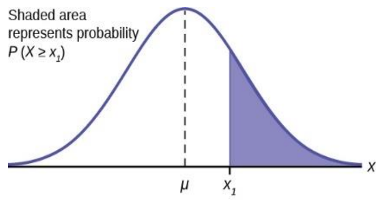
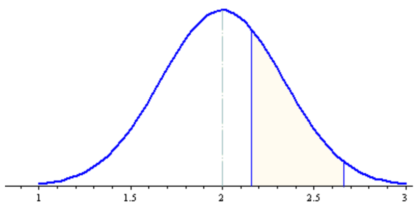
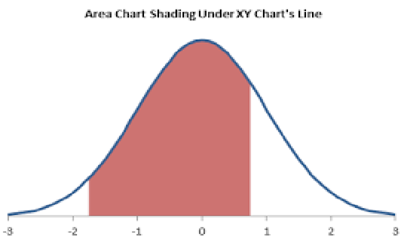
when $z_1 = 0$, $z_2 = z$, then the normal curve over 0 to z is defined as

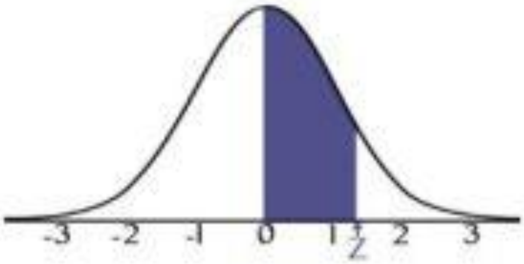
$$A(z) = \varphi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{z^2}{2}} dz$$

where these values will be taken from Area table of normal distribution.



Sl.No.	Probability Range	Result	Graph

<table>
<tr><td>1</td><td>$P(-\infty < z < \infty)$</td><td>1</td><td></td></tr>
<tr><td>2</td><td>$P(-\infty < z < 0) = P(0 < z < \infty)$</td><td>0.5</td><td></td></tr>
<tr><td>3</td><td>$P(-z_1 < z < z_1) = 2P(0 < z < z_1)$</td><td>$2A(z_1)$</td><td></td></tr>
<tr><td>4</td><td>$P(-\infty < z < z_1) = 0.5 + P(0 < z < z_1)$</td><td>0.5+$A(z_1)$</td><td></td></tr>
<tr><td>5</td><td>$P(z_1 < z < \infty) = P(0 < z < \infty) - P(0 < z < z_1)$</td><td>0.5-$A(z_1)$</td><td></td></tr>
<tr><td>6</td><td>$P(z_1 < z < z_2) = P(0 < z < z_2) - P(0 < z < z_1)$</td><td>$A(z_2)$ - $A(z_1)$</td><td></td></tr>
<tr><td>7</td><td>$P(-z_1 < z < z_2) = P(0 < z < z_2) + P(0 < z < z_1)$</td><td>$A(z_2)$ +$A(z_1)$</td><td></td></tr>
</table>



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	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
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

PROBLEMS

1) The marks of 1000 students in an examination follows normal distribution with mean 70 and standard deviation 5. Find the number students whose marks will be
i) Less than 65
i) More than 75
ii) Between 65 and 75. [A(1)=0.3413]

Sol.

Let X be the continuous random variable

Given

Mean of the Normal distribution $\mu = 70$

Standard deviation of the Normal distribution $\sigma = 5$

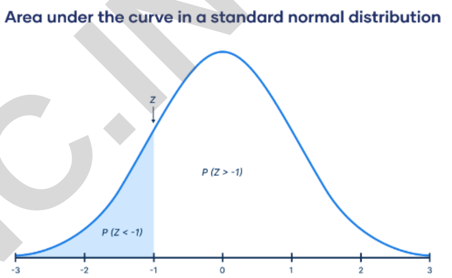
∴The standard normal variate $z = \frac{x-\mu}{\sigma} \Rightarrow z = \frac{x-70}{5}$

When x =65 then $z = \frac{65-70}{5} = -1$

When x =75 then $z = \frac{5-70}{5} = 1$

i) No.0f students scored less than 65 marks=$P(x < 65) = P(z < -1) =$
$P(z > 1) = 0.5 - A(1) = 0.5 - 0.3413 = 0.1587$

No.0f students scored less than 65 marks out of 1000 students=1000x0.1587=158.7=159

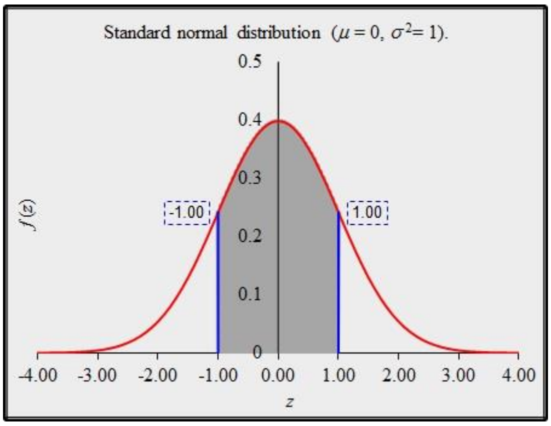


ii) No.0f students scored more than 75 marks=$P(x > 75) = P(z > 1) =$
$P(z > 1) = 0.5 - A(1) = 0.5 - 0.3413 = 0.1587$

No.0f students scored more than 75 marks out of 1000 students=1000x0.1587=158.7=159

iii) No.0f students scored marks between 65 and 75 =$P(65 < x < 75) = P(-1 < z < 1)$
$= 2P(0 < z < 1)$
$= 2A(1)$
$= 2 \times 0.3413$
$= 0.6826$

No.0f students scored between65 and 75 marks out of 1000 students=1000x0.6826
=682.6=683



2) 200 students appeared in an examination, distribution of marks is assumed to be normal with mean 30 and standard deviation 6.25, how many students are expected to get marks .
i) Between 20 and 40
ii) Less than 35 [A(1.6)=0.4452 , A(0.8)=0.2881]

Sol.

Let X be the continuous random variable

Given

Mean of the Normal distribution $\mu = 30$

Standard deviation of the Normal distribution $\sigma = 6.25$

∴The standard normal variate $z = \frac{x-\mu}{\sigma} \Rightarrow z = \frac{x-30}{6.25}$

When x =20 then $z = \frac{20-30}{6.25} = -1.6$

When x =40 then $z = \frac{40-30}{6.25} = 1.6$

When x =35 then $z = \frac{35-30}{6.25} = 0.8$

The probability that number of students expected to score between 20 and 40 marks:

$$P(20 < x < 40) = P(-1.6 < z < 1.6)$$
$$\Rightarrow P(20 < x < 40) = 2 \times P(0 < z < 1.6)$$
$$\Rightarrow P(20 < x < 40) = 2 \times A(1.6)$$
$$\Rightarrow P(20 < x < 40) = 2 \times 0.4452 = 0.8904 = 0.9$$

The number of students expected to score between 20 and 40 marks out of 200:

The probability that number of students expected to score less than 35:

$$= P(x < 35) = P(z < 0.8)$$
$$= A(0.8) = 0.2881 = 0.3$$

The number of students expected to score less than 35 marks out of 200:

=200x0.3=60

3) The weekly wages of workers in a company are normally distributed with mean of Rs.700 and S.D. of Rs.50.Find the probability that the weekly wage of randomly chosen workers is i) Between Rs.650 and Rs.750 ii) More than Rs.750.

Sol.

Let X be the continuous random variable

Given

Mean of the Normal distribution $\mu = 700$

Standard deviation of the Normal distribution $\sigma = 50$

∴The standard normal variate $z = \frac{x-\mu}{\sigma} \Rightarrow z = \frac{x-700}{50}$

When x =650 then $z = \frac{650-700}{50} = -1$

When x =750 then $z = \frac{750-700}{50} = 1$

The probability of the weekly wages between Rs.659 and Rs.750 is:

$$P(650 < x < 750) = P(-1 < z < 1)$$
$$\Rightarrow P(650 < x < 750) = 2 \times P(0 < z < 1)$$
$$\Rightarrow P(650 < x < 750) = 2 \times A(1)$$
$$\Rightarrow P(650 < x < 750) = 2 \times 0.3413 = 0.6826$$

The probability of the weekly wages of more than Rs.750 is:

$$= P(x > 750) = P(z > 1)$$
$$= 0.5 - P(z < 1)$$
$$= 0.5 - 0.3413$$
$$= 0.17065$$

4) In a test on 2000 electric bulbs , it was found that the life of a particular make was normally distributed with an average life of 2040 hours and SD of 60 hours . Estimate the number of bulbs likely to burn for...

i) More than 2150 hours

ii) Less than 1950 hours

iii) Between 1920 and 2160 hours.

Sol.

Let X be the continuous random variable

Given

Mean of the Normal distribution $\mu = 2040$

Standard deviation of the Normal distribution $\sigma = 60$

$\therefore$The standard normal variate $z = \frac{x-\mu}{\sigma} \Rightarrow z = \frac{x-2040}{60}$

When x =2150 then $z = \frac{2150-2040}{60} = 1.83$

When x =1950 then $z = \frac{1950-2040}{60} = -1.5$

When x =1920 then $z = \frac{1920-2040}{60} = -2$

When x =2160 then $z = \frac{2160-2040}{60} = 2$

i) The probability that the number of bulbs likely to burn of more than 2150 hours:

$$P(x > 2150) = P(z > 1.83) =$$

$$P(z > 1.83) = 0.5 - A(1.83) = 0.5 - 0.4664 = 0.0336$$

The number of bulbs likely to burn of more than 2150 hours out of 2000 bulbs :

=2000x0.0336

=67.2=67

ii) The probability that the number of bulbs likely to burn of less than 1950 hours:

$$=P(x < 1950) = P(z < -1.5) =$$

$$P(z > 1.5) = 0.5 - A(1.5) = 0.5 - 0.4332 = 0.0668$$

The number of bulbs likely to burn of less than 1950 hours out of 2000 bulbs :

=2000x0.0668

=133.6=137

iii) The probability that the number of bulbs likely to burn between 1920 and 2160 hours

$$=P(1920 < x < 2160) = P(-2 < z < 2)$$

$$= 2P(0 < z < 2)$$

$$= 2A(2)$$

$$= 2 \times 0.4772$$

$$= 0.9544$$

The number of bulbs likely to burn between 1920 and 2160 hours out of 2000 bulbs :

=2000x0.9544

=1908.8=1909

5) If the life time of a certain types electric bulbs of a particular brand was distributed normally with an average life of 2000 hours and S.D.60 hours. If a firm purchases 2500 bulbs, find the number of bulbs that are likely to last for (i) more than 2100 hours

(ii) less than 1950 hours

(iii) between 1900 and 2100 hours.

Sol.

Let X be the continuous random variable

Given

Mean of the Normal distribution $\mu = 2000$

Standard deviation of the Normal distribution $\sigma = 60$

$\therefore$ The standard normal variate $z = \frac{x-\mu}{\sigma} \Rightarrow z = \frac{x-2000}{60}$

When x =1950 then $z = \frac{1950-2000}{60} = -0.83$

When x=1900 then $z = \dfrac{1900 - 2000}{60} = -1.66$

When x=2100 then $z = \dfrac{2100 - 2000}{60} = 1.66$

i) The probability that the number of bulbs likely to burn of more than 2100 hours:

$P(x > 2100) = P(z > 1.66)$=$0.5 - A(1.66) = 0.5 - 0.4515 = 0.0485$

The number of bulbs likely to burn of more than 2100 hours out of 2500 bulbs:

=2500x0.0485

=121.25=121

ii) The probability that the number of bulbs likely to burn of less than 1950 hours:

=$P(x < 1950) = P(z < -0.83) = 0.5 - A(0.83) = 0.5 - 0.2967 = 0.2033$

The number of bulbs likely to burn of less than 1950 hours out of 2500 bulbs :

=2500x0.2033=508.25=508

iii) The probability that the number of bulbs likely to burn between 1900 and 2100 hours

$$= P(1900 < x < 2100) = P(-1.66 < z < 1.66)$$
$$= 2P(0 < z < 1.66)$$
$$= 2A(1.66)$$
$$= 2 \times 0.4515$$
$$= 0.9030$$

The number of bulbs likely to burn between 1900 and 2100 hours out of 2500 bulbs :

=2500x0.9030

=2257.5=2258

6) In a normal distribution , 7% of items are under 35 and 89% of the items are under 63.Find the mean and standard deviation of the distribution.

Sol.

Let X be the continuous random variable

Given

Let μ andσ be the Mean and Standard deviation of the distribution

∴The standard normal variate $z = \frac{x-\mu}{\sigma}$ --------------(1)

When x=35 the standard normal variate $z = \frac{35-\mu}{\sigma} = z_1(Say)$

When x=63 the standard normal variate $z = \frac{63-\mu}{\sigma} = z_2(Say)$

Given

$P(x < 35) = P(z < z_1) = 0.07$

$\Rightarrow P(z < z_1) = P(-\infty < z < 0) - P(0 < z < z_1) = 0.07$

$\Rightarrow 0.5 - A(z_1) = 0.07$

$\Rightarrow A(z_1) = 0.5 - 0.07$

$\Rightarrow A(z_1) = A(-1.47)$

$\Rightarrow z_1 = -1.47$

$$\Rightarrow \frac{35 - \mu}{\sigma} = -1.47$$

$\Rightarrow \mu - 1.47\sigma = 35 ------(2)$

And $P(x < 63) = P(z < z_2) = 0.89$

$\Rightarrow P(z < z_2) = P(-\infty < z < 0) + P(0 < z < z_2) = 0.89$

$\Rightarrow 0.5 + A(z_2) = 0.89$

$$\Rightarrow A(z_2) = 0.89 - 0.5 = 0.39$$
$$\Rightarrow A(z_2) = A(1.23)$$
$$\Rightarrow z_2 = 1.23$$
$$\Rightarrow \frac{63-\mu}{\sigma} = 1.23$$
$$\Rightarrow \mu + 1.23\sigma = 63 --------(3)$$

Solving eq(2) and (3)
we get

$$\mu = 50.2915$$
$$\sigma = 10.332$$

7) In a normal distribution , 31% of items are under 45 and 8% of the items are 0ver 64.Find the mean and standard deviation of the distribution.

Sol.

Let X be the continuous random variable
Given
Let μ andσ be the Mean and Standard deviation of the distribution
∴The standard normal variate $z = \frac{x-\mu}{\sigma}$ --------------(1)

When x=35 the standard normal variate $z = \frac{45-\mu}{\sigma} = z_1(Say)$
When x=63 the standard normal variate $z = \frac{64-\mu}{\sigma} = z_2(Say)$

Given

$$P(x < 35) = P(z < z_1) = 0.31$$
$$\Rightarrow P(z < z_1) = P(-\infty < z < 0) - P(0 < z < z_1) = 0.31$$
$$\Rightarrow 0.5 - A(z_1) = 0.31$$
$$\Rightarrow A(z_1) = 0.5 - 0.31 = 0.19$$

$$\Rightarrow A(z_1) = A(0.5)$$
$$\Rightarrow z_1 = 0.5$$
$$\Rightarrow \frac{45-\mu}{\sigma} = 0.5$$
$$\Rightarrow \mu + 0.5\sigma = 45 ------(2)$$

And $P(x > 64) = P(z > z_2) = 0.08$

$$\Rightarrow P(z > z_2) = 0.5 - P(0 < z < z_2) = 0.08$$
$$\Rightarrow 0.5 - A(z_2) = 0.08$$
$$\Rightarrow A(z_2) = 0.08 - 0.5 = -0.42$$
$$\Rightarrow A(z_2) = A(1.4)$$
$$\Rightarrow z_2 = 1.4$$
$$\Rightarrow \frac{64-\mu}{\sigma} = 1.4$$
$$\Rightarrow \mu + 1..4\sigma = 64 --------(3)$$

Solving eq(2) and (3)
we get

$$\mu = 50$$
$$\sigma = 10$$

8) In an examination 7% of the students scored less than 35% of the marks and 89% of the students scored less than 60% of the marks. Find the mean and standard deviation if marks are normally distributed.

Sol.

Let X be the continuous random variable
Given

Let μ andσ be the Mean and Standard deviation of the distribution

∴The standard normal variate $z = \frac{x-\mu}{\sigma}$ --------------(1)

When x=35 the standard normal variate $z = \frac{35-\mu}{\sigma} = z_1(Say)$

When x=63 the standard normal variate $z = \frac{60-\mu}{\sigma} = z_2(Say)$

Given

$$P(x < 35) = P(z < z_1) = 0.07$$

$$\Rightarrow P(z < z_1) = P(-\infty < z < 0) - P(0 < z < z_1) = 0.07$$

$$\Rightarrow 0.5 - A(z_1) = 0.07$$

$$\Rightarrow A(z_1) = 0.5 - 0.07$$

$$\Rightarrow A(z_1) = A(-1.47)$$

$$\Rightarrow z_1 = -1.47$$

$$\Rightarrow \frac{35-\mu}{\sigma} = -1.47$$

$$\Rightarrow \mu - 1.47\sigma = 35 ------(2)$$

And $P(x < 60) = P(z < z_2) = 0.89$

$$\Rightarrow P(z < z_2) = P(-\infty < z < 0) + P(0 < z < z_2) = 0.89$$

$$\Rightarrow 0.5 + A(z_2) = 0.89$$

$$\Rightarrow A(z_2) = 0.89 - 0.5 = 0.39$$

$$\Rightarrow A(z_2) = A(1.23)$$

$$\Rightarrow z_2 = 1.23$$

$$\Rightarrow \frac{60-\mu}{\sigma} = 1.23$$

$$\Rightarrow \mu + 1.23\sigma = 60 --------(3)$$

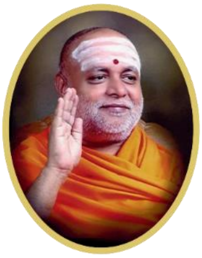
Solving eq(2) and (3)

we get

$\mu = 48.65$

$\sigma = 9.25$

*****



||Jai Sri Gurudev||

S.J.C. INSTITUTE OF TECHNOLOGY, CHICKBALLAPUR

DEPARTMENT OF MATHEMATICS

MATHEMATICS-3 FOR COMPUTER SCIENCE STREAM (BCS301)

MODULE - 2

JOINT DISTRIBUTION, STOCHASTIC PROCESS & MARKOV CHAIN

Prepared by:

Purushotham P
Assistant Professor
SJC Institute of Technology
Email id: purushotham@sjcit.ac.in

Joint Probability:

Let $X = \{x_1, x_2, x_3, \ldots\ldots. x_m\}$ and $Y = \{y_1, y_2, y_3, \ldots\ldots. y_n\}$ are two discrete random variables , then the joint probability function of X and Y is defined as

$$P(X = x_i, Y = y_j) = P(x_i, y_j) = f(x_i, y_j) = p_{ij} = f_{ij}$$

where the function $f(x, y)$ satisfy the conditions

i) $f(x, y) \geq 0$ ii) $\sum_i \sum_j f(x_i, y_j) = 1$

The joint probability table as shown below,

X \ Y	y_1	y_2	y_3	…..	y_n	$f(x_i)$
x_1	p_{11}	p_{12}	p_{13}	…..	p_{1n}	$f(x_1)$
x_2	p_{21}	p_{22}	p_{23}	…..	p_{2n}	$f(x_2)$
x_3	p_{31}	p_{32}	p_{33}	…..	p_{3n}	$f(x_3)$
. . . .	. . . .	. . . .	. . . .	….. . . .	. . . .	. . . .
x_m	p_{m1}	p_{m2}	p_{m3}	…..	p_{mn}	$f(x_m)$
$g(y_i)$	$g(y_1)$	$g(y_2)$	$g(y_3)$	…..	$g(y_n)$	1

Marginal Probability Distributions:

In the joint probability table $f(x_1), f(x_2), f(x_3), \dots\dots\dots, f(x_m)$ *and* $g(y_1), g(y_2), g(y_3), \dots\dots\dots, g(y_n)$ are called the marginal probability distributions respectively and represents the sum of all entries in all the rows and columns.

Independent Random Variables:

The discrete random variables X and Y are said to be independent if,

$P(X = x_i, Y = y_j) = P(X = x_i).P(Y = y_j)$, for every i, j and it is equivalent to

$P(X = x_i, Y = y_j) = P(X = x_i).P(Y = y_j) = f(x_i).g(y_j)$ **or** COV(X,Y) = 0

Expectation, Variance & Covariance:

Let X be the random variable taking the random values $x_1, x_2, x_3, \dots\dots. x_m$, having the probability function $f(x)$. Then,

a) The expectation of X is denoted by E(X) and is defined as, $\mu_X = E(X) = \sum_{i=1}^{m} x_i f(x_i)$

b) The expectation of Y is denoted by E(Y) and is defined as, $\mu_Y = E(Y) = \sum_{j=1}^{n} y_j f(y_j)$

c) The variance of X is denoted by ${\sigma_X}^2$ and is defined as ${\sigma_X}^2 = E(X^2) - [E(X)]^2$

$$\Rightarrow {\sigma_X}^2 = \sum_{i=1}^{m} {x_i}^2 f(x_i) - {\mu_X}^2$$

d) The variance of Y is denoted by ${\sigma_Y}^2$ and is defined as ${\sigma_Y}^2 = E(Y^2) - [E(Y)]^2$

$$\Rightarrow {\sigma_Y}^2 = \sum_{i=1}^{n} {y_j}^2 f(y_j) - {\mu_Y}^2$$

e) The covariance of X and Y is denoted by COV(X,Y) and defined as $COV(X,Y) = E(XY) - E(X).E(Y)$

$$COV(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j f(x_i, y_j) - \mu_X.\mu_Y$$

f) The correlation between X and Y is $\rho(X,Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$

PROBLEMS

1) The joint distribution of two random variables X and Y are as follows:

X \ Y	-4	2	7
1	$1/8$	$1/4$	$1/8$
5	$1/4$	$1/8$	$1/8$

Compute the following,

i) E(X) and E(Y)
ii) E(XY)
iii) σ_X & σ_Y
iv) $\rho(X,Y)$

Sol^n: Given,

$x_1 = 1,\ x_2 = 5,\ y_1 = -4,\ y_2 = 2,\ y_3 = 7$

And the probabilities are

$p_{11} = \frac{1}{8},\ p_{12} = \frac{1}{4},\ p_{13} = \frac{1}{8},\ p_{21} = \frac{1}{4},\ p_{22} = \frac{1}{8},\ p_{23} = \frac{1}{8}$

Given the joint probability distribution is follows as

Y \ X	-4	2	7	$f(x_i)$
1	$1/8$	$1/4$	$1/8$	$1/2$
5	$1/4$	$1/8$	$1/8$	$1/2$
$g(y_i)$	$3/8$	$3/8$	$1/4$	1

The marginal distribution of X and Y are

x_i	1	5
$f(x_i)$	$1/2$	$1/2$

y_i	-4	2	7
$g(y_i)$	$3/8$	$3/8$	$1/4$

i) $\mu_X = E(X) = \sum_{i=1}^{2} x_i f(x_i) = \left(1 \times \frac{1}{2}\right) + \left(5 \times \frac{1}{2}\right) = 3$

$\mu_Y = E(Y) = \sum_{j=1}^{3} y_j g(y_j) = \left(-4 \times \frac{3}{8}\right) + \left(2 \times \frac{3}{8}\right) + \left(7 \times \frac{1}{4}\right) = 1$

ii) $E(XY) = \sum_{i=1}^{2}\sum_{j=1}^{3} x_i y_j f(x_i, y_j)$

$$= \left(1 \times (-4) \times \frac{1}{8}\right) + \left(1 \times 2 \times \frac{1}{4}\right) + \left(1 \times 7 \times \frac{1}{8}\right) + \left(5 \times (-4) \times \frac{1}{4}\right) + \left(5 \times 2 \times \frac{1}{8}\right) + \left(5 \times 7 \times \frac{1}{8}\right)$$
$$= \frac{3}{2}$$

iii) $\sigma_X{}^2 = E(X^2) - \mu_X{}^2 = \sum_{i=1}^{2} x_i{}^2 f(x_i) - \mu_X{}^2 = \left(1^2 \times \frac{1}{2}\right) + \left(5^2 \times \frac{1}{2}\right) - 9 = 13 - 9 = 4 \Rightarrow \sigma_X = 2$

$\sigma_Y{}^2 = E(Y^2) - \mu_Y{}^2 = \sum_{j=1}^{3} y_j{}^2 g(y_j) - \mu_Y{}^2 = \left((-4)^2 \times \frac{3}{8}\right) + \left(2^2 \times \frac{3}{8}\right) + \left(7^2 \times \frac{1}{4}\right) - 1^2 = \frac{75}{4} \Rightarrow \sigma_Y = 4.33$

iv) $COV(X, Y) = E(XY) - \mu_X \mu_Y = \frac{3}{2} - (3)(1) = -\frac{3}{2}$

v) $\rho(X, Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = \frac{-3/2}{2 \times 4.33} = -0.1732$

Hence the given random variables are not independent

2) The joint distribution of two random variables X and Y are as follows:

Y \ X	-2	-1	4	5
1	0.1	0.2	0	0.3
2	0.2	0.1	0.1	0

Find the marginal distribution of X and Y. Also find the covariance of X and Y.

Sol^n: Given,

$x_1 = 1,\ x_2 = 2,\ y_1 = -2,\ y_2 = -1,\ y_3 = 4, y_4 = 5$

And the probabilities are

$p_{11} = 0.1,\ p_{12} = 0.2,\ p_{13} = 0,\ p_{14} = 0.3,\ p_{21} = 0.2,\ p_{22} = 0.1,\ p_{23} = 0.1, p_{13} = 0$

Given the joint probability distribution is follows as

Y \ X	-4	2	7		$f(x_i)$
1	0.1	0.2	0	0.3	0.6
2	0.2	0.1	0.1	0	0.4
$g(y_i)$	0.3	0.3	0.2	0.3	1

The marginal distribution of X and Y are

x_i	1	2
$f(x_i)$	0.6	0.4

y_i	-2	-1	4	5
$g(y_i)$	0.3	0.3	0.1	0.3

i) $\mu_X = E(X) = \sum_{i=1}^{2} x_i f(x_i) = (1 \times 0.6) + (2 \times 0.4) = 1.4$

$\mu_Y = E(Y) = \sum_{j=1}^{3} y_j g(y_j) = (-2 \times 0.3) + (-1 \times 0.3) + (4 \times 0.1) + (5 \times 0.3) = 1$

ii) $E(XY) = \sum_{i=1}^{2} \sum_{j=1}^{3} x_i y_j f(x_i, y_j)$

$= (1)(-2)(0.1) + (1)(-1)(0.2) + (1)(4)(0) + (1)(5)(0.3) + (2)(-2)(0.2) + (2)(-1)(0.1) + (2)(4)(0.1) + (2)(5)(0)$

$= -0.2 - 0.2 + 0 + 1.5 - 0.8 - 0.2 + 0.8 + 0 = 2.3 - 1.4 = 0.9$

iii) ${\sigma_X}^2 = E(X^2) - {\mu_X}^2 = \sum_{i=1}^{2} {x_i}^2 f(x_i) - {\mu_X}^2$

$= (1^2 \times 0.6) + (2^2 \times 0.4) - (1.4)^2 = 2.2 - 1.96 = 0.24 \Rightarrow \sigma_X = 0.4898$

${\sigma_Y}^2 = E(Y^2) - {\mu_Y}^2 = \sum_{j=1}^{3} {y_j}^2 g(y_j) - {\mu_Y}^2$

$= ((-2)^2 \times 0.3) + ((-1)^2 \times 0.3) + (4^2 \times 0.1) + (5^2 \times 0.3) - 1^2 = 9.6 \Rightarrow \sigma_Y = 3.0983$

iv) $COV(X, Y) = E(XY) - \mu_X \mu_Y = 0.9 - (1.4)(1) = -0.5$

vi) $\rho(X, Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = \frac{-0.5}{0.4898 \times 3.0983} = -0.3294$

Hence the given random variables are not independent

3) Determine,
i) Marginal distribution.
ii) Covariance between the discrete random variables X and Y, using the joint probability distribution.

X \ Y	3	4	5
2	$1/6$	$1/6$	$1/6$
5	$1/12$	$1/12$	$1/12$
7	$1/12$	$1/12$	$1/12$

Sol^n: Given,

$x_1 = 2, x_2 = 5, x_3 = 7, y_1 = 3, y_2 = 4, y_3 = 5$

And the probabilities are

$p_{11} = \frac{1}{6}, p_{12} = \frac{1}{6}, p_{13} = \frac{1}{6}, p_{21} = \frac{1}{12}, p_{22} = \frac{1}{12}, p_{23} = \frac{1}{12} p_{31} = \frac{1}{12}, p_{32} = \frac{1}{12}, p_{33} = \frac{1}{12}$

The joint distribution table is as follows:

X \ Y	3	4	5	$f(x_i)$
2	$1/6$	$1/6$	$1/6$	$1/2$
5	$1/12$	$1/12$	$1/12$	$1/4$
7	$1/12$	$1/12$	$1/12$	$1/4$
$g(y_i)$	$1/3$	$1/3$	$1/3$	1

The marginal distributions of X and Y are

x_i	2	5	7
$f(x_i)$	$1/2$	$1/4$	$1/4$

y_i	3	4	5
$g(y_i)$	$1/3$	$1/3$	$1/3$

$$\mu_X = E(X) = \sum_{i=1}^{3} x_i f(x_i)$$

$$\Rightarrow \mu_X = \left(2 \times \frac{1}{2}\right) + \left(5 \times \frac{1}{4}\right) + \left(7 \times \frac{1}{4}\right)$$

$$\Rightarrow \mu_X = 4$$

$$\mu_Y = E(Y) = \sum_{j=1}^{3} y_j g(y_j)$$

$$\Rightarrow \mu_Y = \left(3 \times \frac{1}{3}\right) + \left(4 \times \frac{1}{3}\right) + \left(5 \times \frac{1}{3}\right)$$

$$\Rightarrow \mu_Y = 4$$

$$E(XY) = \sum_{i=1}^{3} \sum_{j=1}^{3} x_i y_j f(x_i, y_j)$$

$$\Rightarrow E(XY) = \left(2 \times 3 \times \frac{1}{6}\right) + \left(2 \times 4 \times \frac{1}{6}\right) + \left(2 \times 5 \times \frac{1}{6}\right) + \left(5 \times 3 \times \frac{1}{12}\right) + \left(5 \times 4 \times \frac{1}{12}\right) + \left(5 \times 5 \times \frac{1}{12}\right)$$
$$+ \left(7 \times 3 \times \frac{1}{12}\right) + \left(7 \times 4 \times \frac{1}{12}\right) + \left(7 \times 5 \times \frac{1}{12}\right) = 16$$

$$\therefore Cov(X, Y) = E(XY) - \mu_X \mu_Y$$

$$\Rightarrow Cov(X, Y) = 16 - 4 \times 4$$

$$\Rightarrow Cov(X, Y) = 0$$

Hence the given random variables are independent

4) The joint probability distribution of discrete random variables X and Y is given below:

X \ Y	1	3	6
1	$1/9$	$1/6$	$1/18$
3	$1/6$	$1/4$	$1/12$
6	$1/18$	$1/12$	$1/36$

Determine,
i) Marginal distribution of X and Y.
ii) Are X and Y statistically independent?

Sol^n: Given,

$$x_1 = 1, x_2 = 3, x_3 = 6, y_1 = 1, y_2 = 3, y_3 = 6$$

And the probabilities are

$$p_{11} = \frac{1}{9}, p_{12} = \frac{1}{6}, p_{13} = \frac{1}{18}, p_{21} = \frac{1}{6}, p_{22} = \frac{1}{4}, p_{23} = \frac{1}{12}, p_{31} = \frac{1}{18}, p_{32} = \frac{1}{12}, p_{33} = \frac{1}{36}$$

The joint distribution table is as follows

X \ Y	1	3	6	$f(x_i)$
1	$1/9$	$1/6$	$1/18$	$1/3$
3	$1/6$	$1/4$	$1/12$	$1/2$
6	$1/18$	$1/12$	$1/36$	$3/18$
$g(y_i)$	$1/3$	$1/2$	$3/18$	1

i) The marginal distributions of X and Y are,

x_i	1	3	6
$f(x_i)$	$1/3$	$1/2$	$3/18$

y_i	1	3	6
$g(y_i)$	$1/3$	$1/2$	$3/18$

ii) $\mu_X = E(X) = \sum_{i=1}^{3} x_i f(x_i) = \frac{1}{3} + \frac{3}{2} + 1 = 2.8333$

$\mu_Y = E(Y) = \sum_{j=1}^{3} y_j g(y_j) = \frac{1}{3} + \frac{3}{2} + 1 = 2.8333$

$E(XY) = \sum_{i=1}^{3}\sum_{j=1}^{3} x_i y_j p_{ij} = \frac{1}{9} + \frac{3}{6} + \frac{6}{18} + \frac{3}{6} + \frac{9}{4} + \frac{18}{12} + \frac{6}{18} + \frac{18}{12} + \frac{36}{36} = 8.0278$

$COV(X,Y) = E(XY) - \mu_X\mu_Y = 8.0278 - (2.8333)(2.8333) = 8.0278 - 8.0276 = 0.0002$

$\therefore$ The given random variables X and Y are not statistically independent.

5) Determine,

i) Marginal distribution.

ii) Covariance between the discrete random variables X and Y along with corelation using the joint probability distribution.

X \ Y	1	3	9
2	$1/8$	$1/24$	$1/12$
4	$1/4$	$1/4$	0
6	$1/8$	$1/24$	$1/12$

Sol^n: Given

$x_1 = 2, x_2 = 4, x_3 = 6, y_1 = 1, y_2 = 3, y_3 = 9$

And the probabilities are

$p_{11} = \frac{1}{8}, p_{12} = \frac{1}{24}, p_{13} = \frac{1}{12}, p_{21} = \frac{1}{4}, p_{22} = \frac{1}{4}, p_{23} = 0, p_{31} = \frac{1}{8}, p_{32} = \frac{1}{24}, p_{33} = \frac{1}{12}$

The joint distribution table is as follows

X \ Y	1	3	9	$f(x_i)$
2	$1/8$	$1/24$	$1/12$	$1/4$
4	$1/4$	$1/4$	0	$1/2$
6	$1/8$	$1/24$	$1/12$	$1/4$
$g(y_i)$	$1/2$	$1/3$	$1/6$	1

The marginal distributions of X and Y are

x_i	2	4	6
$f(x_i)$	$1/4$	$1/2$	$1/4$

y_i	3	4	5
$g(y_i)$	$1/2$	$1/3$	$1/6$

$\mu_X = E(X) = \sum_{i=1}^{3} x_i f(x_i) = \left(2 \times \frac{1}{4}\right) + \left(4 \times \frac{1}{2}\right) + \left(6 \times \frac{1}{4}\right) = 0.5 + 2 + 1.5 = 4$

$\mu_Y = E(Y) = \sum_{j=1}^{3} y_j g(y_j) = \left(1 \times \frac{1}{2}\right) + \left(3 \times \frac{1}{3}\right) + \left(9 \times \frac{1}{6}\right) = 0.5 + 1 + 1.5 = 3$

$E(XY) = \sum_{i=1}^{3}\sum_{j=1}^{3} x_i y_j f(x_i, y_j)$

$= (1)(-2)(0.1) + (1)(-1)(0.2) + (1)(4)(0) + (1)(5)(0.3) + (2)(-2)(0.2) + (2)(-1)(0.1) + (2)(4)(0.1) + (2)(5)(0)$

$= -0.2 - 0.2 + 0 + 1.5 - 0.8 - 0.2 + 0.8 + 0 = 2.3 - 1.4 = 0.9$

$\sigma_X{}^2 = E(X^2) - \mu_X{}^2 = \sum_{i=1}^{3} x_i{}^2 f(x_i) - \mu_X{}^2$

$= \left(2^2 \times \frac{1}{4}\right) + \left(4^2 \times \frac{1}{2}\right) + \left(6^2 \times \frac{1}{4}\right) - 4^2 = 18 - 16 = 2 \Rightarrow \sigma_X = 1.4142$

$\sigma_Y{}^2 = E(Y^2) - \mu_Y{}^2 = \sum_{j=1}^{3} y_j{}^2 g(y_j) - \mu_Y{}^2$

$= \left(1^2 \times \frac{1}{2}\right) + \left(3^2 \times \frac{1}{3}\right) + \left(9^2 \times \frac{1}{6}\right) - 3^2 = 17 - 9 = 8 \Rightarrow \sigma_Y = 2.8284$

$COV(X,Y) = E(XY) - \mu_X\mu_Y = 12 - (4)(3) = 12 - 12 = 0$

$\rho(X,Y) = \frac{COV(X,Y)}{\sigma_X\sigma_Y} = \frac{0}{1.4142 \times 2.8284} = 0$

∴ The given random variables X and Y are not statistically independent.

6) Determine,

i) Marginal distribution.

ii) Covariance between the discrete random variables X and Y along with corelation using the joint probability distribution.

X \ Y	-3	2	4
1	0.1	0.2	0.2
2	0.3	0.1	0.1

Sol^n: Given

$x_1 = 1, x_2 = 2, y_1 = -3, y_2 = 2, y_3 = 4$

And the probabilities are

$p_{11} = 0.1, p_{12} = 0.2, p_{13} = 0.2, p_{21} = 0.3, p_{22} = 0.1, p_{23} = 0.1$

The joint distribution table is as follows

X \ Y	-3	2	4	$f(x_i)$
1	0.1	0.2	0.2	0.5
2	0.3	0.1	0.1	0.5
$g(y_i)$	0.4	0.3	0.3	1

The marginal distributions of X and Y are

x_i	1	2
$f(x_i)$	0.5	0.5

y_i	-3	2	4
$g(y_i)$	0.4	0.3	0.3

$\mu_X = E(X) = \sum_{i=1}^{2} x_i f(x_i) = (1 \times 0.5) + (2 \times 0.5) = 0.5 + 1 = 1.5$

$\mu_Y = E(Y) = \sum_{j=1}^{3} y_j g(y_j) = (-3 \times 0.4) + (2 \times 0.3) + (4 \times 0.3) = -1.2 + 0.6 + 1.2 = 0.6$

$$E(XY) = \sum_{i=1}^{2}\sum_{j=1}^{3} x_i y_j f(x_i, y_j)$$
$$= (1)(-3)(0.1) + (1)(2)(0.2) + (1)(4)(0.2) + (2)(-3)(0.3) + (2)(2)(0.1) + (2)(4)(0.1)$$
$$= -0.3 + 0.4 + 0.8 - 1.8 + 0.4 + 0 = 2.3 - 1.4 = 0.9$$

$$\sigma_X{}^2 = E(X^2) - \mu_X{}^2 = \sum_{i=1}^{2} x_i{}^2 f(x_i) - \mu_X{}^2$$
$$= (1^2 \times 0.5) + (2^2 \times 0.5) - 1.5^2 = 2.5 - 2.25 = 0.25 \Rightarrow \sigma_X = 0.5$$
$$\sigma_Y{}^2 = E(Y^2) - \mu_Y{}^2 = \sum_{j=1}^{3} y_j{}^2 g(y_j) - \mu_Y{}^2$$
$$= (-3^2 \times 0.4) + (2^2 \times 0.3) + (4^2 \times 0.3) - 0.6^2 = 9.6 - 0.36 = 9.24 \Rightarrow \sigma_Y = 3.0397$$

$$COV(X,Y) = E(XY) - \mu_X\mu_Y = 0.9 - (1.5)(0.6) = 0.9 - 0.9 = 0$$

$$\rho(X,Y) = \frac{COV(X,Y)}{\sigma_X\sigma_Y} = \frac{0}{0.5\times 3.0397} = 0$$

$\therefore$ The given random variables X and Y are not statistically independent.

7) X and Y are independent random variables. X takes the values 2,5 and 7 with probabilities $\frac{1}{2}, \frac{1}{4}$ and $\frac{1}{4}$ respectively. Y takes the values 3,4 and 5 with the probabilities $\frac{1}{3}, \frac{1}{3} \& \frac{1}{3}$.

a) Find the JPD of X and Y

b) Show that COV (X, Y) = 0

Sol^n:

Given X & Y are independent random variables follows the marginal probabilities as below.

x	2	5	7
$f(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

y	3	4	5
$g(y)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

The joint distribution table is as follows

X \ Y	3	4	5	$f(x_i)$
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
5	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$
7	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$
$g(y_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

$$\therefore \boldsymbol{\mu_X = E(X) = \sum x_i f(x_i) = \left(2\times\frac{1}{2}\right) + \left(5\times\frac{1}{4}\right) + \left(7\times\frac{1}{4}\right) = 4}$$
$$\therefore \boldsymbol{\mu_Y = E(X) = \sum y_j g(y_j) = \left(3\times\frac{1}{3}\right) + \left(4\times\frac{1}{3}\right) + \left(5\times\frac{1}{3}\right) = 4}$$
$$\therefore E(XY) = \sum_{i=1}^{2}\sum_{j=1}^{3} x_i y_j f(x_i, y_j) = 1 + \frac{8}{6} + \frac{10}{6} + \frac{15}{12} + \frac{20}{12} + \frac{25}{12} + \frac{21}{12} + \frac{28}{12} + \frac{35}{12} = \frac{192}{12} = 16$$
$$\therefore COV(X,Y) = E(XY) - \mu_X\mu_Y = 16 - (4)(4) = 16 - 16 = 0$$

Stochastic Process

Stochastic process consists of sequence of experiments in which each experiment has a finite number of outcomes with the given probabilities.

Probability Vector

A vector $V = [v_1, v_2, v_3, \ldots\ldots, v_n]$ is called the probability vector if each one of its components are non-negative and their sum is equal to unity or 1.

Ex: $= [0.1, 0.6, 0.3]$, $V = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$, etc...

Stochastic Matrix

A square matrix P is called a stochastic matrix if all the entries of P are non-negative and the sum of all the entries of any row is 1

(or)

A square matrix P is called a stochastic matrix where each row is in the form of the probability vector.

Ex: $= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$, $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$

Regular Stochastic Matrix

A matrix P is said to be a Regular Stochastic Matrix, if all the entries of some power (P^n) are positive. The Regular Stochastic Matrix P has a unique probability vector Q such that QP=Q and all the sum of the probabilities of a fixed vector matrix should be equal to 1.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nn} \end{bmatrix}$$

Transition Matrix

A transition matrix is also known as a stochastic or probability matrix, is a square matrix (n x n) representing the transition probabilities of a stochastic system.

Ex: $P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$

PROBLEMS

1) Verify that the matrix A $= \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \end{bmatrix}$ is a regular stochastic matrix.

Sol^n: Given matrix A , each element is nonnegative and the sum of the elements in each row is equal to 1.

∴A is stochastic matrix.

$$\text{Let } A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1/8 & 5/16 & 9/16 \\ 1/2 & 1/4 & 1/4 \end{bmatrix}$$

$$A^3 = A^2 \times A = \begin{bmatrix} 0 & 1 & 0 \\ 1/8 & 5/16 & 9/16 \\ 1/2 & 1/4 & 1/4 \end{bmatrix} . \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 5/32 & 41/64 & 13/64 \\ 1/8 & 5/16 & 9/16 \end{bmatrix}$$

∴Hence, all the entries in A^3 are nonnegative or positive and the sum of each row =1.

∴ The given matrix A is regular stochastic matrix.

2) Prove that the Markov Chain with Transition matrix $\mathbf{A}=\begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$ **is irreducible.**

Sol^n: Given matrix A is a stochastic matrix (Being a transition matrix).

Also, all the elements of given matrix does have non-negative and the sum of each row=1.

$$\therefore A^2=\begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}.\begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}=\begin{bmatrix} 1/2 & 1/6 & 1/3 \\ 1/4 & 7/12 & 1/16 \\ 1/4 & 1/3 & 5/12 \end{bmatrix}$$

∴Hence, all the entries in A^2 are nonnegative or positive and the sum of each row =1.

Hence the given transition matrix A is regular, consequently it follows that the given Markov Chain is irreducible.

3) Find the fixed probability vector for the regular stochastic matrix $\mathbf{A}=\begin{bmatrix} 1/3 & 2/3 \\ 1/4 & 3/4 \end{bmatrix}$.

Sol^n: Given $A=\begin{bmatrix} 1/3 & 2/3 \\ 1/4 & 3/4 \end{bmatrix}$

Since, the given matrix A is of second order.

Let $Q=[x \quad y]$ be the fixed probability vector, for every $x\geq 0, y\geq 0 \& x+y=1$

$$\therefore \boldsymbol{QA}=[x \quad y].\begin{bmatrix} 1/3 & 2/3 \\ 1/4 & 3/4 \end{bmatrix}=\left[\tfrac{1}{3}x+\tfrac{1}{4}y \quad \tfrac{2}{3}x+\tfrac{3}{4}y\right]$$

Since QA=Q

$$\Rightarrow \left[\tfrac{1}{3}x+\tfrac{1}{4}y \quad \tfrac{2}{3}x+\tfrac{3}{4}y\right]=[x \quad y]$$

$$\Rightarrow \frac{1}{3}x+\frac{1}{4}y=x, \frac{2}{3}x+\frac{3}{4}y=y$$

$$\Rightarrow \frac{2}{3}x=\frac{1}{4}y, \frac{2}{3}x=\frac{1}{4}y \dots (1), We have x+y=1 \Rightarrow y=1-x$$

$$\therefore (1) \Rightarrow \frac{2}{3}x=\frac{1}{4}(1-x)$$

$$\Rightarrow \frac{2}{3}x+\frac{1}{4}x=\frac{1}{4}$$

$$\Rightarrow \frac{2}{3}x+\frac{1}{4}x=\frac{1}{4}$$

$$\Rightarrow \frac{8x+3x}{12}=\frac{1}{4}$$

$$\Rightarrow \frac{11}{3}x=1 \Rightarrow x=\frac{3}{11}$$

$$\therefore y=1-x \Rightarrow y=1-\frac{3}{11} \Rightarrow y=\frac{8}{11}$$

Thus, the required fixed probability vector is $Q=[x \quad y]=\left[\frac{3}{11} \quad \frac{8}{11}\right]$

4) Find the fixed probability vector of the regular stochastic matrix $\mathbf{P}=\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$.

Sol^n: Since the given matrix P is of order 3x3, the required fixed probability vector Q must be also order of 3x3.

Let $Q=[x \quad y \quad z]$, $\quad For\ every\ x\geq 0, y\geq 0, z\geq 0 \& x+y+z=1$

Also, QP=Q

$$\therefore [x \quad y \quad z]\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0\end{bmatrix} = \begin{bmatrix}\frac{z}{2} & x+\frac{z}{2} & y\end{bmatrix}$$

$$\Rightarrow \begin{bmatrix}\frac{z}{2} & x+\frac{z}{2} & y\end{bmatrix} = [x \quad y \quad z]$$

$$\Rightarrow \frac{z}{2} = x, x + \frac{z}{2} = y, y = z$$

$$\Rightarrow \frac{1}{2}(1-x-y) = x.......(1), x + \frac{1}{2}(1-x-y) = y....(2), y = 1-x-y....(3)$$

$$\Rightarrow 3x + y = 1, x - 3y = -1, x + 2y = 1$$

$$\Rightarrow x = \frac{1}{5}, y = \frac{2}{5} \Rightarrow z = 1 - \frac{1}{5} - \frac{2}{5} = \frac{2}{5}$$

Hence the required fixed probability vector is $Q = [x \quad y \quad z] = \begin{bmatrix}\frac{1}{5} & \frac{2}{5} & \frac{2}{5}\end{bmatrix}$

5) Find the fixed probability vector of the regular stochastic matrix $\mathbf{P} = \begin{bmatrix}\frac{1}{2} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 1 & 0\end{bmatrix}$.

Sol^n:

Given, $P = \begin{bmatrix}\frac{1}{2} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 1 & 0\end{bmatrix}$

Since the given matrix P is of order 3x3, the required fixed probability vector Q must be also order of 3x3.

Let $Q = [x \quad y \quad z] For every x \geq 0, y \geq 0, z \geq 0 \& x + y + z = 1$

Also, QP=Q

$$\therefore QP = [x \quad y \quad z]\begin{bmatrix}\frac{1}{2} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 1 & 0\end{bmatrix}$$

$$\Rightarrow QP = \begin{bmatrix}\frac{1}{2}x + \frac{1}{2}y & \frac{1}{4}x + z & \frac{1}{4}x + \frac{1}{2}y\end{bmatrix}$$

WKT

$QP = Q$

$$\Rightarrow \begin{bmatrix}\frac{1}{2}x + \frac{1}{2}y & \frac{1}{4}x + z & \frac{1}{4}x + \frac{1}{2}y\end{bmatrix} = [x \quad y \quad z]$$

$$\Rightarrow x = \frac{x}{2} + \frac{y}{2}, y = \frac{x}{4} + z, z = \frac{1}{4}x + \frac{1}{2}y$$

$$\Rightarrow \frac{x}{2} + \frac{y}{2} = 0, \frac{x}{4} + (1-x-y) - y = 0, \frac{1}{4}x + \frac{1}{2}y = 1 - x - y$$

$$\Rightarrow x + y = 0, 3x + 8y = 4.....(1), 5x + 6y = 4......(2)$$

By solving eq (1) & (2)

$$\Rightarrow x = \frac{4}{11}, y = \frac{4}{11}, z = 1 - \frac{4}{11} - \frac{4}{11} = 1 - \frac{8}{11} = \frac{3}{11}$$

$$Q = [x \quad y \quad z] = \begin{bmatrix}\frac{4}{11} & \frac{4}{11} & \frac{3}{11}\end{bmatrix}$$

6) Find the fixed probability vector of the regular stochastic matrix $\mathbf{P} = \begin{bmatrix}0 & 1 & 0\\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3}\\ 0 & \frac{2}{3} & \frac{1}{3}\end{bmatrix}$.

Sol^n:

Given, $P = \begin{bmatrix}0 & 1 & 0\\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3}\\ 0 & \frac{2}{3} & \frac{1}{3}\end{bmatrix}$

Since the given matrix P is of order 3x3, the required fixed probability vector Q must be also order of 3x3.

Let $Q = [x \quad y \quad z] For every x \geq 0, y \geq 0, z \geq 0 \& x + y + z = 1$

Also, QP=Q

$$\therefore QP = [x \quad y \quad z]\begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

$$\Rightarrow QP = \begin{bmatrix} \frac{1}{6}y & x + \frac{1}{2}y + \frac{2}{3}z & \frac{1}{3}y + \frac{1}{3}z \end{bmatrix}$$

$WKT, \ QP = Q$

$\Rightarrow \begin{bmatrix} \frac{1}{6}y & x + \frac{1}{2}y + \frac{2}{3}z & \frac{1}{3}y + \frac{1}{3}z \end{bmatrix} = [x \quad y \quad z]$

$\Rightarrow x = \frac{1}{6}y, y = x + \frac{1}{2}y + \frac{2}{3}z, z = \frac{1}{3}y + \frac{1}{3}z$

$\Rightarrow 6x - y = 0, 2x - 7y = -4 \ldots. (1) 2x + 3y = 2 \ldots\ldots (2)$

By solving eq (1) & (2)

$$\Rightarrow x = \frac{1}{10}, y = \frac{6}{10}, z = 1 - \frac{1}{10} - \frac{6}{10} = 1 - \frac{7}{10} = \frac{3}{10}$$

$$\therefore Q = [x \quad y \quad z] = \begin{bmatrix} \frac{1}{10} & \frac{6}{10} & \frac{3}{10} \end{bmatrix}$$

7) Find the fixed probability vector of the regular stochastic matrix $P = \begin{bmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$.

Sol^n:

Given, $P = \begin{bmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$

Since the given matrix P is of order 3x3, the required fixed probability vector Q must be also order of 3x3.

Let $Q = [x \quad y \quad z] For every x \geq 0, y \geq 0, z \geq 0 \& x + y + z = 1$

Also, QP=Q

$$\therefore QP = [x \quad y \quad z]\begin{bmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$$\Rightarrow QP = \begin{bmatrix} \frac{1}{2}y + \frac{1}{2}z & \frac{2}{3}x + \frac{1}{2}z & \frac{1}{3}x + \frac{1}{2}y \end{bmatrix}$$

WKT

$QP = Q$

$\Rightarrow \begin{bmatrix} \frac{1}{2}y + \frac{1}{2}z & \frac{2}{3}x + \frac{1}{2}z & \frac{1}{3}x + \frac{1}{2}y \end{bmatrix} = [x \quad y \quad z]$

$\Rightarrow x = \frac{1}{2}y + \frac{1}{2}z, y = \frac{2}{3}x + \frac{1}{2}z, z = \frac{1}{3}x + \frac{1}{2}y$

$\Rightarrow 3x - 1 = 0, x - 9y = -3, 8x + 9y = 6$

$\Rightarrow x = \frac{9}{27}, y = \frac{10}{27}, z = \frac{8}{27}$

$\therefore Q[x \quad y \quad z] = \begin{bmatrix} \frac{9}{27} & \frac{10}{27} & \frac{8}{27} \end{bmatrix}$

8) If $P_1 = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$ and $P_2 = \begin{bmatrix} 1-b & b \\ a & 1-a \end{bmatrix}$. Show that P_1, P_2 and $P_1 P_2$ are stochastic matrices.

Sol^n: In P_1 we have $a + (1-a) = 1$ and $b + (1-b) = 1$

In P_2 we have $b + (1-b) = 1$ and $a + (1-a) = 1$

Thus, P_1 and P_2 are stochastic matrices.

Now, $\mathbf{P_1P_2} = \begin{bmatrix} 1-a & a \\ b & 1-\mathrm{b} \end{bmatrix}\begin{bmatrix} 1-b & b \\ a & 1-a \end{bmatrix}$

$$= \begin{bmatrix} (1-a)(1-b)+a^2 & b(1-a)+a(1-a) \\ b(1-b)+a(1-b) & (1-a)(1-b)+b^2 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} (say)$$

We shall know that $a_1 + b_1 = 1$ and $a_2 + b_2 = 1$

Now,

$$\begin{aligned} a_1 + b_1 &= (1-a)(1-b) + a^2 + b(1-a) + a(1-a) \\ &= (1-a)\{1-b+b\} + a\{a+1-a\} \\ &= 1-a+a \\ &= 1 \end{aligned}$$

$\therefore a_1 + b_1 = 1$

Also,

$$\begin{aligned} a_2 + b_2 &= b(1-b) + a(1-b) + (1-b)(1-a) + b^2 \\ &= b\{1-b+b\} + (1-b)\{a+1-a\} \\ &= b+1-b \\ &= 1 \end{aligned}$$

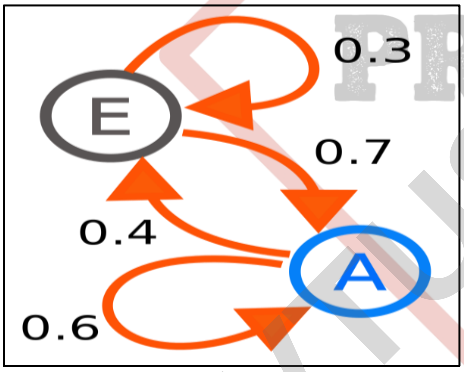
$\therefore a_2 + b_2 = 1$

Thus, $\mathbf{P_1P_2}$ is a stochastic matrix.

Markov Chain

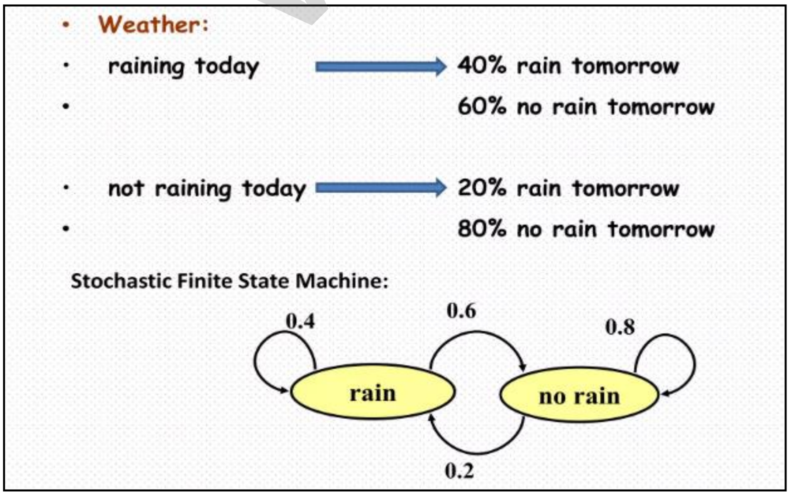
A Markov Chian or Markov process is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

Ex (1).



$$A = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0.7 & 0.3 \end{bmatrix}$$

Ex (2).



$$A = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix}$$

Higher transition probabilities:

Let P be a $n\times n$ transition probability matrix of the Markov chain with the probabilities $p_{ij}, 1\le i,j\le n$, is called changes from a_i to the state a_j, that is $a_i\to a_j$.The probabilities that the system changes from a_i to the state a_j in exists n steps is denoted by $p^{(n)}{}_{ij}$ and the matrix formed by the probabilities $p^{(n)}{}_{ij}$is called the n - step transition matrix , denoted by $P^{(n)}$ and initial probabilities are defined as,

$$p^{(0)}=[p_1{}^{(0)},p_2{}^{(0)},p_3{}^{(0)},p_4{}^{(0)}\ldots\ldots p_n{}^{(0)}]$$
$$p^{(1)}=[p_1{}^{(1)},p_2{}^{(1)},p_3{}^{(1)},p_4{}^{(1)}\ldots\ldots p_n{}^{(1)}]$$
$$p^{(2)}=[p_1{}^{(2)},p_2{}^{(2)},p_3{}^{(2)},p_4{}^{(2)}\ldots\ldots p_n{}^{(2)}]$$
$$-------------------$$
$$p^{(n)}=[p_1{}^{(n)},p_2{}^{(n)},p_3{}^{(n)},p_4{}^{(n)}\ldots\ldots p_n{}^{(n)}]$$

And $p^{(1)},p^{(2)},p^{(3)},p^{(4)}\ldots.$ will be evaluated as

$$p^{(1)}=p^{(0)}P,\ p^{(2)}=p^{(1)}P=p^{(0)}P^2\ ,\ p^{(3)}=p^{(2)}P=p^{(0)}P^3\ldots..\ p^{(n)}=p^{(n-1)}P=p^{(0)}P^n$$

1) Consider the $t.p.m.$ of the $P=\begin{array}{c} t\\ b\end{array}\begin{bmatrix}0 & 1\\ 1/2 & 1/2\end{bmatrix}$, hence find P^2 , P^3, also find $p^{(3)}$take the initial probability distribution the person rolled a die and decided that he will go by bus if the number appeared on the face is divisible by 3.

Sol^n: Given,

$$P=\begin{array}{c} t\\ b\end{array}\begin{bmatrix}0 & 1\\ 1/2 & 1/2\end{bmatrix}=\begin{bmatrix}p_{tt} & p_{tb}\\ p_{tb} & p_{bb}\end{bmatrix}$$

$$\therefore P^2=\begin{bmatrix}0 & 1\\ 1/2 & 1/2\end{bmatrix}\cdot\begin{bmatrix}0 & 1\\ 1/2 & 1/2\end{bmatrix}$$

$$\Rightarrow P^2=\begin{bmatrix}1/2 & 1/2\\ 1/4 & 3/4\end{bmatrix}=\begin{bmatrix}p^{(2)}{}_{tt} & p^{(2)}{}_{tb}\\ p^{(2)}{}_{tb} & p^{(2)}{}_{bb}\end{bmatrix}$$

$$\therefore P^3=P^2.P=\begin{bmatrix}1/2 & 1/2\\ 1/4 & 3/4\end{bmatrix}\cdot\begin{bmatrix}0 & 1\\ 1/2 & 1/2\end{bmatrix}$$

$$\Rightarrow P^3=\begin{bmatrix}1/4 & 3/4\\ 3/8 & 5/8\end{bmatrix}=\begin{bmatrix}p^{(3)}{}_{tt} & p^{(3)}{}_{tb}\\ p^{(3)}{}_{tb} & p^{(3)}{}_{bb}\end{bmatrix}$$

$\therefore p^{(2)}{}_{tb}=\frac{1}{2}$ Means that the probability that the system changes from the state $t\to b$ in exactly 2 steps is $\frac{1}{2}$.

$\therefore p^{(3)}{}_{bt}=\frac{3}{8}$ Means that the probability that the system changes from the state $b\to t$ in exactly 3 steps is $\frac{3}{8}$.

Given, the probability distribution is evaluated from the person rolled a die and decided that he will go by bus if the number appeared on the face is divisible by 3.

$$\therefore p(b)=\frac{2}{6}=\frac{1}{3}\ ,\Rightarrow p(t)=1-p(b)=1-\frac{1}{3}=\frac{2}{3}$$

$$p^{(0)}=[p(t)\quad p(b)]=\begin{bmatrix}\frac{2}{3} & \frac{1}{3}\end{bmatrix}$$

$$\therefore p^{(2)}=p^{(0)}P^2=\begin{bmatrix}\frac{2}{3} & \frac{1}{3}\end{bmatrix}\cdot\begin{bmatrix}1/2 & 1/2\\ 1/4 & 3/4\end{bmatrix}=\begin{bmatrix}\frac{5}{12} & \frac{7}{12}\end{bmatrix}$$

$$\therefore p^{(3)}=p^{(0)}P^3=\begin{bmatrix}\frac{2}{3} & \frac{1}{3}\end{bmatrix}\cdot\begin{bmatrix}1/4 & 3/4\\ 3/8 & 5/8\end{bmatrix}=\begin{bmatrix}\frac{7}{24} & \frac{17}{24}\end{bmatrix}=[p_t{}^{(3)}\quad p_b{}^{(3)}]$$

$\therefore$ The probability of travelling by train after 3 days $=\frac{7}{24}$

$\therefore$ The probability of travelling by bus after 3 days $=\frac{17}{24}$

2) The transition matrix P of a Markov chain is given by $\begin{bmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \end{bmatrix}$ with initial probability distribution $p^{(0)} = [1/4 \quad 3/4]$. Define and find the following i) $p_{21}{}^{(2)}$ ii) $p_{12}{}^{(2)}$ iii) $p^{(2)}$ iv) $p_1{}^{(2)}$ v) the vector $p^{(0)}P^n$ approaches. Vi) The matrix approaches.

Sol^n:

Given transition matrix $P = \begin{bmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \end{bmatrix}$

$$P^2 = P.P = \begin{bmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \end{bmatrix} . \begin{bmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \end{bmatrix}$$

$$\Rightarrow P^2 = \begin{bmatrix} 5/8 & 3/8 \\ 9/16 & 7/16 \end{bmatrix} = \begin{bmatrix} p_{11}{}^{(2)} & p_{12}{}^{(2)} \\ p_{21}{}^{(2)} & p_{22}{}^{(2)} \end{bmatrix}$$

$$\therefore p_{21}{}^{(2)} = \frac{9}{16}\ ,\ p_{12}{}^{(2)} = \frac{3}{8}$$

Given initial probability distribution is $p^{(0)} = [1/4 \quad 3/4]$

$$\therefore p^{(2)} = p^{(0)}P^2 = [1/4 \quad 3/4] . \begin{bmatrix} 5/8 & 3/8 \\ 9/16 & 7/16 \end{bmatrix}$$

$$\Rightarrow p^{(2)} = \begin{bmatrix} \frac{37}{64} & \frac{27}{64} \end{bmatrix} = [p_1{}^{(2)} \quad p_2{}^{(2)}]$$

$$\therefore p_1{}^{(2)} = \frac{37}{64}$$

$p^{(0)}P^n$ Approaches the unique probability vector $Q = [x \quad y]$ for which $QP = Q$

$$\Rightarrow [x \quad y] \begin{bmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \end{bmatrix} = [x \quad y]$$

$$\Rightarrow \begin{bmatrix} \frac{x}{2} + \frac{3y}{4} & \frac{x}{2} + \frac{y}{4} \end{bmatrix} = [x \quad y]$$

$$\Rightarrow \frac{x}{2} + \frac{3y}{4} = x, \frac{x}{2} + \frac{y}{4} = y$$

$$\Rightarrow \frac{x}{2} + \frac{3y}{4} = x, \frac{x}{2} + \frac{y}{4} = y$$

$$\Rightarrow -\frac{x}{2} + \frac{3(1-x)}{4} = 0$$

$$\Rightarrow -\frac{x}{2} - \frac{3x}{4} = -\frac{3}{4}$$

$$\Rightarrow \frac{5}{4}x = \frac{3}{4}$$

$$\Rightarrow x = \frac{3}{5} \Rightarrow y = \frac{2}{5}$$

$$\therefore Q[x \quad y] = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \end{bmatrix}$$

Therefore, the vector $p^{(0)}P^n$ approaches the vector $\begin{bmatrix} \frac{3}{5} & \frac{2}{5} \end{bmatrix}$

Therefore, the vector P^n approaches the matrix $\begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{3}{5} & \frac{2}{5} \end{bmatrix}$

3) The t.p.m. of a Markov chain is given by $P = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$ and the initial probability distribution is $p^{(0)} = \left(\frac{1}{2} \quad \frac{1}{2} \quad 0\right)$. Find $p_{13}{}^{(2)}$, $p_{23}{}^{(2)}$, $p^{(2)} p_1{}^{(2)}$.

Sol^n:

Given transition matrix of Markov chain $P = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$

And the initial probability distribution $p^{(0)} = \left(\frac{1}{2} \quad \frac{1}{2} \quad 0\right)$

$$\therefore P^2 = P.P = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix} . \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$$

$$\Rightarrow P^2 == \begin{bmatrix} 3/8 & 1/4 & 3/8 \\ 1/2 & 0 & 1/2 \\ 11/16 & 1/8 & 3/16 \end{bmatrix} = \begin{bmatrix} p^{(2)}{}_{11} & p^{(2)}{}_{12} & p^{(2)}{}_{13} \\ p^{(2)}{}_{21} & p^{(2)}{}_{22} & p^{(2)}{}_{23} \\ p^{(2)}{}_{31} & p^{(2)}{}_{32} & p^{(2)}{}_{33} \end{bmatrix}$$

$$\therefore p^{(2)}{}_{13} = 3/8, p^{(2)}{}_{23} = 1/2$$

$$\therefore p^{(2)} = p^{(0)} P^2 = \left(\frac{1}{2} \quad \frac{1}{2} \quad 0\right) . \begin{bmatrix} 3/8 & 1/4 & 3/8 \\ 1/2 & 0 & 1/2 \\ 11/16 & 1/8 & 3/16 \end{bmatrix}$$

$$\therefore p^{(2)} = [p_1{}^{(2)} \quad p_2{}^{(2)} \quad p_3{}^{(2)}] = \left[\frac{7}{16} \quad \frac{1}{8} \quad \frac{7}{16}\right]$$

$$\therefore p_1{}^{(2)} = \frac{7}{16}$$

4) Prove that the Markov chain whose t.p.m is $P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$ is irreducible. Find the corresponding stationary probability vector.

Sol^n:

Given transition matrix of Markov chain

$$P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

$$\Rightarrow P = \frac{1}{6} \begin{bmatrix} 0 & 4 & 2 \\ 3 & 0 & 3 \\ 3 & 3 & 0 \end{bmatrix}$$

$$\Rightarrow P^2 = P.P = \frac{1}{6} \begin{bmatrix} 0 & 4 & 2 \\ 3 & 0 & 3 \\ 3 & 3 & 0 \end{bmatrix} . \frac{1}{6} \begin{bmatrix} 0 & 4 & 2 \\ 3 & 0 & 3 \\ 3 & 3 & 0 \end{bmatrix}$$

$$\Rightarrow P^2 == \frac{1}{36} \begin{bmatrix} 18 & 6 & 12 \\ 9 & 21 & 6 \\ 9 & 12 & 15 \end{bmatrix}$$

Since all the entries of P^2 are non- negative, thus the given t.p.m P is regular and hence the Markov chain having t.p.m P is irreducible.

Let the unique probability vector $Q = [x \quad y \quad z]$ for which $QP = Q$, $\forall x + y + z = 1$

$$QP = [x \quad y \quad z]\begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} = [x \quad y \quad z]$$

$$\Rightarrow \left[\frac{y}{2}+\frac{z}{2} \quad \frac{2x}{3}+\frac{z}{2} \quad \frac{x}{3}+\frac{y}{2}\right] = [x \quad y \quad z]$$

$$\Rightarrow \frac{y}{2}+\frac{z}{2} = x, \frac{2x}{3}+\frac{z}{2} = y, \frac{x}{3}+\frac{y}{2} = z$$

$$\Rightarrow 2x - y - z = 0, 4x - 6y + 3z = 0, 2x + 3y - 6z = 0$$

$$\Rightarrow 2x - y - (1 - x - y) = 0, 4x - 6y + 3(1 - x - y) = 0$$

$$\Rightarrow 3x = 1 \& x - 9y = -3$$

$$\Rightarrow x = \frac{1}{3}, \frac{1}{3} - 9y = -3 \Rightarrow 9y = \frac{10}{3} \Rightarrow y = \frac{10}{27}$$

$$\therefore z = 1 - x - y = 1 - \frac{1}{3} - \frac{10}{27} = \frac{8}{27}$$

$\therefore Q[x \quad y \quad z] = \left[\frac{1}{3} \quad \frac{10}{27} \quad \frac{8}{27}\right]$ is the required stationary probability vector.

5) A student's study habits are as follows. If he studies one night, he is 30% sure to study the next night. On the other hand, if he does not study one night, he is 40% sure to study the next night. Find the transition matrix for the chain of his study.

Sol^n: We have two possible states

a_1= Studying a_2=Not studying

Therefore, given that

p_{11} =Probability of studying on night, given that he has studied in the previous night=30%=0.3

p_{12} =Probability of not studying on night, given that he has studied the previous night=70%=0.7

p_{21} =Probability of studying on night, given that he has not studied the previous night=40%=0.4

p_{22} =Probability of not studying on night, given that he has not studied the previous night=60%=0.6

Accordingly, the transition matrix of the chain of study is $P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix}$

Let the unique probability vector $Q = [x \quad y]$ for which $QP = Q$

$$\therefore [x \quad y]\begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix} = [x \quad y]$$

$$\Rightarrow [0.3x + 0.4y \quad 0.7x + 0.6y] = [x \quad y]$$

$$\Rightarrow 0.3x + 0.4y = x, 0.7x + 0.6y = y$$

$$\Rightarrow 0.7x - 0.4y = 0$$

$$\Rightarrow 0.7x - 0.4(1 - x) = 0$$

$$\Rightarrow 1.1x - 0.4 = 0$$

$$\Rightarrow x = \frac{0.4}{1.1} = \frac{4}{11} \Rightarrow y = 1 - \frac{4}{11} = \frac{7}{11}$$

$$\therefore Q[x \quad y] = \left[\frac{4}{11} \quad \frac{7}{11}\right] = [p_{a_1} p_{a_2}]$$

Thus, we conclude that in the long run the student will study $\frac{4}{11}$ of the time or 36.36 % of the time.

6) A software engineer goes to his work-place every day by motor bike or by car. He never goes by a bike on two consecutive days; but if he goes by car on a day then he is equally likely to go by car or bike on the next day. Find the transition matrix for the chain of the mode of transport he uses. If car is used on the first day of a week, find the probability that, (i) Bike is used, (ii) Car is used on the fifth day.

Sol^n: Given the Markov chain of the mode of transport has the following two states:

a_1= Using bike a_2=Using car

And to find,

p_{11} =Probability of using bike on a day, given that bike has been used on the previous day=0 (Because bike is not used on two consecutive days)

p_{12} =Probability of using car on a day, given that the bike has been used on the previous day=1 (Because it is certain that car is used on a day if bike is used on the previous day)

p_{21} =Probability of using bike on a day, given that car is used on the previous day= ½ (Because using car or bike on a day are equally likely if car is used on the previous day)

p_{22} =Probability of using car on a day, given that car is used on the previous day=1/2

Hence the transition matrix for the chain of the mode of transport is $P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

∴ The initial probability distribution vector of the mode of transport is given by $p^{(0)} = [p_1{}^{(0)} \quad p_2{}^{(0)}] = [0 \quad 1]$

$$\therefore P^2 = P.P = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} . \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$$

$$\Rightarrow P^2 = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$$

$$\Rightarrow P^4 = P^2.P^2 = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} . \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$$

$$\Rightarrow P^4 = \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix}$$

$$\therefore p^{(4)} = p^{(0)}P^4 = [0 \quad 1]. \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix}$$

$$\Rightarrow p^{(4)} = [p_1{}^{(4)} \quad p_2{}^{(4)}] = \begin{bmatrix} \frac{5}{16} & \frac{11}{16} \end{bmatrix}$$

Therefore, on the fifth day the probability of using the bike is $p_1{}^{(4)} = \frac{5}{16}$, the probability of using the car is $p_2{}^{(4)} = \frac{11}{16}$.

7) A man's smoking habits are as follows. If he smokes filter cigarettes one week, he switches to non-filter cigarettes the next week with the probability 0.2. On the other hand, if he smokes non filter cigarettes one week there is a probability of 0.7that he will smoke non filter cigarettes the next week as well. In the long run how often does he smoke filter cigarettes?

Sol^n:

Let A= Smoking filter cigarettes B= Smoking non filter cigarettes

Therefore, the associated transition probability matrix is as follows

$$P = \begin{bmatrix} p_{AA}{}^{(1)} & p_{AB}{}^{(1)} \\ p_{BA}{}^{(1)} & p_{BB}{}^{(1)} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

Let the unique probability vector $Q = [x \quad y]$ for which $QP = Q, \forall x + y = 1$

$$\therefore [x \quad y] \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} = [x \quad y]$$

$$\Rightarrow [0.8x + 0.3y \quad 0.2x + 0.7y] = [x \quad y]$$

$$\Rightarrow 0.8x + 0.3y = x, 0.2x + 0.7y = y$$

$$\Rightarrow 0.2x - 0.3y = 0, 0.2x - 0.3y = 0$$

$$\Rightarrow 0.2x - 0.3(1 - x) = 0$$

$$\Rightarrow 0.2x + 0.3x - 0.3 = 0$$

$$\Rightarrow 0.5x = 0.3$$

$$\Rightarrow x = \frac{0.3}{0.5} = \frac{3}{5} \Rightarrow y = \frac{0.2}{0.5} = \frac{2}{5}$$

$$\therefore Q = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \end{bmatrix} = [p_A \quad p_B]$$

Thus, in the long run, he will smoke filter cigarettes $\frac{3}{5}$ or 60% of the time.

(SMOKING IS INJURIOUS TO HEALTH, IT CAUSES CANCER AND TOBACCO CAUSES PAINFUL DEATH)

8) Three boys A, B, C are throwing ball to each other. "A" always throws the ball to "B" and "B" always throws ball to "C". "C" is just as likely to throw the ball to "B" as to "A". If, "C" was the first person to throw the ball, find the probabilities that after three throws.

i) A has the ball

ii) B has the ball

iii) C has the ball

Sol^n: Given three boys A, B, C are throwing a ball associated with the transition probability matrix of the Markov chain as below,

$$P = \begin{bmatrix} p_{AA}^{(1)} & p_{AB}^{(1)} & p_{AC}^{(1)} \\ p_{BA}^{(1)} & p_{BB}^{(1)} & p_{BC}^{(1)} \\ p_{CA}^{(1)} & p_{CB}^{(1)} & p_{CC}^{(1)} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

$$\Rightarrow P^2 = P.P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix} . \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

$$\Rightarrow P^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

$$\therefore P^3 = P^2.P = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix} . \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

$$\Rightarrow P^3 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

Initially if C has the ball, associated with the initial probability vector is given by $p^{(0)} = [0 \quad 0 \quad 1]$

$$\therefore p^{(3)} = p^{(0)}P^3 = [0 \quad 0 \quad 1] . \left[\begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}\right] = [1/4 \quad 1/4 \quad 1/2]$$

$$\therefore p^{(3)} = [p_A^{(3)} \quad p_B^{(3)} \quad p_C^{(3)}] = [1/4 \quad 1/4 \quad 1/2]$$

Thus, after three throws, the probability that the ball is with A is $p_A^{(3)} = \frac{1}{4}$, with B is $p_B^{(3)} = \frac{1}{4}$ and with C is $p_C^{(3)} = \frac{1}{2}$

9) A gambler's luck follows a pattern: if he wins a game, the probability of winning next game is 0.6. However, he loses the game, the probability of losing the next game is 0.7. There is an even chance of gambler winning the first game if so,

i) What is the probability of winning second game.

ii) What is the probability of winning the third game.

iii) In the long run, how often he will win.

Sol^n: Let W = Win the game , L = Lose the game

The transition probability matrix is given,

$$P = \begin{bmatrix} p_{ww} & p_{wl} \\ p_{lw} & p_{ll} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$$

And we know that the probability of winning and losing have the equal priority.

$\therefore$ The initial probability vector $p^{(0)} = [p_w^{(0)} \quad p_l^{(0)}] = [0.5 \quad 0.5]$

$$\therefore p^{(1)} = p^{(0)}P = [0.5 \quad 0.5] . \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix} = [0.45 \quad 0.55]$$

$\therefore p^{(2)} = p^{(1)}P = [0.45 \quad 0.55].[$

$\therefore p^{(3)} = p^{(2)}P = [0.435 \quad 0.565].\begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix} = [0.4305 \quad 0.5605]$

i) $\therefore {p_w}^{(2)} = 0.435 = 43.5\%$

ii) $\therefore {p_w}^{(3)} = 0.4305 = 43.05\%$

iii) Let Q=[x y] be the probability vector for which x+y=1

$\therefore QP = Q$

$\therefore [x \quad y].\begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix} = [x \quad y]$

$\Rightarrow 0.6x + 0.3y = x \;\;,\;\; 0.4x + 0.7y = y$

$\Rightarrow 0.4x - 0.3y = 0$

$\Rightarrow 0.4x - 0.3(1 - x) = 0$

$\Rightarrow 0.4x - 0.3 + 0.3x = 0$

$\Rightarrow 0.7x = 0.3$

$\Rightarrow x = \frac{3}{7}$

$\Rightarrow y = 1 - x \Rightarrow y = 1 - \frac{3}{7} \Rightarrow y = \frac{4}{7}$

$\therefore Q = [p_w \quad p_l] = \begin{bmatrix} \frac{3}{7} & \frac{4}{7} \end{bmatrix}$

10) A Salesman's territory consists of three cities A, B, C. He never sells in the same city on successive days. If he sells in city A then the next day he sells in city B. If he sells in B or C then the next day is twice as likely to sell in city A as than other cities. In long run, how often does he sells in each of the city.

Sol^n:

Given a salesman can move to the cities A, B, C with the probabilities as below,

$$P = \begin{array}{c} A \\ B \\ C \end{array}\begin{bmatrix} {p_{AA}}^{(1)} & {p_{AB}}^{(1)} & {p_{AC}}^{(1)} \\ {p_{BA}}^{(1)} & {p_{BB}}^{(1)} & {p_{BC}}^{(1)} \\ {p_{CA}}^{(1)} & {p_{CB}}^{(1)} & {p_{CC}}^{(1)} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 2/3 & 0 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix}$$

Let $Q = [x \quad y \quad z]$ be the probability vector for which x+y+z=1

$\therefore QP = Q$

$\therefore [x \quad y \quad z].\begin{bmatrix} 0 & 1 & 0 \\ 2/3 & 0 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix} = [x \quad y \quad z]$

$\Rightarrow \begin{bmatrix} \frac{2y}{3} + \frac{2z}{3} & x + \frac{z}{3} & \frac{y}{3} \end{bmatrix} = [x \quad y \quad z]$

$\Rightarrow \frac{2y}{3} + \frac{2z}{3} = x \;,\; x + \frac{z}{3} = y \;,\; \frac{y}{3} = z$

$\Rightarrow 3x - 2y - 2z = 0 \;\;,\;\; 3x - 3y + z = 0$

$\Rightarrow 3x - 2y - 2(1 - x - y) = 0 \;\;,\;\; 3x - 3y + (1 - x - y) = 0$

$\Rightarrow 3x - 2y - 2 + 2x + 2y = 0 \;\;,\;\; 3x - 3y + 1 - x - y = 0$

$\Rightarrow 5x = 2 \;\;,\;\; 2x - 4y = -1$

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

$\Rightarrow 4y = \frac{9}{5} \Rightarrow y = \frac{9}{20}$

$\Rightarrow z = 1 - x - y \Rightarrow z = 1 - \frac{2}{5} - \frac{9}{20} \Rightarrow z = \frac{3}{20}$

$\therefore Q = [x \quad y \quad z] = \begin{bmatrix} \frac{2}{5} & \frac{9}{20} & \frac{3}{20} \end{bmatrix}$

Thus, the salesman in the long run sells,

$\frac{2}{5}\ in\ city\ A = 40\%$, $\frac{9}{20}\ in\ city\ B = 45\%$, $\frac{3}{20}\ in\ city\ C = 15\%$

11) Every year, a man trades his car for a new car. If he has a Maruthi, he trades it for an Ambassador. If he has an Ambassador, he trades it for Santro. However if he had a Santro, he is just as likely to trade it for a Maruthi or an Ambassador. In 2000 he bought his first car which was a Santro. Find the probability that he has,

i) 2002 Santro

ii) 2002 Maruthi

iii) 2003 Ambassador

iv) 2003 Santro

Sol^n:

Given a man trades his car for a new car with the probabilities as below,

$$P = \begin{array}{c} M \\ A \\ S \end{array}\begin{bmatrix} p_{MM}{}^{(1)} & p_{MA}{}^{(1)} & p_{MS}{}^{(1)} \\ p_{AM}{}^{(1)} & p_{AA}{}^{(1)} & p_{AS}{}^{(1)} \\ p_{SM}{}^{(1)} & p_{SA}{}^{(1)} & p_{SS}{}^{(1)} \end{bmatrix} = \begin{array}{c} M \\ A \\ S \end{array}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

Also given, he has bought his first car in 2000 was Santro.

∴ The initial probability vector $p^{(0)} = [p_M^{(0)} \quad p_A^{(0)} \quad p_S^{(0)}] = [0 \quad 0 \quad 1]$

$$\therefore P^2 = P.P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix}.\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

$$\Rightarrow P^3 = P^2.P = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix}.\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

$$\therefore p^{(2)} = p^{(0)}P^2 = [0 \quad 0 \quad 1].\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix} = [0 \quad 1/2 \quad 1/2] = [p_M^{(2)} \quad p_A^{(2)} \quad p_S^{(2)}]$$

$$\therefore p^{(3)} = p^{(0)}P^3 = [0 \quad 0 \quad 1].\begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} = [1/4 \quad 1/4 \quad 1/2] = [p_M^{(3)} \quad p_A^{(3)} \quad p_S^{(3)}]$$

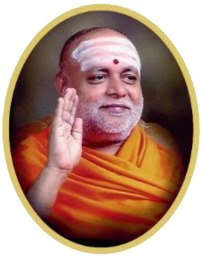
i) ∴ The probability to have a Santro car in the year 2002, $p_S^{(2)} = 1/2 = 50\%$

ii) ∴ The probability to have a Maruthi car in the year 2002,$p_M^{(2)} = 0 = 0\%$

iii) ∴ The probability to have an Ambassador car in the year 2003,$p_A^{(3)} = 1/4 = 25\%$

iv) ∴ The probability to have a Santro car in the year 2003,$p_S^{(3)} = 1/2 = 50\%$

***



||Jai Sri Gurudev||

S.J.C. INSTITUTE OF TECHNOLOGY, CHICKBALLAPUR

DEPARTMENT OF MATHEMATICS

MATHEMATICS-3 FOR COMPUTER SCIENCE (BCS301)

MODULE - 3

STATISTICAL INFERENCE -1

Prepared by:

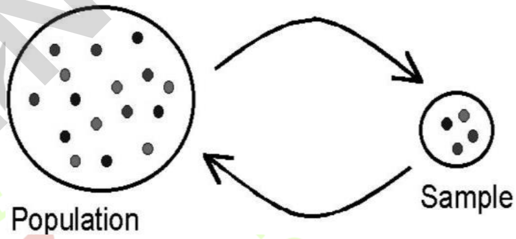
Purushotham P
Assistant Professor
SJC Institute of Technology
Email id: purushotham@sjcit.ac.in

Introduction:

Sampling is a statistical method of obtaining representative data (observations) from a group. We have been using sampling concepts in our day to day lives knowingly or unknowingly; for instance we take a handful of rice to check the rice quality of the full lot. This is an example of random sampling from a large population.

Population (Universe):

The group of objects (individuals) under study is called population or universe. Universe may be finite or infinite.

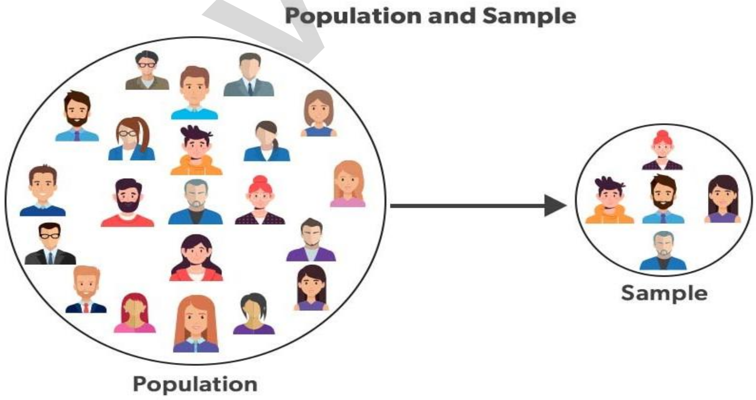


Sample:

A part containing objects(individuals), selected from the population is called a sample.

Sample size:

The number of individuals in a sample is called a sample size. If the sample size n is less than or equal to 30, then the sample is aid to be small, otherwise it is called a large sample.



Random Sampling:

The selection of objects (individuals) from the universe in such a way that each object (individual) of the universe has the same chance of being selected is called random sampling. Lottery system is the most common example of random sampling.

Every random sampling need not be simple. For example, if balls are drawn without replacement from a bag of balls containing different balls; the probability of success changes in every trial. Thus, the sampling though random is not simple.

Simple Sampling:

Simple sampling is a special case of random sampling in which each event has same probability of success or failure.

Hypothesis:

A hypothesis is an assumption based on insubstantial evidences that lends itself to further testing and experimentation. For example a farmer claims significant increase in crop production after using a particular fertilizer and after a season of experimenting, his hypothesis may be proved true or false. Any hypothesis may be accepted or rejected as per specific confidence levels and must be admissible to refutation.

Null Hypothesis:

The **null hypothesis** is a general statement or default position that there is no relationship between two measured phenomena or no association among groups.

Example: Given the test scores of two random samples, one of men and one of women, does one group differ from the other? A possible null hypothesis is that the mean male score is the same as the mean female score:

$H_0: \mu_1 = \mu_2$

where

H_0 = the null hypothesis,

μ_1 = the mean of population 1, and

μ_2 = the mean of population 2.

A stronger null hypothesis is that the two samples are drawn from the same population, such that the variances and shapes of the distributions are also equal.

Alternative Hypothesis:

It is the opposite statement of null hypothesis and denoted by $H_1: \mu_1 \neq \mu_2$

Significance levels (α):

The significance level of an event (such as a statistical test) is the probability that the event could have occurred by chance. If the level is quite low, that is, the probability of occurring by chance is quite small, we say the event is significant.

The level of significance is the measurement of the statistical significance. It defines whether the null hypothesis is assumed to be accepted or rejected. It is expected to identify if the result is statistically significant for the null hypothesis to be false or rejected.

$$\alpha = 5\% \quad \alpha = 1\% \quad \alpha = 0.27\%$$

Example: A level of significance of p=0.05 means that there is a 95% probability that the results found in the study are the result of a true relationship/difference between groups being compared. It also means that there is a 5% chance that the results were found by chance alone and no true relationship exists between groups.

Standard Error:

The standard deviation of the sampling distribution of a statistic is Known as Standard Error (S.E.).

Precision:

Reciprocal of standard error is known as precision.

Confidence Limits:

In short, confidence limits show how accurate an estimation of the mean is or is likely to be. Confidence limits are the lowest and the highest numbers at the end of a **confidence interval**.

Confidence Interval:

A confidence interval is a range around a measurement that conveys how precise the measurement is. A confidence interval, in statistics, refers to the probability that a population parameter will fall between a set of values for a certain proportion of times. Analysts often use confidence intervals that contain either 95% or 99% of expected observations.

Critical Value:

A critical value is the value of the test statistic which defines the upper and lower bounds of a confidence interval, or which defines the threshold of statistical significance in a statistical test.

	Level of Significance		
Types of test	*1%*	*5%*	*10%*
Two tailed test	2.58	1.96	1.645
One tailed test	2.33	1.645	1.28

Critical Region:

A critical region, also known as the rejection region, is a set of values for the test statistic for which the null hypothesis is rejected. i.e. if the observed test statistic is in the critical region then we reject the null hypothesis and accept the alternative hypothesis.

Type I and Type II Errors:

When we test a statistic at specified confidence level, there are chances of taking wrong decisions due to small sample size or sampling fluctuations etc.

Type I error is the incorrect rejection of a true null hypothesis, i.e. we reject H_0, when itis true, whereas Type II error is the incorrect acceptance of a false null hypothesis, i.e. we accept H_0 when it is false.

One Tailed and Two Tailed Tests:

While testing statistical significance levels; one- tailed test and a two-tailed test are used for accepting or rejecting a hypothesis. One- tailed tests are used for asymmetric distributions (reference value is unidirectional) whichhave a single tail; such as the chi-square distribution.

A two-tailed test is appropriate if the estimated value may lie on both sides of reference value. Two-tailed tests are only applicable when the probability curve has two tails; such as normal distribution.

Test of hypothesis:

Let x be the observed number of successes in a sample size of n and $\mu = np$ be the expected number of successes .Then the standard normal variate $Z$is defined as

$$Z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$$

Test of hypothesis for means:

Let μ_1 , μ_2 be the means, σ_1, σ_2 be the standard deviations of two populations and $\bar{x}_1, \bar{x}_2$ are the means of the samples, then

$Z = \frac{(\bar{x}_2 - \bar{x}_1)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}}$ If the samples are drawn from the same population, then $\sigma_1 = \sigma_2 = \sigma$ we have

$$Z = \frac{(\bar{x}_2 - \bar{x}_1)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

PROBLEMS

1) A coin is tossed 1000 times and head turns up 540 times. Decide on the hypothesis that the coin is unbiased at 1 % level of significance.

Sol.

Let us suppose that the coin is unbiased.
and let p= the probability of getting a head in one toss=1/2=0.5
Since p+q=1, q=1-p=1/2=0.5
Expected number of heads in 1000 tosses=np=1000x0.5=500 , npq=250
∴ The difference is $x - \mu$ =540-500=40
∴ Consider $Z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$
$\Rightarrow Z = \frac{40}{\sqrt{250}} = 2.53 < 2.58$
1% level of significance = 99% confidence level.
Therefore accept the hypothesis that the coin is unbiased.

2) A coin is tossed 400 times and turns up head 216 times. Test the hypothesis that the coin is unbiased at 5%level of significance.

Sol.

Let us suppose that the coin is unbiased.
and let p= the probability of getting a head in one toss=1/2=0.5
Since p+q=1, q=1-p=1/2=0.5

Expected number of heads in 400 tosses=np=400x0.5=200 , npq=100

∴ The difference is $x - \mu$ =216-200=16

∴ Consider $Z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$

$\Rightarrow Z = \frac{16}{\sqrt{100}} = \frac{16}{10} = 1.6 < 1.96$

Critical value of z at alpha = 0.05 is 1.96

Therefore accept the hypothesis that the coin is unbiased at the 5% level of significance.

3) A coin was tossed 1600 times and the tailed turned up 864 times. Test the hypothesis that the Coin is unbiased at 1% level of significance.

Sol.

Let us suppose that the coin is unbiased .

and let p= the probability of getting a tail in one toss=1/2=0.5

Since p+q=1, q=1-p=1/2=0.5

Expected number of tailed in 1600 tosses=np=1600x0.5=800, npq=400

∴ The difference is $x - \mu$ =864-800=64

∴ Consider $Z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$

$\Rightarrow Z = \frac{64}{\sqrt{400}} = 3.2 > 2.58$

1% level of significance = 99% confidence level.

Therefore accept the hypothesis that the coin is biased.

4) In 324 throws of a six faced 'die' , an odd number turned up 181 times. Is it possible to think that the 'die' is an unbiased one?

Sol.

Let us suppose that the die is unbiased.

and let p= the probability of the turn up of an odd number is=3/6=1/2=0.5

Since p+q=1, q=1-p=1/2=0.5

Expected number of successes=np=324x0.5=162, npq=81

∴ The difference is $x - \mu$ =181-162=19

∴ Consider $Z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$

$\Rightarrow Z = \frac{19}{\sqrt{81}} = \frac{19}{9} = 2.11 < 2.58$

Thus we can that the die is unbiased

5) A die is thrown 9000 times and a throw of 3 or 4 was observed 3240 times. Show that the die Can not be regarded as an unbiased one.

Sol.

The probability of getting 3 or 4 in a single through is $p = {}^{2}/_{6} = {}^{1}/_{3}$

And $q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

∴ Expected number of success $= \frac{1}{3} \times 9000 = 3000$

∴ The difference =3240-3000=240

$$Z = \frac{x-np}{\sqrt{npq}}$$

Consider $\Rightarrow Z = \frac{(3240)-\left(9000\times\frac{1}{3}\right)}{\sqrt{9000\times\frac{1}{3}\times\frac{2}{3}}}$

$$\Rightarrow Z = \frac{240}{\sqrt{2000}}$$

$$\Rightarrow Z = 5.37$$

Since Z=5.37>2.58 ,

We conclude that the die is biased.

Test of significance for proportion:

Test of significance of single proportion:

To test the significant difference between the sample proportion p and the population proportion P, we use the statistic.

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

The formulated Null and Alternative hypothesis is , $H_0; P = a$ specified value , $H_1: P \neq a$ specified value

Test of significance of Difference between two sample proportions:

To test the significance of the difference between the samples proportions, the test statistic under the null hypothesis H_0 that there is no significance difference between the two sample proportions,

We have $Z = \frac{p_1-p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} Where P = \frac{n_1p_1+n_2p_2}{n_1+n_2} or P = \frac{x_1+x_2}{n_1+n_2}$ and P+Q=1,

Here $p_1 and p_2$ are the sample proportions in respect of an attribute corresponding to two large samples of size $n_1 and n_2$ drawn from the two populations.

PROBLEMS

1. A coin is tossed 400 times and it turns up head 216 times. Discuss whether the coin may be regarded as unbiased one.

Sol.

Set the null hypothesis $H_0; P = \frac{1}{2}$

Set the Alternative hypothesis $H_1: P \neq \frac{1}{2}$

The level of significance $\alpha = 0.05$ (5%)

$\therefore$ The test statistic $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$, where P+Q=1 => Q=1-P

Given, the coin is tossed and it turns up in the equal proportion

$$P = \frac{1}{2} \Rightarrow Q = 1 - P$$

$$\Rightarrow Q = 1 - \frac{1}{2} = \frac{1}{2}$$

And the coin turns up head 216 times when it tossed $n = 400$times

$$\therefore p = \frac{216}{400} = 0.54$$

$$\therefore Z = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}}$$

$$\Rightarrow Z = \frac{0.04}{\sqrt{0.000625}} = 1.6$$

At 5% level, the tabulated value of $Z_\alpha$is 1.96

Since $|Z| = 1.6 < 1.96$

Hence, the null hypothesis is accepted at 5% level of significance and the coin may be regarded as unbiased.

2. In a city of sample of 500 people, 280 are tea drinkers and the rest are coffee drinkers. Can we assume that both coffee and tea are equally popular in this city at 5% Los.

Sol.

Set the null hypothesis $H_0; P = \frac{1}{2}$ (Both coffee and tea drinkers are equally popular)

Set the Alternative hypothesis $H_1: P \neq \frac{1}{2}$

The level of significance $\alpha = 0.05$ (5%)

$\therefore$ The test statistic $= \frac{p-P}{\sqrt{\frac{PQ}{n}}}$, where P+Q=1 => Q=1-P

$$P = \frac{1}{2} \Rightarrow Q = 1 - P$$

$$\Rightarrow Q = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\therefore p = \frac{280}{500} = 0.56 \text{ , where } n = 500$$

$$\therefore Z = \frac{0.56 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{500}}}$$

$$\Rightarrow Z = \frac{0.06}{\sqrt{0.0005}} = 2.68$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 2.68 > 1.96$

Hence, the null hypothesis is rejected at 5% level of significance and both the drinkers are not popular.

3. A manufacturing company claims that at least 95% of its products supplied confirm to the specifications out of a sample of 200 products, 18 are defective. Test the claim at 5% Los.

Sol.

Set the null hypothesis $H_0; P = 95\% = 0.95$

Set the Alternative hypothesis $H_1: P \neq 0.95$

The level of significance $\alpha = 0.05\ (5\%)$

$\therefore$ The test statistic $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$, where P+Q=1 => Q=1-P

Given,

$$P = 95\% = 0.95 \Rightarrow Q = 1 - P$$

$$\Rightarrow Q = 1 - 0.95 = 0.05$$

Found 18 products are defective out of 200 sample products

$\therefore$ The total defective less products(Non defective)=200-18=182

$$\therefore p = \frac{182}{200} = 0.91$$

$$\therefore Z = \frac{0.91 - 0.95}{\sqrt{\frac{0.95 \times 0.05}{200}}}$$

$$\Rightarrow Z = -\frac{0.04}{\sqrt{0.0002375}} = -2.5955$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 2.5955 > 1.96$

Hence, the null hypothesis is rejected at 5% level of significance

4. If a sample of 300 units of a manufactured product 65 units were found to be defective and in another sample of 200 units, there were 35 defectives. Is there significant difference in the proportion of defectives in the samples at 5% Los.

Sol.

Set the null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.05\ (5\%)$

Given

$$n_1 = 300, n_2 = 200$$

The sample of 300 units of a manufactured product 65 units were found to be defective

$$\therefore p_1 = \frac{65}{300} = 0.2166 = 0.22$$

The sample of 200 units of a manufactured product 35 units were found to be defective

$$\therefore p_2 = \frac{35}{200} = 0.1750$$

We know that $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{65+35}{300+200}$$

$$\Rightarrow P = \frac{100}{500}$$

$$\Rightarrow P = 0.2$$

$$\Rightarrow Q = 1 - P = 1 - 0.2 = 0.8$$

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\Rightarrow Z = \frac{0.22-0.1750}{\sqrt{(0.2\times0.8)\left(\frac{1}{300}+\frac{1}{200}\right)}}$$

$$\Rightarrow Z = \frac{0.045}{\sqrt{(0.16)(0.00833)}}$$

$$\Rightarrow Z = \frac{0.045}{\sqrt{0.001328}}$$

$$\Rightarrow Z = \frac{0.045}{0.03644}$$

$$\Rightarrow Z = 1.233$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 1.233 < 1.96$

Hence, the null hypothesis is accepted at 5% level of significance

5. In a large city A, 20% of a random sample of 900 school boys had a slight physical defect. In another large city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference between the proportions significant?

Sol.

Set the null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.05$ (5%)

Given

$n_1 = 900, n_2 = 1600$

$x_1 =$ 20%of random sample of 900=0.2x900=180

$x_2 =$ 18.5%of random sample of 1600=0.185x1600=296

$$\therefore p_1 = 20\% = \frac{20}{100} = 0.2, p_2 = 18.5\% = \frac{18.5}{100} = 0.185$$

We know that $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{180+296}{900+1600}$$

$$\Rightarrow P = \frac{476}{2500}$$

$$\Rightarrow P = 0.1904 \Rightarrow Q = 1 - P = 1 - 0.1904 = 0.8096$$

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\Rightarrow Z = \frac{0.2-0.185}{\sqrt{(0.1904\times0.8096)\left(\frac{1}{900}+\frac{1}{1600}\right)}}$$

$$\Rightarrow Z = \frac{0.015}{\sqrt{(0.1541)(0.00173)}}$$

$$\Rightarrow Z = \frac{0.015}{\sqrt{0.00026}}$$

$$\Rightarrow Z = \frac{0.015}{0.01612}$$

$$\Rightarrow Z = 0.9305$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 0.9305 < 1.96$

Hence, the null hypothesis H_0 is accepted at 5% level of significance and hence there is no significant difference.

6. Before an increase in excise duty on tea, 800 persons out of a sample of 1000 persons were found to be tea drinkers. After an increase is excise duty. 800 people were tea drinkers in a sample of 1200 people. Test whether there is a significant decrease in the consumption of tea after the increase in excise duty at 5% Los.

Sol.

Set the null hypothesis $H_0; P_1 = P_2$
Set the Alternative hypothesis $H_1: P_1 \neq P_2$
The level of significance $\alpha = 0.05$ (5%)
Given

$$n_1 = 1000, n_2 = 1200 \ \& \ x_1 = 800, x_2 = 800$$

$$\therefore p_1 = \frac{800}{1000} = 0.8, p_2 = \frac{800}{1200} = 0.6670$$

We know that $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{800+800}{1000+1200}$$

$$\Rightarrow P = \frac{1600}{2200}$$

$$\Rightarrow P = 0.7272 \Rightarrow Q = 1 - P = 1 - 0.7272 = 0.2728$$

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\Rightarrow Z = \frac{0.8 - 0.6670}{\sqrt{\left((0.7272 \times 0.2728)\left(\frac{1}{1000} + \frac{1}{1200}\right)\right)}}$$

$$\Rightarrow Z = \frac{0.133}{\sqrt{(0.1983)(0.00183)}}$$

$$\Rightarrow Z = \frac{0.133}{\sqrt{0.00036}}$$

$$\Rightarrow Z = \frac{0.133}{0.0189}$$

$$\Rightarrow Z = 7.037$$

At 5% level, the tabulated value of Zα is 1.645.
Since |Z| = 7.037 > 1.645
Hence Null Hypothesis H_0 is rejected at 5% level of significance.
There is a significance decrease in the consumption of tea due to increase in excise duty.

7. In a sample of 600 men from a certain city, 450 are found smokers. In another sample of 900 men from another city, 450 are smokers. Do the indicate that the cities are significantly different with respect to the habit of smoking among men. Test at 5% significance level.

(Warning: Smoking is injurious to health, causes cancer, Tabaco causes painful death)

Sol.

Set the null hypothesis $H_0; P_1 = P_2$
Set the Alternative hypothesis $H_1: P_1 \neq P_2$
The level of significance $\alpha = 0.05$ (5%)
Given

$$n_1 = 600, n_2 = 900 \ \& \ x_1 = 450, x_2 = 450$$

$$\therefore p_1 = \frac{450}{600} = 0.75, p_2 = \frac{450}{900} = 0.5$$

We know that $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{450+450}{600+900}$$

$$\Rightarrow P = \frac{900}{1500} = 0.6$$

$$\Rightarrow P = 0.6 \Rightarrow Q = 1 - P = 1 - 0.6 = 0.4$$

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\Rightarrow Z = \frac{0.75 - 0.5}{\sqrt{\left(0.6 \times 0.4)\left(\frac{1}{600} + \frac{1}{900}\right)\right)}}$$

$$\Rightarrow Z = \frac{0.25}{\sqrt{(0.24)(0.00277)}}$$

$$\Rightarrow Z = \frac{0.25}{\sqrt{0.0006648}}$$

$$\Rightarrow Z = \frac{0.25}{0.02578}$$
$$\Rightarrow Z = 9.69$$

At 5% level, the tabulated value of Zα is 1.645.
Since |Z| = 9.69 > 1.645
Hence Null Hypothesis H_0 is rejected at 5% level of significance.

8. One type of air craft is found to develop engine trouble in 5 flights out of a total of 100 and another type in 7 flights out of a total of 200 flights. Is there a significance difference in the two types of air craft's so far as engine defects are concerned? Test at 5% significance level.

Sol.

Set the null hypothesis $H_0; P_1 = P_2$
Set the Alternative hypothesis $H_1: P_1 \neq P_2$
The level of significance $\alpha = 0.05$ (5%)
Given

$$n_1 = 100, n_2 = 200 \ \ \& \ x_1 = 5, x_2 = 7$$
$$\therefore p_1 = \frac{5}{100} = 0.05, p_2 = \frac{7}{200} = 0.35$$

We know that $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{5+7}{100+200}$$
$$\Rightarrow P = \frac{12}{300}$$
$$\Rightarrow P = 0.04 \Rightarrow Q = 1 - P = 1 - 0.04 = 0.96$$
$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$
$$\Rightarrow Z = \frac{0.05-0.35}{\sqrt{(0.04\times0.96)\left(\frac{1}{100}+\frac{1}{200}\right)}}$$
$$\Rightarrow Z = -\frac{0.3}{\sqrt{(0.384)(0.015)}}$$
$$\Rightarrow Z = -\frac{0.3}{\sqrt{0.00576}}$$
$$\Rightarrow Z = -\frac{0.3}{0.07589}$$
$$\Rightarrow Z = -3.953$$

At 5% level, the tabulated value of Zα is 1.645.
Since |Z| = 3.953 > 1.645
Hence Null Hypothesis H_0 is rejected at 5% level of significance.

9. A machine produced 16 defective articles in a batch of 500. After overhauling it produced 3 defectives in a batch of 100. Has the machine improved?

Sol.

Set the null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.01$ (1%)

Given

$$n_1 = 500, n_2 = 100 \ \ \& \ x_1 = 16, x_2 = 3$$
$$\therefore p_1 = \frac{16}{500} = 0.032, p_2 = \frac{3}{100} = 0.03$$

We know that $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{16+3}{500+100}$$
$$\Rightarrow P = \frac{19}{600}$$
$$\Rightarrow P = 0.03166 \Rightarrow Q = 1 - P = 1 - 0.03166 = 0.96834$$
$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\Rightarrow Z = \frac{0.032-0.03}{\sqrt{\left(0.03166\times0.96834\right)\left(\frac{1}{500}+\frac{1}{100}\right)}}$$

$$\Rightarrow Z = \frac{0.002}{\sqrt{(0.03065)(0.012)}}$$

$$\Rightarrow Z = \frac{0.002}{\sqrt{0.0003678}}$$

$$\Rightarrow Z = \frac{0.002}{0.01917}$$

$$\Rightarrow Z = 0.1047$$

At 1% level, the tabulated value of Zα is 1.96.
Since |Z| = 0.1047 <1.96
Hence Null Hypothesis H_0 is accepted at 1% level of significance
∴the machine is not improved after overhauling.

10. A machine produced 25 defective articles in a batch of 400. After over hauling it produced 15 defectives in a batch of 200. Test at 1% level of significance whether there is a reduction of defective articles after overhauling.

Sol.

The null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.01$ (1%)

Given

$$n_1 = 400, n_2 = 200 \ \ \& \ x_1 = 25, x_2 = 15$$

$$\therefore p_1 = \frac{25}{400} = 0.0625, p_2 = \frac{15}{200} = 0.075$$

We know that $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{25+15}{400+200}$$

$$\Rightarrow P = \frac{40}{600}$$

$$\Rightarrow P = 0.0666 \Rightarrow Q = 1 - P = 1 - 0.0666 = 0.9334$$

$$\therefore Z = \frac{p_1-p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\Rightarrow Z = \frac{0.0625-0.075}{\sqrt{\left(0.0666\times0.9334\right)\left(\frac{1}{400}+\frac{1}{200}\right)}}$$

$$\Rightarrow Z = -\frac{0.0125}{\sqrt{(0.0621)(0.0075)}}$$

$$\Rightarrow Z = -\frac{0.0125}{\sqrt{0.00046}}$$

$$\Rightarrow Z = -\frac{0.0125}{0.0214}$$

$$\Rightarrow Z = -0.5841$$

At 1% level, the tabulated value of Zα is 1.96.
Since |Z| = 0.5841 <1.96
Hence Null Hypothesis H_0 is accepted at 1% level of significance

11. In an examination given to students at a large number of different schools the mean grade was 74.5 and S.D grade was 8. At one particular school where 200 students took the examination the mean grade was 75.9. Discuss the significance of this result at both 5% and 1% level of significance.

Sol.

The level of significance $\alpha = 0.05$ (5%)$\Rightarrow Z_{0.05}$ =1.96
The level of significance $\alpha = 0.01$ (1%)$\Rightarrow Z_{0.01}$=1.64
Given

$$n = 200$$
$$\sigma = 8$$
$$\mu = 74.5 and \bar{x} = 75.9$$

We calculate Z through Test Statistic,

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$
$$\Rightarrow Z = \frac{75.9 - 74.5}{8/\sqrt{200}}$$
$$\Rightarrow Z = \frac{1.4}{8/14.1421}$$
$$\Rightarrow Z = \frac{1.4 \times 14.1421}{8}$$
$$\Rightarrow Z = 2.4748$$

i) Thus At 5% level, the tabulated value of Zα is 1.645.
Since |Z| = 2.4748 > 1.96
Hence Null Hypothesis H_0 is rejected at 5% level of significance.
ii) Thus At 1% level, the tabulated value of Zα is 1.645.
Since |Z| = 2.4748 > 1.645
Hence Null Hypothesis H_0 is rejected at 1% level of significance.

12. Intelligent tests were given to the two groups of boys and girls,

	Mean	S.D	Size
Girls	75	8	60
Boys	73	10	100

Find out if the two mean significantly differ at 5% level of significance.

Sol^n:

Set The null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

where, P1 refers the girls and P2 refers the boys

Given, the means, S.D's & sizes of both the groups of girls and boys are as follows,

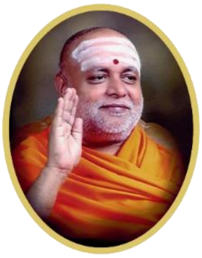
$$\overline{x_1} = 75 \;,\; \overline{x_2} = 73 \;,\; \sigma_1 = 8 \;,\; \sigma_2 = 10 \;, n_1 = 60 \;,\; n_2 = 100$$

WKT, $Z = \frac{(\bar{x}_2 - \bar{x}_1)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}} = \frac{(73-75)}{\sqrt{\frac{64}{60} + \frac{100}{100}}} = -\frac{2}{\sqrt{2.07}} = -1.3898.$

Thus At 5% level, the tabulated value of Zα is 1.96.
Since |Z| = 1.3898 < 1.96

Hence, the null hypothesis is accepted at 5% level of significance, i.e., there is no significant difference.

***



||Jai Sri Gurudev||

S.J.C. INSTITUTE OF TECHNOLOGY, CHICKBALLAPUR

DEPARTMENT OF MATHEMATICS

MATHEMATICS-3 FOR COMPUTER SCIENCE (BCS301)

MODULE - 4

STATISTICAL INFERENCE -2

Prepared by:

Purushotham P
Assistant Professor
SJC Institute of Technology
Email id: purushotham@sjcit.ac.in

Statistics: Any function of the sample values is known as a statistics.
Eg: Sample mean, Sample median, Sample variance etc. are all statistics.

Sampling Distribution: A sampling distribution is a distribution of a statistic over all possible samples. That is sampling distribution is the probability distribution of the statistics.

Sampling Variables: Variables sampling is the process used to predict the value of a specific variable within a population. For example, a limited sample size can be used to compute the average accounts receivable balance, as well as a statistical derivation of the plus or minus range of the total receivables value that is under review.

The Central Limit Theorem: Suppose that a sample of size n is selected from a population that has mean μ and the standard deviation σ, then Let $x_1, x_2, x_3, x_4 \ldots\ldots, x_n$ be the n observations, they are independent and identically distributed with mean $\bar{X} = \frac{x_1+x_2+x_3+\ldots\ldots x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$, the central limit theorem states that the sample mean $\bar{x}$ follows approximately the normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ (is also called Standard error) , i.e. $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, where μ , σ are mean and standard deviation of the population from where the sample was selected and the sample size becomes large ($n \geq 30$).

Degrees of freedom: Degrees of freedom refer to the maximum number of logically independent values, which may vary in a data sample. Degrees of freedom are calculated by subtracting one from the number of items within the data sample (n -1).

Description	Population notation	Sample Notation
Size	N	n
Mean	μ	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Variance	σ^2	$s^2 = \frac{1}{n-1}\sum (x - \bar{x})^2$

Standard deviation	σ	$s = \sqrt{\frac{1}{n-1}\sum(x-\bar{x})^2}$

Confidence Intervals:

Suppose we want to estimate an actual population meanμ. As you know, we can only obtain$\bar{x}$, the mean of a sample randomly selected from the population of interest. We can use $\bar{x}$ to find a range of values:

Lower value<population meanμ <Upper value

That we can be really confident contains the population meanμ. The range of values is called a "**confidence interval**."

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

Confidence interval $C.I. == Mean \pm Z(Standard\ Deviation/\sqrt{SampleSize} = \mu \pm Z\frac{\sigma}{\sqrt{n}}$ or $\bar{X} = \mu \pm Z\frac{\sigma}{\sqrt{n}}$

Confidence Level	99%	98%	95%	90%	50%
Z	2.58	2.33	1.96	1.645	0.6745

PROBLEMS

1. State Central limit theorem. Use the theorem to evaluate P[50 < $\bar{X}$ < 56] where $\bar{X}$ represents the mean of a random sample of size 100 from an infinite population with mean μ = 53 and variance σ^2 = 400.

Sol.

The central limit theorem states that the sample mean $\bar{x}$ follows approximately the normal distribution with mean $\mu$and standard deviation $\frac{\sigma}{\sqrt{n}}$ (is also called Standard error), i.e., $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, where μ , $\sigma$are mean and standard deviation of the population from where the sample.

Given,

Sample size n=100

Mean of the population $\mu = 53$

Variance of the population $\sigma^2 = 400 \Rightarrow \sigma = \sqrt{400} = 20$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

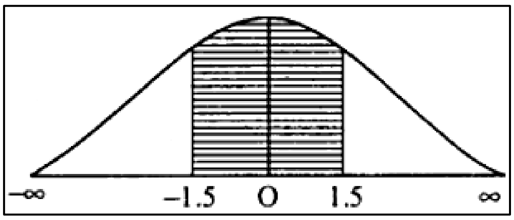
$$\Rightarrow \bar{X} \sim N\left(53, \frac{20}{\sqrt{100}}\right)$$

$$\Rightarrow \bar{X} \sim N(53,2)$$

$\therefore$ we know that $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

$$\Rightarrow Z = \frac{\bar{X}-53}{20/\sqrt{100}}$$

$$\Rightarrow Z = \frac{\bar{X}-53}{2}$$



$\therefore$ At $\bar{X}$=50 $\Rightarrow Z = \frac{50-53}{2} = -\frac{3}{2} = -1.5 = z_1$

At $\bar{X}$=56 $\Rightarrow Z = \frac{56-53}{2} = \frac{3}{2} = 1.5 = z_2$

$$\therefore P(50 < \bar{X} < 56) = P(-1.5 < z < 1.5)$$
$$= 2P(0 < z < 1.5)$$
$$= 2A(1.5)$$

$= 2 \times 0.4332$

$\therefore P(50 < \bar{X} < 56) = 0.8664$

2. An unknown distribution has a mean of 90 and a standard deviation of 15. Samples of size n = 25 are drawn randomly from the population. Find the probability that the sample mean is between 85 and 92.

Sol.

Given,

Sample size n=25

Mean of the population $\mu = 90$

Variance of the population $\Rightarrow \sigma = 15$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

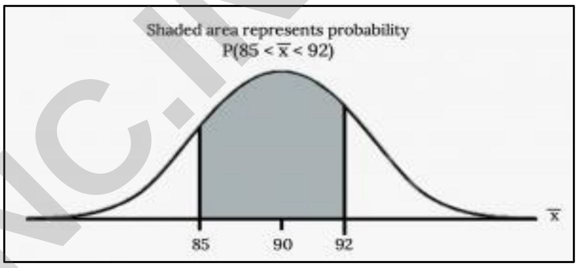
$$\Rightarrow \bar{X} \sim N\left(90, \frac{15}{\sqrt{25}}\right)$$

$$\Rightarrow \bar{X} \sim N(90,3)$$

$$\therefore \text{we know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

$$\Rightarrow Z = \frac{\bar{X}-90}{15/\sqrt{25}}$$

$$\Rightarrow Z = \frac{\bar{X}-90}{3}$$



$\therefore \text{At } \bar{X} = 85 \Rightarrow z = \frac{85-90}{3} = -\frac{5}{3} = -1.66$

$\therefore \text{At } \bar{X} = 92 \Rightarrow z = \frac{92-90}{3} = \frac{2}{3} = 0.66$

$\therefore P(85 < \bar{X} < 92) = P(-1.66 < z < 0.66)$

$\Rightarrow P(-1.66 < z < 0.66) = P(0 < z < 1.66) + P(0 < z < 0.66)$

$= 0.4515 + 0.2454$

$\Rightarrow P(-1.66 < z < 0.66) = 0.6965$

3. A random sample of size 64 is taken from an infinite population having mean 112 and variance 144. Using central limit theorem, find the probability of getting the sample mean $\bar{X}$ greater than 114.5.

Sol.

Given,

Sample size n=64

Mean of the population $\mu = 112$

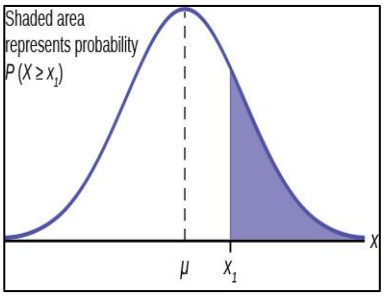
Variance of the population $\Rightarrow \sigma^2 = 144 \Rightarrow \sigma = 12$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\Rightarrow \bar{X} \sim N\left(112, \frac{12}{\sqrt{64}}\right)$$

$$\Rightarrow \bar{X} \sim N(90,1.5)$$

$$\therefore \text{we know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$



$$\Rightarrow Z = \frac{\bar{X}-112}{12/\sqrt{64}}$$

$$\Rightarrow Z = \frac{\bar{X}-112}{1.5}$$

$\therefore \text{At } \bar{X} = 114.5 \Rightarrow z = \frac{114.5-112}{1.5} = 1.66$

$$\therefore P(\bar{X} > 114.5) = P(z > 1.66)$$
$$\Rightarrow P(z > 1.66) = 0.5 - P(0 < z < 1.66)$$
$$= 0.5 - 0.4515$$
$$\Rightarrow P(z > 1.66) = 0.0489$$

4. Let $\bar{X}$denote the mean of a random sample of size 100 from a distribution, that is $\chi^2(50)$. Compute an approximate value of P(49<$\bar{X}$<51).

Sol.

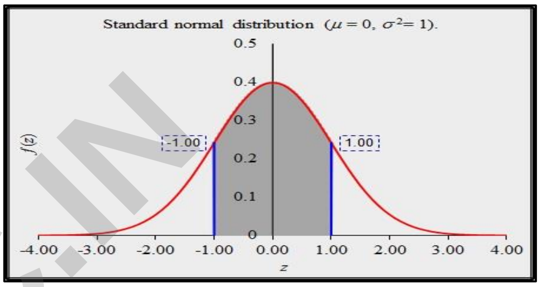
The sample size n is=100

The chi-square distribution is given as ,X~ $\chi^2(50)$, where d.f.=50

The mean and variance of chi-square distribution is given as, μ = 50

Therefore $\Rightarrow \sigma^2 = 2 \times d.f. = 2 \times 50 = 100 \Rightarrow \sigma = 10$

The sample mean of chi-square distribution follows normal distribution with mean and standard error $\frac{\sigma}{\sqrt{n}}$.



$$\therefore \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$
$$\Rightarrow \bar{X} \sim N\left(50, \frac{10}{\sqrt{100}}\right)$$
$$\Rightarrow \bar{X} \sim N(50,1)$$
$$\therefore \text{we know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

$$\therefore \text{At } \bar{X}\text{=50} \Rightarrow Z = \frac{49-50}{1} = -\frac{1}{1} = -1 = z_1$$
$$\text{At } \bar{X}\text{=51} \Rightarrow Z = \frac{51-50}{1} = \frac{1}{1} = 1 = z_2$$

$$\therefore P(49 < \bar{X} < 51) = P(-1 < z < 1)$$
$$= 2P(0 < z < 1)$$
$$=A(1)$$
$$= 2 \times 0.3416$$
$$\therefore P(50 < \bar{X} < 56) = 0.6826$$

5. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distribute with mean 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

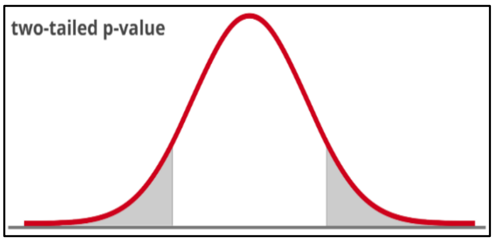
Sol.

Total number of bulbs n=16

An average life of bulbs μ = 800

Standard deviation of the bulbs $\Rightarrow \sigma = 40$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$
$$\Rightarrow \bar{X} \sim N\left(800., \frac{40}{\sqrt{16}}\right)$$
$$\Rightarrow \bar{X} \sim N(800,10)$$



$$\therefore \text{We know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-800}{10}$$

$$\therefore \text{At } \bar{X}\text{=775} \Rightarrow Z = \frac{775-800}{10} = -\frac{25}{10} = -2.5.$$

$$\therefore P(\bar{X} < 775) = P(z < -2.5)$$
$$\Rightarrow P(z < -2.5) = P(z > 2.5)$$
$$\Rightarrow P(z < -2.5) = 0.5 - P(0 < z < 2.5)$$
$$\Rightarrow P(z < -2.5) = 0.5 - A(2.5)$$

$$\Rightarrow P(z < -2.5) = 0.5 - 0.4938$$
$$\Rightarrow P(z < -2.5) = 0.0062$$

6. The heights of a random sample of 50 college students showed a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Construct a 99% confidence interval for the mean height of all college students.

Sol.

Given the sample size n=50
Average height of Students (Mean) $\mu = 174.5c.m.$
Standard deviation of the Students $\sigma = 6.9c.m.$
We know that, Confidence level of 99%, the corresponding z value is 2.576. This is determined from the normal distribution table.

Confidence interval $C.I. == Mean \pm Z(S\tan d\,ardDeviation/\sqrt{SampleSize} = \mu \pm Z\frac{\sigma}{\sqrt{n}}$

$$\therefore C.I. = 174.5 \pm \left(2.576 \times \frac{6.9}{\sqrt{50}}\right)$$
$$\Rightarrow C.I. = 174.5 \pm (2.576 \times 0.9758)$$
$$\Rightarrow C.I. = 174.5 \pm 2.5136$$

The lower end of the confidence interval is $= 174.5 - 2.5136 = 171.9864$
The upper end of the confidence interval is $= 174.5 + 2.5136 = 177.0136$

Therefore, with 99% confidence interval, the mean height of all college students is between 171.9864 centimeters and 177.0136 centimeters.

7. The mean and SD of the diameters of a sample of 250 rivet heads manufactured by a company are 7.2642 mm and 0.0058 mm respectively. Find,

a) 99% b) 98% c) 95% d) 90% e) 50%

Confidence limits for the mean diameter of all the rivet heads manufactured by the company.

Sol.

Given the sample size n=250
Mean of a diameter $\mu = 7.2642mm.$
Standard deviation of the diameter $\sigma = 0.0058mm$
We know that,

Confidence Level	99%	98%	95%	90%	50%
Z	2.58	2.33	1.96	1.645	0.6745

Confidence interval $\boldsymbol{C.I. = Mean \pm Z(S\tan d\,ardDeviation/\sqrt{SampleSize} = \mu \pm Z\frac{\sigma}{\sqrt{n}}}$

Confidence Level	$\boldsymbol{C.I. = \mu \pm Z\frac{\sigma}{n}}$	Final C.I.	Interval
99%	$7.2642 \pm \left(2.58 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00094	(7.26326 , 7.26514)

98%	$7.2642 \pm \left(2.33 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00086	(7.26334 , 7.26504)
95%	$7.2642 \pm \left(1.96 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00073	(7.26347 , 7.26493)
90%	$7.2642 \pm \left(1.645 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00061	(7.26359 , 7.26481)
50%	$7.2642 \pm \left(0.6745 \times \frac{0.0058}{\sqrt{250}}\right)$	7.2642 ± 0.00025	(7.26395 , 7.26445)

8. A random sample of size 25 from a normal distribution ($\sigma^2 = 4$) yields, sample mean $\bar{X} = 78.3$. Obtain a 99% confidence interval for μ.

Sol.

Given the sample size n=25

Mean of sample $\bar{X} = 78.3$

Standard deviation $\sigma = 2$

We know, Confidence level of 99%, the corresponding z value is 2.58. This is determined from the normal distribution table.

Confidence interval $C.I. = \mu = Mean \pm Z(S\tan d\,ardDeviation/\sqrt{SampleSize} = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$

$$\therefore C.I. = \mu = 78.3 \pm \left(2.58 \times \frac{2}{\sqrt{25}}\right)$$

$$\Rightarrow \mu = 78.3 \pm 1.032$$

$$\Rightarrow C.I \Rightarrow (78.3 - 1.032\,, 78.3 + 1.032) = (77.268\,, 79.332)$$

9. Let the observed value of the mean $\bar{X}$of a random sample of size 20 from a normal distribution with $mean\ \mu$ and variance $\sigma^2 = 80$ be 81.2. Find a 90% and 95% confidence intervals for μ.

Sol.

Given the sample size n=20

Mean of sample $\bar{X} = 81.2$

Variance $\sigma^2 = 80 \Rightarrow \sigma = \sqrt{80} = 8.9442$

We know, Confidence level of 95%, 90% the corresponding z values are 1.96 , 1.645. This is determined from the normal distribution table.

Confidence interval $C.I. = \mu = Mean \pm Z(S\tan d\,ardDeviation/\sqrt{SampleSize} = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$

For 95%:

$$\therefore C.I. = \mu = 81.2 \pm \left(1.96 \times \frac{8.9442}{\sqrt{20}}\right)$$

$$\Rightarrow \mu = 81.2 \pm 3.92$$

$$\Rightarrow C.I = (81.2 - 3.92, 81.2 + 3.92) = (77.28, 85.12)$$

For 90%:

$$\therefore C.I. = \mu = 81.2 \pm \left(1.645 \times \frac{8.9442}{\sqrt{20}}\right)$$
$$\Rightarrow \mu = 81.2 \pm 3.29$$
$$\Rightarrow C.I = (81.2 - 3.29, 81.2 + 3.29) = (77.91, 84.49)$$

10. Suppose that 10, 12, 16, 19 is a sample taken from a normal population with variance 6.25. Find at 95% confidence interval for the population mean.

Sol.

Given samples are 10, 12, 16 and 19

Therefore, sample size n=4

Mean $\bar{X}$=14.25

Variance $\sigma^2 = 6.25 \Rightarrow \sigma = \sqrt{6.25} = 2.5$

We know, Confidence level of 95%, the corresponding z value is 1.96, This is determined from the normal distribution table.

Confidence interval $C.I. = \mu = Mean \pm Z(S\tan d\,ardDeviation/\sqrt{SampleSize} = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$

$$\therefore C.I. = \mu = 14.25 \pm \left(1.96 \times \frac{2.5}{\sqrt{4}}\right)$$
$$\Rightarrow \mu = 14.25 \pm 2.45$$
$$\Rightarrow C.I = (14.25 - 2.45, 14.25 + 2.45) = (11.80, 16.70)$$

SAMPLING DISTRIBUTIONS

Student's t -distribution:

Let μbe the mean of population, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the mean and $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ be the standard deviation of a sample, then the Student's t -distribution is defined as

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{\bar{x}-\mu}{s}\sqrt{n}$$

Another formula for t - test of two samples is

$$t = \frac{(\bar{x}_2 - \bar{x}_1)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

where, $s^2 = \frac{n_1{\sigma_1}^2 + n_2{\sigma_2}^2}{n_1+n_2-2}$ or $s = \sqrt{\frac{1}{n_1+n_2-2}\left[\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_i - \bar{x}_2)^2\right]}$

Chi-square distribution:

Let $O_i(i = 1,2,3\ldots n)$ and $E_i(i = 1,2,3\ldots n)$ be the set of observed frequencies and expected frequencies respectively, then the Chi-square distribution is defined as

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \frac{(O_3-E_3)^2}{E_3} + \ldots\ldots + \frac{(O_n-E_n)^2}{E_n}$$
$$\Rightarrow \chi^2 = \sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i}$$

F-Distribution:

The F-distribution is useful in hypothesis testing. Hypothesis testing is used by scientists to statistically compare data from two or more populations. The F-distribution is needed to determine whether the F-value for a study indicates any statistically significant differences between two populations.

F-test is to determine whether the two independent estimates of population variance differ significantly. In this case, F-ratio is : $F = \frac{{\sigma_1}^2}{{\sigma_2}^2}$ where $\sigma^2 = \frac{1}{n}\sum(x-\mu)^2$ or

$$F = \frac{{s_1}^2/{\sigma_1}^2}{{s_2}^2/{\sigma_2}^2},$$

where σ_1=Standard deviation of population-1
σ_2=Standard deviation of population-2
s_1=Standard deviation of sample-1
s_2=Standard deviation of sample-2

To find out whether the two samples drawn from the normal population have the same variance. In this case, F-ratio is,

$F = \frac{{s_1}^2}{{s_2}^2}$ Where ${s_1}^2 = \frac{1}{n_1-1}\sum(x-\bar{x})^2, {s_2}^2 = \frac{1}{n_2-1}\sum(y-\bar{y})^2$

It should be noted that numerator is always greater than the denominator in F-ratio

$$F = \frac{Larger\ Variance}{Smaller\ Variance}$$

n_1= d.f for sample having larger variance
n_2 = d.f for sample having smaller variance

Expected value of F:

$F_E = \frac{{s_2}^2}{{s_1}^2}$ follows F- distribution with $v_1 = n_1 - 1$, $v_2 = n_2 - 1$ d.f.

11. A certain stimulus administered to each of the 12 patients resulted in the following change in the blood pressure 5,2,8,-1,3,0,6,-2,1,5,0,4. Can it be concluded that the stimulus will increase the blood pressure? (*Note:* $t_{0.05}$ for 11 d.f. is 2.201).

Sol.

Given the change in blood pressure

x: 5,2,8,-1,3,0,6,-2,1,5,0,4

$\therefore \bar{x} = \frac{1}{n}\sum x = \frac{31}{12} = 2.5833$

Variance, $s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2$

$\Rightarrow s^2 =$

$$\frac{1}{11}\begin{Bmatrix}(5-2.58)^2+(2-2.58)^2+(8-2.58)^2+(-1-2.58)^2+(0-2.58)^2+(6-2.58)^2\\+(-2-2.58)^2+(1-2.58)^2+(5-2.58)^2+(0-2.58)^2+(4-2.58)^2\end{Bmatrix}$$

$\Rightarrow s^2 = 9.538 \Rightarrow s = 3.088$

Let us suppose that the stimulus administration is not accompanied with increase in blood pressure, we can take $\mu = 0$

we have,

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

$$\Rightarrow t = \frac{2.5833-0}{\left(\frac{3.088}{\sqrt{12}}\right)}$$

$$\Rightarrow t = 2.8979 \approx 2.9 > 2.201$$

Hence the hypothesis is rejected at 5%level of significance. We conclude with 95% Confidence that the stimulus in general is accompanied with increase of blood pressure.

12. A random sample of 10 boys had the following I.Q: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Does this data support the assumption of a population mean I.Q. of 100 at 5% level of Significance? (*Note:*$t_{0.05}$ =2.262 for 9 d.f.).

Sol.

Given the I.Q. of 10 boys

x: 70, 120, 110,101,88,83,95,98,107,100

$\therefore \bar{x} = \frac{1}{n}\sum x = \frac{972}{10} = 97.2$

Variance, $s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2$

$\Rightarrow s^2 = \frac{1}{9} \times 1833.6$

$\Rightarrow s^2 = 203.73333$

$\Rightarrow s = 14.2735$

Given the mean of population $\mu = 100$

We have,

$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

$\Rightarrow t = \frac{97.2-100}{\left(\frac{14.2735}{\sqrt{10}}\right)}$

$\Rightarrow t = \frac{-2.8}{4.5136} \approx -0.6203 < 2.262$

13. Ten individuals are chosen at random from a population and their heights in inches are found to be 63, 63,66,67,68,69,70,70,71,71 . Test the hypothesis that the mean height of the universe is 66inches ($t_{0.05}$ =2.262 for 9 d.f.).

Sol.

Given the heights of the population in inches

x: 63, 63,66,67,68,69,70,70,71,71

$\therefore \bar{x} = \frac{1}{n}\sum x = \frac{678}{10} = 67.8$

Variance, $s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2$

$\Rightarrow s^2 =$

$$\frac{1}{9}\begin{Bmatrix}(63-67.8)^2 + (63-67.8)^2 + (66-67.8)^2 + (67-67.8)^2 + (68-67.8)^2 + (69-67.8)^2 \\ +(70-67.8)^2 + (70-67.8)^2 + (71-67.8)^2 + (71-67.8)^2\end{Bmatrix}$$

$\Rightarrow s^2 = 9.067 \Rightarrow s = 3.011$

And given the mean of population $\mu = 66$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

$$\Rightarrow t = \frac{67.8-66}{\left(\frac{3.011}{\sqrt{10}}\right)}$$

We have $\Rightarrow t = 1.8979 \approx 1.89 > 2.262$

Thus, the hypothesis is accepted at 5% level of significance.

14. The nine items of a sample have the following values: 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these differ significantly from the assumed mean of 47.5 at 5% significance level?

Sol.

Given sample values: 45, 47, 50, 52, 48, 47, 49, 53, 51

Therefore, sample size n=9

Population Mean $\mu = 47.50$

$\therefore$Sample mean $\bar{x} = \frac{1}{n}\sum x = \frac{442}{9} = 49.11$

Variance, $s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2$

$\Rightarrow s^2 =$

$$\frac{1}{8}\left\{\begin{matrix}(45-49.11)^2+(47-49.11)^2+(50-49.11)^2+(52-49.11)^2+(48-49.11)^2 \\ +(47-49.11)^2+(49-49.11)^2+(53-49.11)^2+(51-49.11)^2\end{matrix}\right\}$$

$\Rightarrow s^2 = \frac{54.9}{8} = 6.8625 \Rightarrow s = \sqrt{6.8625} = 2.6196$

∴Thee Null hypothesis $H_0: \mu = 47.5$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

$$\Rightarrow t = \frac{49.11-47.5}{\left(2.6196/\sqrt{9}\right)}$$

$$\Rightarrow t = \frac{1.61}{0.8732}$$

$$\Rightarrow t = 1.8437$$

∴ Level of significance =5%

Critical value at 5 % level of significance for v=9-1=8 degrees of freedom is 2.3060.

Since the calculated value 1.8437 is less than the tabulated value 2.3060.

Hence the Null hypothesis is accepted.

15. Two types of batteries are tested for their length of life and the following results are obtained:

Battery A: $n_1 = 10, \bar{x}_1 = 500hrs., \sigma_1{}^2 = 100$

Battery B: $n_2 = 10, \bar{x}_2 = 560hrs., \sigma_2{}^2 = 121$

Compute Student's t and test whether there is a significant difference in the two means.

Sol.

Given

Battery A: $n_1 = 10, \bar{x}_1 = 500hrs., \sigma_1{}^2 = 100$

Battery B: $n_2 = 10, \bar{x}_2 = 560hrs., \sigma_2{}^2 = 121$

We know that,

$$s^2 = \frac{n_1\sigma_1{}^2+n_2\sigma_2{}^2}{n_1+n_2-2}$$

$$\Rightarrow s^2 = \frac{(10\times100)+(10\times121)}{10+10-2}$$

$$\Rightarrow s^2 = 122.78$$

$$\Rightarrow s = 11.0805$$

We have,

$$t = \frac{(\bar{x}_2-\bar{x}_1)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$\Rightarrow t = \frac{560-500}{11.0805\sqrt{0.1+0.1}}$$

$$\Rightarrow t = 12.1081 \approx 12.11$$

The value of t is greater than the table value of t for 18d.f.at all levels of significance.

16. A group of boys and girls were given an intelligence test. The mean score , SD score and numbers in each group are as follows.

	Boys	**Girls**
Mean	**74**	**70**
SD	**8**	**10**
n	**12**	**10**

Is the difference between the means of the two groups significant at 5% level of significance ($t_{0.05} = 2.086$ for 20 $d.f.$)

Sol.

Given

$\bar{x}_1 = 74, \sigma_1 = 8, n_1 = 12\{Boys\}$

$\bar{x}_2 = 70, \sigma_2 = 10, n_2 = 10\{Girls\}$

We know that

$$s^2 = \frac{n_1{\sigma_1}^2 + n_2{\sigma_2}^2}{n_1+n_2-2}$$

$$\Rightarrow s^2 = \frac{(12\times64)+(10\times100)}{12+10-2}$$

$$\Rightarrow s^2 = \frac{1768}{20} = 88.4$$

$$\Rightarrow s = 9.402 \approx 9.4$$

We have

$$t = \frac{|(\bar{x}_2 - \bar{x}_1)|}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$\Rightarrow t = \frac{74-70}{9.4\sqrt{\frac{1}{12}+\frac{1}{10}}} = \frac{4}{9.4\times0.4281} = \frac{4}{4.0244} = 0.9939$$

Thus, the hypothesis that there is a difference between the means of the two groups is accepted at 5% level of significance.

17. Two horses A and B were tested according to the time (In Seconds) to run a particular race with the following results:

Horse A	**28**	**30**	**32**	**33**	**33**	**29**	**34**
Horse B	**29**	**30**	**30**	**24**	**27**	**29**	**-**

Test whether you can discriminate between the two horses.

Sol.

Let the variables x and y respectively correspond to Horse A and B

x: 28,30,32,33,,33,29,34

y: 29,30,30,24,27,29

$\therefore \bar{x} = \frac{1}{n_1}\sum_i x_i = \frac{219}{7} = 31.30$, $\therefore \bar{y} = \frac{1}{n_2}\sum_i y_i = \frac{169}{6} = 28.20$

$\sum(x-\bar{x})^2 = (28-31.3)^2 + (30-31.3)^2 + (32-31.3)^2 + (33-31.3)^2 + (33-31.3)^2 + (29-31.3)^2 + (34-31.3)^2 = 31.4$

$\sum(y-\bar{y})^2 = (29-28.20)^2 + (30-28.20)^2 + (30-28.20)^2 + (24-28.20)^2 + (27-28.20)^2 + (29-28.20)^2 = 26.84$

$$\therefore s^2 = \frac{1}{n_1+n_2-2}\left[\sum(x-\bar{x})^2 + \sum(y-\bar{y})^2\right]$$

$$\Rightarrow s^2 = \frac{31.4+26.84}{7+6-2} = 5.2973$$

$$\Rightarrow s = 2.3016$$

We have,

$$t = \frac{|(\bar{x}_2-\bar{x}_1)|}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \Rightarrow t = \frac{31.30-28.20}{2.3016\sqrt{\frac{1}{7}+\frac{1}{6}}} \Rightarrow t = 2.42 \begin{cases} > t_{0.05} = 2.2 \\ < t_{0.02} = 2.72 \end{cases}$$

18. Four coins are tossed 100 times and the following results were obtained:

No. of Heads	**0**	**1**	**2**	**3**	**4**
Frequency	**5**	**29**	**36**	**25**	**5**

Fit a binomial distribution for the data and test the goodness of fit $\chi 0_{.05}{}^{2}$= 9.49 for 4 d.f.

Sol.

Given the 4 coins are tossed 100 times

The probability of getting head is p=0.5, q=0.5

The probability mass function of a binomial distribution is

$P(X = x) = 4_{Cx}(0.5)^x(0.5)^{4-x}$

$P(0) = 4_{C0}(0.5)^0(0.5)^{4-0} = 0.0625$

$P(1) = 4_{C1}(0.5)^1(0.5)^{4-1} = 0.25$

$P(2) = 4_{C2}(0.5)^2(0.5)^{4-2} = 0.375$

$P(3) = 4_{C3}(0.5)^3(0.5)^{4-3} = 0.25$

$P(4) = 4_{C4}(0.5)^4(0.5)^{4-4} = 0.0625$

$\therefore E_0 = 100 \times 0.0625 = 6.25$

$E_1 = 100 \times 0.25 = 25$

$E_2 = 100 \times 0.375 = 37.5$

$E_3 = 100 \times 0.25 = 25$

$E_4 = 100 \times 0.0625 = 6.25$

where 100 is the sum of frequency

O_i	5	29	36	25	5
E_i	6.25	25	37.5	25	6.25

$\therefore \chi^2 = \sum_i \left[\frac{(O_i-E_i)^2}{E_i}\right]$

$\Rightarrow \chi^2 = \frac{1.5625}{6.25} + \frac{16}{25} + \frac{2.25}{37.5} + 0 + \frac{1.5625}{6.25}$

$\Rightarrow \chi^2 = 0.25 + 0.64 + 0.06 + 0.25$

$\Rightarrow \chi^2 = 1.2 < \chi^2{}_{0.05} = 9.49$

Hence the fitness is good.

19. A dice thrown 264 times and the number appearing on the face (x) follows the following frequency (f) distribution.

x	**1**	**2**	**3**	**4**	**5**	**6**
f	**40**	**32**	**28**	**58**	**54**	**60**

Calculate the value of χ^2.

Sol.

The frequencies in the given data are the observed frequencies. assuming that dice is unbiased, the expected number of frequencies for the numbers 1,2,3,4,5,6 to appear on the face is $\frac{264}{6} = 44$ each.

Now the data is as follows:

x	1	2	3	4	5	6
O_i	40	32	28	58	54	60
E_i	44	44	44	44	44	44

$$\therefore \chi^2 = \sum_i \left[\frac{(O_i-E_i)^2}{E_i}\right]$$

$$\Rightarrow \chi^2 = \frac{(40-44)^2}{44} + \frac{(32-44)^2}{44} + \frac{(28-44)^2}{44} + \frac{(58-44)^2}{44} + \frac{(54-44)^2}{44} + \frac{(60-44)^2}{44}$$

$$\Rightarrow \chi^2 = \frac{1}{\cancel{44}}[16 + 144 + 256 + 196 + 100 + 256]\frac{968}{44}$$
$$\Rightarrow \chi^2 = 22$$

20. A die was thrown 60 times and the following frequency distribution was observed:

Faces	1	2	3	4	5	6
Frequency	15	6	4	7	11	17

Test whether the die is unbiased at 5% significance level.

Sol.

The frequencies in the given data are the observed frequencies. Assuming that dice is unbiased, the expected number of frequencies for the numbers 1,2,3,4,5,6 to appear on the face is $\frac{60}{6} = 10$ each.

Now the data is as follows:

x	1	2	3	4	5	6
O_i	15	6	4	7	11	17
E_i	10	10	10	10	10	10

$$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right]$$
$$\Rightarrow \chi^2 = \frac{(15-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(4-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(17-10)^2}{10}$$
$$\Rightarrow \chi^2 = \frac{1}{10}[25 + 16 + 36 + 9 + 1 + 49] = \frac{136}{10}$$
$$\Rightarrow \chi^2 = 13.6$$

21. A survey of 320 families with 5 children each revealed the following distribution.

No. of boys	5	4	3	2	1	0
No. of girls	0	1	2	3	4	5
No. of families	14	56	110	88	40	12

Is the result consistent with the hypothesis that male and female births are equally probable at 5% level of significance?

Sol.

Given,

Number of families selected for the survey = 320

The probability of female and male birth is equal, $p = \frac{1}{2} = 0.5 \Rightarrow q = 1 - p = 1 - 0.5 = 0.5$

Number of children in the selected families, n = 5

No. of boys	5	4	3	2	1	0
No. of girls	0	1	2	3	4	5
No. of families	14	56	110	88	40	12

The statistical hypothesis is,

H_0: The probability of female and male birth is equal.

H_1: The probability of female and male birth is not equal.

Here Chi square distribution is used to test the hypothesis.

Therefore, by the Binomial distribution.

We have,

$$P(x) = nC_x p^x q^{n-x}$$
$$\therefore P(x) = 5C_x (0.5)^x (0.5)^{5-x}$$
$$\Rightarrow P(x) = 5C_x (0.5)^5$$

The expected frequencies can be calculated for 320 families as

$E(x) = 320 \times P(x) = 320 \times 5C_x(0.5)^5$

$\therefore E(0) = 320 \times P(0) = 320 \times 5C_0(0.5)^5 = 320 \times (0.5)^5 = 10 = E_0$

$E(1) = 320 \times P(1) = 320 \times 5C_1(0.5)^5 = 320 \times 5 \times (0.5)^5 = 50 = E_1$

$E(2) = 320 \times P(2) = 320 \times 5C_2(0.5)^5 = 320 \times 5C_2 \times (0.5)^5 = 100 = E_2$

$E(3) = 320 \times P(3) = 320 \times 5C_3(0.5)^5 = 320 \times 5C_3 \times (0.5)^5 = 100 = E_3$

$E(4) = 320 \times P(4) = 320 \times 5C_4(0.5)^5 = 320 \times 5C_4 \times (0.5)^5 = 50 = E_4$

$E(5) = 320 \times P(5) = 320 \times 5C_5(0.5)^5 = 320 \times 5C_5 \times (0.5)^5 = 10 = E_5$

No. of Boys	No. of Girls	Total Observed Frequencies (O_i)	Expected Frequencies(E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
5	0	14	10	4	16	1.6
4	1	56	50	6	36	0.72
3	2	110	100	10	100	1
2	3	88	100	-12	144	1.44
1	4	40	50	-10	100	2
0	5	12	10	2	4	0.4

We have the Table value ofχ^2 for 5 degrees of freedom at level of significance 5% from the chi-square table is 11.07.

$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right] \Rightarrow \chi^2 = 7.16 < 11.02$

Since the calculated χ^2value is less than tabulated χ^2 value then the decision is fail to reject the H_0 (Accept H_0) that means both the male and female birth is equal.

22. The theory predicts the proportion of beans in the four groups A, B, C and D should be 9:3:3:1. In an experiment among 1600 beans, the number in four groups were 882, 313, 287 and 118. The chi square value is approximately equal to.

Sol.

Given,

The total number of beans: 882+313+287+118=1600

Sum of the ratios: 9+3+3+1=16

$E(A) = 1600 \times \frac{9}{16} = 900$

$E(B) = 1600 \times \frac{3}{16} = 300$

$E(C) = 1600 \times \frac{3}{16} = 300$

$E(D) = 1600 \times \frac{1}{16} = 100$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{\sum(O_i - E_i)^2}{E_i}$
882	900	-18	324	0.36
313	300	13	169	0.5633
287	300	-13	169	0.5633
118	100	18	324	3.24

$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right]$

$\Rightarrow \chi^2 = 4.72$

23. Two random samples drawn from two normal populations are:

Sample-I	**20**	**16**	**26**	**27**	**22**	**23**	**18**	**24**	**19**	**25**	**-**	**-**
Sample-II	**27**	**33**	**42**	**35**	**32**	**34**	**38**	**28**	**41**	**43**	**30**	**37**

Obtain the estimates of the variance of the population and test 5% level of significance whether the two populations have the same variance.

Sol.

Set Null Hypothesis: $H_0: {\sigma_1}^2 = {\sigma_2}^2$

i.e., The two samples are drawn from two populations having the same variance.

Alternate Hypothesis: $H_1: {\sigma_1}^2 \neq {\sigma_2}^2$

Given,

Sample-I	20	16	26	27	22	23	18	24	19	25	-	-
Sample-II	27	33	42	35	32	34	38	28	41	43	30	37

$$\bar{x}_1 = \frac{\sum_{i=1}^{n} {}^1 x_i}{n_1} \Rightarrow \bar{x}_1 = \frac{20+16+26+27+22+23+18+24+19+25}{10} \Rightarrow \bar{x}_1 = \frac{220}{10} \Rightarrow \bar{x}_1 = 22$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_i}{n_2} \Rightarrow \bar{x}_2 = \frac{27+33+42+35+32+34+38+28+41+43+30+37}{12} \Rightarrow \bar{x}_2 = \frac{420}{12} \Rightarrow \bar{x}_2 = 35$$

x_1	$x_1 - \overline{x_1}$	$(x_1 - \overline{x_1})^2$	x_2	$x_2 - \overline{x_2}$	$(x_2 - \overline{x_2})^2$
20	–2	4	27	–8	64
16	–6	36	33	–2	4
26	4	16	42	7	49
27	5	25	35	0	0
22	0	0	32	–3	9
23	1	1	34	–1	1
18	–4	16	38	3	9
24	2	4	28	–7	49
19	–3	9	41	6	36
25	3	9	43	8	64
220	**0**	**120**	30	–5	25
			37	2	4
			420	**0**	**314**

∴The statistic F is defined by the ratio:

$$F_0 = \frac{{S_1}^2}{{S_2}^2} ------(1)$$

where ${S_1}^2 = \frac{1}{n_1-1}\sum(x_1 - \bar{x}_1)^2 = \frac{120}{9} = 13.33$

$${S_2}^2 = \frac{1}{n_2-1}\sum(x_2 - \bar{x}_2)^2 = \frac{314}{11} = 28.54$$

Since ${S_2}^2 > {S_1}^2$, $F = \frac{L\,arg\,erVariance}{SmallerVariance}$

$$F_0 = \frac{{S_2}^2}{{S_1}^2} = \frac{28.54}{13.33} = 2.14$$

Expected Value:

$F_E = \frac{{S_2}^2}{{S_1}^2}$, follows F-distribution with the degrees of freedom as given below for 5% level of significance:

$v_1 = n_1 - 1 = 10 - 1 = 9, v_2 = n_2 - 1 = 12 - 1 = 11$ is 3.10

Since $F_0 < F_E$ we accept null hypothesis at 5% level of significance and conclude that the two samples may be regarded as drawn from the populations having same variance.

24. The table shows the standard Deviation and Sample Standard Deviation for both men and women. Find the f statistic considering the Men population in numerator.

Population	Population Standard Deviation	Sample Standard Deviation
Men	30	35
Women	50	45

Sol.

Given,

σ_1=Standard deviation of population-1=30

σ_2=Standard deviation of population-2=50

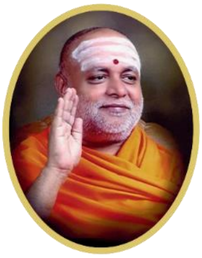
s_1=Standard deviation of sample-1=35

s_2=Standard deviation of sample-2=45

We know that,

$$F = \frac{\frac{{s_1}^2}{{\sigma_1}^2}}{\frac{{s_2}^2}{{\sigma_2}^2}} \Rightarrow F = \frac{\left({35^2}/{30^2}\right)}{\left({45^2}/{50^2}\right)} \Rightarrow F = \frac{\left({1225}/{900}\right)}{\left({2025}/{2500}\right)} \Rightarrow F = \frac{1.3610}{0.81} \Rightarrow F = 1.68$$

***



||JAI SRI GURUDEV||

S.J.C. INSTITUTE OF TECHNOLOGY, CHICKBALLAPUR

Department of Mathematics

LECTURE NOTES

MATHEMATICS-3 FOR COMPUTER SCIENCE STREAM (BCS301)

MODULE - 5

DESIGN OF EXPERIMENTS AND ANOVA

Prepared by:

Purushotham P
Assistant Professor
SJC Institute of Technology
Email id: purushotham@sjcit.ac.in

Experimental unit:

For conducting an experiment, the experimental material is divided into smaller parts and each part is referred to as an experimental unit. The experimental unit is randomly assigned to treatment is the experimental unit. The phrase "randomly assigned" is very important in this definition.

Experiment:

A way of getting an answer to a question which the experimenter wants to know.

Treatment

Different objects or procedures which are to be compared in an experiment are called treatments.

Sampling unit:

The object that is measured in an experiment is called the sampling unit. This may be different from the experimental unit.

Factor:

A factor is a variable defining a categorization. A factor can be fixed or random in nature. A factor is termed as a fixed factor if all the levels of interest are included in the experiment. A factor is termed as a random factor if all the levels of interest are not included in the experiment and those that are can be considered to be randomly chosen from all the levels of interest.

Replication:

It is the repetition of the experimental situation by replicating the experimental unit.

Experimental error:

The unexplained random part of the variation in any experiment is termed as experimental error. An estimate of experimental error can be obtained by replication.

Treatment design:

A treatment design is the manner in which the levels of treatments are arranged in an experiment.

ANOVA:

Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not.

ANOVA stands for Analysis of Variance. It is a statistical method used to analyze the differences between the means of two or more groups or treatments. It is often used to determine whether there are any statistically significant differences between the means of different groups

There are two main types of ANOVA: one-way (or unidirectional) and two-way. There also variations of ANOVA.

Real Life Applications of ANOVA:

- In social sciences, ANOVA tests can be used to study the statistical significance of various study environments on test scores. Medical research. In medical research, the ANOVA test can be used to identify the relationship between various types or brands of medications on individuals with migraines or depression.
- We can use the ANOVA test to compare different suppliers and select the best available. ANOVA (Analysis of Variance) is used when we have more than two sample groups and determine whether there are any statistically significant differences between the means of two or more independent sample groups.

CRD: A completely randomized design (CRD) is one where the treatments are assigned completely at random so that each experimental unit has the same chance of receiving any one treatment.

RBD: A randomized block design is a restricted randomized design, in which experimental units are first organized into homogeneous blocks and then the treatments are assigned at random to these units within these blocks. The main advantage of this design is, if done properly, it provides more precise results.

LSD: The Latin Square Design gets its name from the fact that we can write it as a square with Latin letters to correspond to the treatments. The treatment factor levels are the Latin letters in the Latin square design. The number of rows and columns has to correspond to the number of treatment levels.

ONE WAY CLASSIFICATION:

- Define the problem for different varieties and different treatments.

Verities					**Sum**	**Squares**
x_{11}	x_{12}	x_{13}	…	x_{1n_1}	T_1	${T_1}^2$
x_{21}	x_{22}	x_{23}	…	x_{2n_2}	T_2	${T_2}^2$
x_{31}	x_{32}	x_{33}	…	x_{3n_3}	T_3	${T_3}^2$
---	---	----	…	---		---
x_{k1}	x_{k2}	x_{k3}	…	x_{kn_k}	T_k	${T_k}^2$

- Define the null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n$ for the level of significance.
- Find the sum of all the verities (Row wise) Find the sum of all the contents of N varieties, say T.
- Find the correction factor $CF = \frac{T^2}{N}$
- Find the sum of squares of individual items $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$
- Find the sum of the squares of between the treatments $SST = \sum_i \frac{{T_i}^2}{n_i} - CF$
- Find the sum of squares with in the class or sum of squares due to error by subtraction SEE=TSS-SST.
- Here k represents total number of verities, N represents the total number of observations.
- Plot the ANOVA table

Sources variation	**d.f**	**SS**	**MSS**	**F Ratio**
Between treatments	k-1	SST	$MST = \frac{SST}{k-1}$	$F = \frac{MST}{MSE}$
Error	N-k	SSE	$MSE = \frac{SSE}{N-k}$	
Total	N-1	-	-	

Critical values of F for the 0.05 significance level:

	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.39	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.97	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.97	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.56	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.33	3.47	3.07	2.84	2.69	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.38	2.32	2.28
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.26
25	4.24	3.39	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.17
31	4.16	3.31	2.91	2.68	2.52	2.41	2.32	2.26	2.20	2.15
32	4.15	3.30	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14
33	4.14	3.29	2.89	2.66	2.50	2.39	2.30	2.24	2.18	2.13
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11

PROBLEMS:

1. Three processes A, B and C are tested to see whether their outputs are equivalent. The following observations of outputs are made:

A	10	12	13	11	10	14	15	13
B	9	11	10	12	13	-	-	-
C	11	10	15	14	12	13	-	-

Carry out the analysis of variance and state your conclusion.

Sol. To carry out the analysis of variance, we form the following tables

									Total	Squares
A	10	12	13	11	10	14	15	13	$T_1=98$	$T^2_1=9604$
B	9	11	10	12	13				$T_2=55$	$T^2_2=3025$
C	11	10	15	14	12	13			$T_3=75$	$T^2_3=5625$
Total T									228	-

The squares are as follows

									Sum of Squares
A	100	144	169	121	100	196	225	169	1224
B	81	121	100	144	169				615
C	121	100	225	196	144	169			955
Grand Total - $\sum_i \sum_j {x_{ij}}^2$									2794

Set the null hypotheses $H_0: \mu_1 = \mu_{2=}\mu_3$

Correction Factor $CF = \frac{T^2}{N} = \frac{(228)^2}{19} = \frac{51984}{19} = 2736$

Therefore, Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 2794 - 2736$$
$$\Rightarrow TSS = 58$$

Sum of the squares of between the treatments $SST = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SST = \frac{9604}{8} + \frac{3025}{5} + \frac{5625}{6} - 2736$$
$$\Rightarrow SST = 1200.5 + 605 + 937.5 - 2736$$
$$\Rightarrow SST = 2743 - 2736$$
$$\Rightarrow SST = 7$$

Therefore, sum of squares due to error SEE=TSS-SST

$$\Rightarrow SSE = 58 - 7$$
$$\Rightarrow SSE = 51$$

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**
Between treatments	3-1=2	SST=7	$MST = \frac{7}{2} = 3.5$	$F = \frac{3.5}{3.1875} = 1.0980$
Error	19-3=16	SSE=51	$MSE = \frac{51}{16} = 3.1875$	
Total	19-1=18	-	-	

Since evaluated value 1.0980<3.63 for F(2,16) at 5% level of significance
Hence the null hypothesis is accepted, there is no significance between the three process.

2. A test was given to five students taken at random from the fifth class of three schools of a town. The individual scores are

School I	9	7	6	5	8
School II	7	4	5	4	5
School III	6	5	6	7	6

Carry out the analysis of variance.

Sol.

To carry out the analysis of variance, we form the following tables

						Total	Squares
S1	9	7	6	5	8	T_1=35	T^2_1=1225
S2	7	4	5	4	5	T_2=25	T^2_2=625
S3	6	5	6	7	6	T_3=30	T^2_3=900
Total T=						90	-

The squares are as follows

						Sum of Squares
S1	81	49	36	25	64	255
S2	49	16	25	16	25	131
S3	36	25	36	49	36	182
Grand Total - $\sum_i \sum_j x_{ij}{}^2$						568

Set the null hypotheses $H_0: \mu_1 = \mu_2 = \mu_3$

Correction Factor $CF = \frac{T^2}{N} = \frac{(90)^2}{15} = \frac{8100}{15} = 540$

Therefore, Total sum of squares $TSS = \sum_i \sum_j x_{ij}{}^2 - CF$

$$\Rightarrow TSS = 568 - 540$$
$$\Rightarrow TSS = 28$$

Sum of the squares of between the treatments $SST = \sum_i \frac{T_i{}^2}{n_i} - CF$

$$SST = \frac{1225}{5} + \frac{625}{5} + \frac{900}{5} - 540$$
$$\Rightarrow SST = 245 + 125 + 180 - 540$$
$$\Rightarrow SST = 550 - 540$$
$$\Rightarrow SST = 10$$

Therefore sum of squares due to error SEE=TSS-SST

$$\Rightarrow SSE = 28 - 10 \Rightarrow SSE = 18$$

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**
Between treatments	3-1=2	SST=10	$MST = \frac{10}{2} = 5$	$F = \frac{5}{1.5} = 3.33$
Error	15-3=12	SSE=18	$MSE = \frac{18}{12} = 1.5$	
Total	15-1=14	-	-	

Since evaluated value 3.33<3.63 for F(2,12) at 5% level of significance

Hence the null hypothesis is accepted, there is no significance between the three process.

3. Three different kinds of food are tested on three groups of rats for 5 weeks. The objective is to check the difference in mean weight (in grams) of the rats per week. Apply one-way ANOVA using a 0.05 significance level to the following data:

Food 1	8	12	19	8	6	11
Food 2	4	5	4	6	9	7
Food 3	11	8	7	13	7	9

Sol. To carry out the analysis of variance, we form the following tables

							Total	Squares
F1	8	12	19	8	6	11	T_1=64	T^2_1=4096
F2	4	5	4	6	9	7	T_2=35	T^2_2=1225
F3	11	8	7	13	7	9	T_3=55	T^2_3=3025
Total T							154	-

The squares are as follows

							Sum of Squares
F1	64	144	361	64	36	121	790
F2	16	25	16	36	81	49	223
F3	121	64	49	169	49	81	533
Grand Total - $\sum_i \sum_j x_{ij}^2$							1546

Set the null hypotheses $H_0: \mu_1 = \mu_2 = \mu_3$

Correction Factor $CF = \frac{T^2}{N} = \frac{(154)^2}{18} = \frac{23716}{18} = 1317.55$

Therefore Total sum of squares $TSS = \sum_i \sum_j x_{ij}^2 - CF$

$$\Rightarrow TSS = 1546 - 1317.55$$
$$\Rightarrow TSS = 228.45$$

Sum of the squares of between the treatments $SST = \sum_i \frac{T_i{}^2}{n_i} - CF$

$$SST = \frac{4096}{6} + \frac{1225}{6} + \frac{3025}{6} - 1317.55$$
$$\Rightarrow SST = 682.66 + 204.166 + 504.166 - 1317.55$$
$$\Rightarrow SST = 1391 - 1317.55$$
$$\Rightarrow SST = 73.45$$

Therefore, sum of squares due to error SEE=TSS-SST

$$\Rightarrow SSE = 228.45 - 73.45 \Rightarrow SSE = 155$$

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**
Between treatments	3-1=2	SST=73.45	$MST = \frac{73.45}{2} = 36.725$	$F = \frac{36.725}{10.33} = 3.55$
Error	18-3=15	SSE=155	$MSE = \frac{155}{15} = 10.33$	
Total	18-1=17	-	-	

Since evaluated value 3.55 <3.68 for F(2,15) at 5% level of significance
Hence the null hypothesis is accepted , there is no significance between the three process.

4. Three types of fertilizers are used on three groups of plants for 5 weeks. We want to check if there is a difference in the mean growth of each group. Using the data given below apply a one-way ANOVA test at 0.05 significant level

Fertilizer 1	6	8	4	5	3	4
Fertilizer 2	8	12	9	11	6	8
Fertilizer 3	13	9	11	8	7	12

Sol.
To carry out the analysis of variance, we form the following tables

							Total	Squares
F1	6	8	4	5	3	4	T_1=30	T^2_1=900
F2	8	12	9	11	6	8	T_2=54	T^2_2=2916
F3	13	9	11	8	7	12	T_3=60	T^2_3=3600
Total T							144	-

The squares are as follows

							Sum of Squares
F1	36	64	16	25	9	16	166
F2	64	144	81	121	36	64	510
F3	169	81	121	64	49	144	628
Grand Total - $\sum_i \sum_j {x_{ij}}^2$							1304

Set the null hypotheses $H_0: \mu_1 = \mu_{2=}\mu_3$

Correction Factor $CF = \frac{T^2}{N} = \frac{(144)^2}{18} = \frac{20736}{18} = 1152$

Therefore Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 1304 - 1152$$
$$\Rightarrow TSS = 152$$

Sum of the squares of between the treatments $SST = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SST = \frac{900}{6} + \frac{2916}{6} + \frac{3600}{6} - 1152$$
$$\Rightarrow SST = 150 + 486 + 600 - 1152$$
$$\Rightarrow SST = 1236 - 1152$$
$$\Rightarrow SST = 84$$

Therefore sum of squares due to error SEE=TSS-SST

$$\Rightarrow SSE = 152 - 84 \Rightarrow SSE = 68$$

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**
Between treatments	3-1=2	SST=84	$MST = \frac{84}{2} = 42$	$F = \frac{42}{4.533} = 9.2653$
Error	18-3=15	SSE=68	$MSE = \frac{68}{15} = 4.533$	
Total	18-1=17	-	-	

Since evaluated value 9.2653 >3.68 for F(2,15) at 5% level of significance
Hence the null hypothesis is rejected, there is significance between the tree process.

5. Set an analysis of variance table for the following data.

A	6	7	3	8
B	5	5	3	7
C	5	4	3	4

Sol.

To carry out the analysis of variance, we form the following tables

					Total	Squares
A	6	7	3	8	$T_1=24$	$T_1^2=576$
B	5	5	3	7	$T_2=20$	$T_2^2=400$
C	5	4	3	4	$T_3=16$	$T_3^2=256$
Total T					60	-

The squares are as follows

					Total Squares
A	36	49	9	64	158
B	25	25	9	49	108
C	25	16	9	16	66
Grand Total - $\sum_i \sum_j x_{ij}{}^2$					332

Set the null hypotheses $H_0: \mu_1 = \mu_2 = \mu_3$

Correction Factor $CF = \frac{T^2}{N} = \frac{(60)^2}{12} = \frac{3600}{12} = 300$

Therefore, Total sum of squares $TSS = \sum_i \sum_j x_{ij}{}^2 - CF$

$$\Rightarrow TSS = 332 - 300$$
$$\Rightarrow TSS = 32$$

Sum of the squares of between the treatments $SST = \sum_i \frac{T_i{}^2}{n_i} - CF$

$$SST = \frac{576}{4} + \frac{400}{4} + \frac{256}{4} - 300$$
$$\Rightarrow SST = 144 + 100 + 64 - 300$$
$$\Rightarrow SST = 308 - 300$$
$$\Rightarrow SST = 8$$

Therefore sum of squares due to error SEE=TSS-SST

$$\Rightarrow SSE = 32 - 8$$
$$\Rightarrow SSE = 24$$

Sources variation	d.f.	SS	MSS	F Ratio
Between treatments	3-1=2	SST=8	$MST = \frac{8}{2} = 4$	$F = \frac{4}{2.66} = 1.5037$
Error	12-3=9	SSE=24	$MSE = \frac{24}{9} = 2.66$	
Total	12-1=11	-	-	

Since evaluated value 1.5037<4.26 for F(2,9) at 5% level of significance
Hence the null hypothesis is accepted, there is no significance between the three process.

6. A trial was run to check the effects of different diets. Positive numbers indicate weight loss and negative numbers indicate weight gain. Check if there is an average difference in the weight of people following different diets using an ANOVA Table.

Low Fat	Low Calorie	Low protein	Low carbohydrate
8	2	3	2
9	4	5	2
6	3	4	-1
7	5	2	0
3	1	3	3

Sol.

To carry out the analysis of variance, we form the following tables

	Low Fat	Low Calorie	Low protein	Low carbohydrate	
	8	2	3	2	
	9	4	5	2	
	6	3	4	-1	
	7	5	2	0	
	3	1	3	3	
T	33	15	17	6	71
T^2	1089	225	289	36	-

The squares are as follows

	Low Fat	Low Calorie	Low protein	Low carbohydrate	
	64	4	9	4	
	81	16	25	4	
	36	9	16	1	
	49	25	4	0	
	9	1	9	9	
$\sum_i \sum_j {x_{ij}}^2$	239	55	63	18	375

Set the null hypotheses $H_0: \mu_1 = \mu_{2=}\mu_3$

Correction Factor $CF = \frac{T^2}{N} = \frac{(71)^2}{20} = \frac{5041}{20} = 252$

Therefore Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 375 - 252$$
$$\Rightarrow TSS = 123$$

Sum of the squares of between the treatments $SST = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SST = \frac{1089}{5} + \frac{225}{5} + \frac{289}{5} + \frac{36}{5} - 252$$
$$\Rightarrow SST = 217.8 + 45 + 57.8 + 7.2 - 252$$
$$\Rightarrow SST = 327.8 - 252$$
$$\Rightarrow SST = 75.80$$

Therefore sum of squares due to error SEE=TSS-SST

$$\Rightarrow SSE = 123 - 75.80$$
$$\Rightarrow SSE = 47.2$$

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**
Between treatments	4-1=3	SST=75.80	$MST = \frac{75.80}{3} = 25.26$	$F = \frac{25.26}{2.95} = 8.56$
Error	20-4=16	SSE=47.20	$MSE = \frac{47.20}{16} = 2.95$	
Total	20-1=19	-	-	

Since evaluated value 8.56 >3.24 for $F_{(3,16)}$ at 5% level of significance
Hence the null hypothesis is rejected, there is significance between the four process.

7. The following data show the number of worms quarantined from the GI areas offour groups of muskrats in a carbon tetrachloride anthelmintic study. Conduct a two-way ANOVA test.

I	II	III	IV
33	41	12	38
32	38	35	43
26	40	46	25
14	23	22	13
30	21	11	26

Sol.

Given

I	II	III	IV
33	41	12	38
32	38	35	43
26	40	46	25
14	23	22	13
30	21	11	26

Subtract 30 from all the observations, we get

I	II	III	IV
3	11	-18	8
2	8	5	13
-4	10	16	-5
-16	-7	-8	-17
0	-9	-19	-4

	I	II	III	IV	
	3	11	-18	8	
	2	8	5	13	
	-4	10	16	-5	
	-16	-7	-8	-17	
	0	-9	-19	-4	
T	-15	13	-24	-5	-31
T^2	225	169	576	25	

The squares are as follows

	I	II	III	IV	
	9	121	324	64	
	4	64	25	169	
	16	100	256	25	
	256	49	64	289	
	0	81	361	16	
$\sum_i \sum_j {x_{ij}}^2$	285	415	1030	563	2293

Set the null hypotheses $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

Correction Factor $CF = \frac{T^2}{N} = \frac{(-31)^2}{20} = \frac{961}{20} = 48$

Therefore, Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 2293 - 48$$
$$\Rightarrow TSS = 2245$$

Sum of the squares of between the treatments $SST = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SST = \frac{225}{5} + \frac{169}{5} + \frac{576}{5} + \frac{25}{5} - 48$$

$$\Rightarrow SST = 45 + 33.8 + 115.2 + 5 - 48$$
$$\Rightarrow SST = 199 - 48$$
$$\Rightarrow SST = 151$$

Therefore sum of squares due to error SEE=TSS-SST

$$\Rightarrow SSE = 2245 - 151$$
$$\Rightarrow SSE = 2094$$

Sources variation	d.f.	SS	MSS	F Ratio
Between treatments	4-1=3	SST=151	$MST = \frac{151}{3} = 50.33$	$F = \frac{130.87}{50.33} = 2.6$
Error	20-4=16	SSE=2094	$MSE = \frac{2094}{16} = 130.87$	
Total	20-1=19	-	-	

Since evaluated value 2.6<3.24 for F(3,16) at 5% level of significance
Hence the null hypothesis is accepted, there is no significance between the four process.

TWO WAY CLASSIFICATION:

- Define the problem for different varieties and different treatments.

	Verities					Sum	Squares
	x_{11}	x_{12}	x_{13}	$\cdots$	x_{1n_1}	T_1	${T_1}^2$
	x_{21}	x_{22}	x_{23}	$\cdots$	x_{2n_2}	T_2	${T_2}^2$
	x_{31}	x_{32}	x_{33}	$\cdots$	x_{3n_3}	T_3	${T_3}^2$
	---	---	----	$\cdots$	---	---	---
	x_{k1}	x_{k2}	x_{k3}	$\cdots$	x_{kn_k}	T_k	${T_k}^2$
Sum	P_1	P_2	P_3	---	P_K	=G	
Squares	P_1^2	P_2^2	P_3^2	---	P_K^2		

- Define the null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n$ for the level of significance.
- Find the sum of all the verities (Row wise) Find the sum of all the observations of N varieties, say T.
- Find the correction factor $CF = \frac{T^2}{N}$
- Find the sum of squares of individual items $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$
- Find the sum of the squares of rows $SSR = \sum_i \frac{{T_i}^2}{n_i} - CF$
- Find the sum of the squares of columns $SCC = \sum_i \frac{{P_i}^2}{n_i} - CF$
- Find the sum of squares with in the class or sum of squares due to error by subtraction SEE=TSS-SSR-SSC.
- Plot the ANOVA table

Sources variation	d.f.	SS	MSS	F Ratio
Rows	r-1	SSR	$MSR = \frac{SSR}{r-1}$	$F_r = \frac{MSR}{MSE}$
Columns	c-1	SSC	$MSC = \frac{SSC}{c-1}$	
				$F_c = \frac{MSC}{MSE}$
Error	(r-1)(c-1)	SSE	$MSE = \frac{SSE}{(r-1)(c-1)}$	
Total	N-1	-	-	

1. Set up an analysis of variance table for the following per acre production data for three varieties of wheat, each grown on 4 plots and state it the variety differences are significant at 5% significant level.

Per acre production data			
Plot of land	Variety of wheat		
	A	B	C
1	6	5	5
2	7	5	4
3	3	3	3
4	8	7	4

Sol.

To carry out the analysis of variance, we form the following tables

Per acre production data				T	T^2
Plot of land	Variety				
	A	B	C		
1	6	5	5	16	256
2	7	5	4	16	256
3	3	3	3	9	81
4	8	7	4	19	361
P	24	20	16	=60	-
P^2	576	400	256		

The squares are as follows

Variety			
A	B	C	
36	25	25	
49	25	16	
9	9	9	
64	49	16	Grand Total - $\sum_i \sum_j {x_{ij}}^2$=332

Set the null hypotheses $H_0: \mu_1 = \mu_2 = \mu_3$, N=12

Correction Factor $CF = \frac{T^2}{N} = \frac{(60)^2}{12} = \frac{3600}{12} = 300$

Therefore, Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 332 - 300$$
$$\Rightarrow TSS = 32$$

Sum of the row squares $SSR = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SSR = \frac{256}{3} + \frac{256}{3} + \frac{81}{3} + \frac{361}{3} - 300$$
$$\Rightarrow SSR = 85.33 + 85.33 + 27 + 120.33 - 300$$
$$\Rightarrow SSR = 318 - 300$$
$$\Rightarrow SSR = 18$$

Sum of the column squares $SSC = \sum_i \frac{{P_i}^2}{n_i} - CF$

$$SSC = \frac{576}{4} + \frac{400}{4} + \frac{256}{4} - 300$$
$$\Rightarrow SSC = 144 + 100 + 64 - 300$$
$$\Rightarrow SSC = 308 - 300$$
$$\Rightarrow SSC = 8$$

Therefore SSE=TSS-SSR-SSC

SSE=32-18-8=6

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**
Rows	4-1=3	SSR=18	$MSR = \frac{18}{3} = 6$	$F_r = \frac{6}{1} = 6$
Columns	3-1=2	SSC=8	$MSC = \frac{8}{2} = 4$	
Error	3X2=6	SSE=6	$MSE = \frac{6}{6}$=1	$F_c = \frac{4}{1} = 4$
Total	12-1=11	-	-	

$F_r = 6 >$ F(3,6)=4.76 &

$F_c = 4 <$ F(6,2)=19.33

2. Three varieties of coal were analysed by four chemists and the ash-content in the varieties was found to be as under.

Varieties	Chemists			
	1	2	3	4
A	8	5	5	7
B	7	6	4	4
C	3	6	5	4

Carry out the analysis of variance.

Sol. To carry out the analysis of variance, we form the following tables

Chemists					T	T^2
Variety	1	2	3	4		
A	8	5	5	7	25	625
B	7	6	4	4	21	441
C	3	6	5	4	18	324
P	18	17	14	15	=64	-
P^2	324	289	196	225		

The squares are as follows

Chemists				
1	2	3	4	
64	25	25	49	
49	36	16	16	
9	36	25	16	Grand Total - $\sum_i \sum_j {x_{ij}}^2$=366

Set the null hypotheses $H_0: \mu_1 = \mu_{2=}\mu_3$, N=12

Correction Factor $CF = \frac{T^2}{N} = \frac{(64)^2}{12} = \frac{4096}{12} = 341.33$

Therefore, Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 366 - 341.33$$
$$\Rightarrow TSS = 24.67$$

Sum of the row squares $SSR = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SSR = \frac{625}{4} + \frac{441}{4} + \frac{324}{4} - 341.33$$
$$\Rightarrow SSR = 156.25 + 110.25 + 81 - 341.33$$
$$\Rightarrow SSR = 347.50 - 341.33$$
$$\Rightarrow SSR = 6.17$$

Sum of the column squares $SSC = \sum_i \frac{{P_i}^2}{n_i} - CF$

$$SSC = \frac{324}{3} + \frac{289}{3} + \frac{196}{3} + \frac{225}{3} - 341.33$$
$$\Rightarrow SSC = 108 + 96.33 + 65.33 + 75 - 341.33$$
$$\Rightarrow SSC = 344.66 - 341.33$$
$$\Rightarrow SSC = 3.33$$

Therefore SSE=TSS-SSR-SSC

SSE=24.67-6.17-3.33=15.17

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**
Rows	3-1=2	SSR=6.17	$MSR = \frac{6.17}{2} = 3.085$	$F_r = \frac{3.085}{2.53} = 1.22$
Columns	4-1=3	SSC=3.33	$MSC = \frac{3.33}{3} = 1.11$	
Error	3X2=6	SSE=15.17	$MSE = \frac{15.17}{6}$=2.53	$F_c = \frac{2.53}{1.11} = 2.28$
Total	12-1=11	-	-	

$F_r = 1.22 < F_{(2,6)}$ &

$F_c = 2.28 < F_{(6,3)}$

3. Perform ANOVA and test at 0.05 level of significant whether these are differences in the detergent or in the engines for the following data:

Detergent	Engine		
	I	II	III
A	45	43	51
B	47	46	52
C	48	50	55
D	42	37	49

Sol.

Given the data

Detergent	Engine		
	I	II	III
A	45	43	51
B	47	46	52
C	48	50	55
D	42	37	49

Subtract 45 from all the observations, we get

Detergent	Engine			T	T^2
	I	II	III		
A	0	-2	6	**4**	**16**
B	2	1	7	**10**	**100**
C	3	5	10	**18**	**324**
D	-3	-8	4	**-7**	**49**
P	**2**	**-4**	**27**	**2**	**=25**
P^2	**4**	**16**	**729**	**4**	**-**

The squares are

Detergent	Engine			Sum
	I	II	III	
A	0	4	36	**40**
B	4	1	49	**54**
C	9	25	100	**134**
D	9	64	16	**89**
Grand Total - $\sum_i \sum_j {x_{ij}}^2 =$				**317**

Set the null hypotheses $H_0: \mu_1 = \mu_{2=}\mu_3$, N=12

Correction Factor $CF = \frac{T^2}{N} = \frac{(25)^2}{12} = \frac{625}{12} = 52.08$

Therefore Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 317 - 52.08$$
$$\Rightarrow TSS = 264.92$$

Sum of the row squares $SSR = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SSR = \frac{16}{3} + \frac{100}{3} + \frac{324}{3} + \frac{49}{3} - 52.08$$
$$\Rightarrow SSR = 5.33 + 33.33 + 108 + 16.33 - 52.08$$
$$\Rightarrow SSR = 163 - 52.08$$
$$\Rightarrow SSR = 110.92$$

Sum of the column squares $SSC = \sum_i \frac{{P_i}^2}{n_i} - CF$

$$SSC = \frac{4}{4} + \frac{16}{4} + \frac{729}{4} - 52.08$$
$$\Rightarrow SSC = 1 + 4 + 182.25 - 52.08$$
$$\Rightarrow SSC = 187.25 - 52.08$$
$$\Rightarrow SSC = 135.17$$

Therefore SSE=TSS-SSR-SSC

SSE=264.92-110.92-135.17=18.83

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**
Rows	4-1=3	SSR=110.92	$MSR = \frac{110.92}{3} = 36.97$	$F_r = \frac{36.97}{3.14} = 11.77$
Columns	3-1=2	SSC=135.17	$MSC = \frac{135.17}{2} = 67.58$	
Error	3X2=6	SSE=18.83	$MSE = \frac{18.83}{6}$=3.14	$F_c = \frac{67.58}{3.14} = 21.52$
Total	12-1=11	-	-	

$F_r = 11.77 >$ F(3,6) &
$F_c = 21.52>$ F(6,2)

Since the null hypothesis is rejected and there is a significance between Detergent and Engine.

4. Analyze and interpret the following statistics concerning output of wheat for field obtained as result of experiment conducted to test for Four varieties of wheat viz. A,B,C and D under Laton square design.

C 25	B 23	A 20	D 20
A 19	D 19	C 21	B 18
B 19	A 14	D 17	C 20
D 17	C 20	B 21	A 15

Sol.

Given observations are

C 25	B 23	A 20	D 20
A 19	D 19	C 21	B 18
B 19	A 14	D 17	C 20
D 17	C 20	B 21	A 15

Null hypothesis H_o : There is no significant difference between rows, columns and treatment

Code the data by subtracting 20 from each value, we get

					T	T^2
	C 5	B 3	A 0	D 0	8	64
	A -1	D -1	C 1	B -2	-3	9
	B -1	A -6	D -3	C 0	-10	100
	D -3	C 0	B 1	A -5	-7	49
P	0	-4	-1	-7	=- 12	
P^2	0	16	1	49	-	-

The squares are as follows

C 25	B 9	A 0	D 0	
A 1	D 1	C 1	B 4	
B 1	A 36	D 9	C 0	
D 9	C 0	B 1	A 25	
36	46	11	29	$\sum_i \sum_j {x_{ij}}^2$ =122

Correction Factor $CF = \frac{T^2}{N} = \frac{(-12)^2}{16} = \frac{144}{16} = 9$

Therefore, Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 122 - 9$$
$$\Rightarrow TSS = 113$$

Sum of the row squares $SSR = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SSR = \frac{64}{4} + \frac{9}{4} + \frac{100}{4} + \frac{49}{4} - 9$$
$$\Rightarrow SSR = 16 + 2.25 + 25 + 12.25 - 9$$
$$\Rightarrow SSR = 55.5 - 9$$
$$\Rightarrow SSR = 4$$

Sum of the column squares $SSC = \sum_i \frac{{P_i}^2}{n_i} - CF$

$$SSC = 0 + \frac{16}{4} + \frac{1}{4} + \frac{49}{4} - 9$$
$$\Rightarrow SSC = 4 + 0.25 + 12.25 - 9$$
$$\Rightarrow SSC = 16.5 - 9$$
$$\Rightarrow SSC = 7.5$$

To find the sum of the treatments

Observations					$Q = \sum(Observations)$	Q^2
A	0	-1	-6	-5	-12	144
B	3	-2	-1	1	1	1
C	5	1	0	0	6	36
D	0	-1	-3	-3	-7	49

Sum of the squares of treatments $SST = \sum_i \frac{{Q_i}^2}{n_i} - CF$

$$SST = \frac{144}{4} + \frac{1}{4} + \frac{36}{4} + \frac{49}{4} - 9$$
$$\Rightarrow SST = 36 + 0.25 + 9 + 12.25 - 9$$

$$\Rightarrow SST = 57.50 - 9$$
$$\Rightarrow SST = 48.50$$

$\therefore SSE = TSS - SSR - SSC - SST \Rightarrow SSE = 113 - 46.5 - 7.5 - 48.50 = 10.5$, We know that $F(3,6) = 4.76$

Sources variation	**d.f**	**SS**	**MSS**	**F Ratio**	**Conclusion**
Rows	4-1=3	SSR=46.5	$MSR = \frac{46.5}{3} = 15.5$	$F_r = \frac{15.5}{1.75} = 8.85$	$F_r > F(3,6)$ H_0-Rejected
Columns	4-1=3	SSC=7.5	$MSC = \frac{7.5}{3} = 2.5$	$F_c = \frac{2.5}{1.75} = 1.428$	$F_c < F(3,6)$ H_0-Accepted
Treatments	4-1=3	SST=48.5	$MST = \frac{48.5}{3} = 16.16$	$F_T = \frac{16.16}{1.75} = 9.23$	$F_T > F(3,6)$ H_0-Rejected
Error	3x2=6	SSE=10.5	$MSE = \frac{10.5}{6} = 1.75$	-	-
Total	25-1=24	-	-	-	-

5. Five varieties of paddy A, B, C, D, and E are tried. The plan, the varieties shown in each plot and yields obtained in Kg are given in the following table (LSD)

B 95	E 85	C 139	A 117	D 97
E 90	D 89	B 75	C 146	A 87
C 116	A 95	D 92	B 89	E 74
A 85	C 130	E 90	D 81	B 77
D 87	B 65	A 99	E 89	C 93

Test whether there is a significant difference between rows and columns at 5% LOS.

Sol. Given observations are

B 95	E 85	C 139	A 117	D 97
E 90	D 89	B 75	C 146	A 87
C 116	A 95	D 92	B 89	E 74
A 85	C 130	E 90	D 81	B 77
D 87	B 65	A 99	E 89	C 93

Null hypothesis H_o: There is no significant difference between rows, columns and treatment, Code the data by subtracting 100 from each value, we get

						T	T^2
	B -5	E -15	C 39	A 17	D -3	33	1089
	E -10	D -11	B -25	C 46	A -13	-13	169
	C 16	A -5	D -8	B -11	E -26	-34	1156
	A -15	C 30	E -10	D -19	B -23	-37	1369
	D -13	B -35	A -1	E -11	C -7	-67	4489
P	-27	-36	-5	22	-72	= - 118	
P^2	729	1296	25	484	5184	-	-

The squares are as follows:

B 25	E 225	C 1521	A 289	D 9	
E 100	D 121	B 625	C 2116	A 169	
C 256	A 25	D 64	B 121	E 676	
A 225	C 900	E 100	D 361	B 529	
D 169	B 1225	A 1	E 121	C 49	
775	2496	2311	3008	1432	$\sum_i \sum_j {x_{ij}}^2 = 10022$

Correction Factor $CF = \frac{T^2}{N} = \frac{(-118)^2}{25} = \frac{13924}{25} = 557$

Therefore Total sum of squares $TSS = \sum_i \sum_j {x_{ij}}^2 - CF$

$$\Rightarrow TSS = 10022 - 557$$
$$\Rightarrow TSS = 9465$$

Sum of the row squares $SSR = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SSR = \frac{1089}{5} + \frac{169}{5} + \frac{1156}{5} + \frac{1369}{5} + \frac{4489}{5} - 557$$

$$\Rightarrow SSR = 217.8 + 33.8 + 231.2 + 273.8 + 897.8 - 557$$

$$\Rightarrow SSR = 1654.4 - 557$$

$$\Rightarrow SSR = 1097.4$$

Sum of the column squares $SSC = \sum_i \frac{{P_i}^2}{n_i} - CF$

$$SSC = \frac{729}{5} + \frac{1296}{5} + \frac{25}{5} + \frac{484}{5} + \frac{5184}{5} - 557$$

$$\Rightarrow SSC = 145.8 + 259.2 + 5 + 96.8 + 1036.8 - 557$$

$$\Rightarrow SSC = 1543.6 - 557$$

$$\Rightarrow SSC = 986.6$$

To find the sum of the treatments,

	Observations					$Q = \sum(Observations)$	Q^2
A	17	-13	-5	-15	-1	-17	289
B	-5	-25	-11	-23	-35	-99	9801
C	39	46	16	30	-7	124	15376
D	-3	-11	-8	-19	-13	-54	2916
E	-15	-10	-26	-10	-11	-72	5184

Sum of the squares of treatments $SST = \sum_i \frac{{Q_i}^2}{n_i} - CF$

$$SST = \frac{289}{5} + \frac{9801}{5} + \frac{15376}{5} + \frac{2916}{5} + \frac{5184}{5} - 557$$

$$\Rightarrow SST = 57.8 + 1960.2 + 3075.2 + 583.2 + 1036.8 - 557$$

$$\Rightarrow SST = 6713.2 - 557$$

$$\Rightarrow SST = 6156.2$$

$$\therefore SSE = TSS - SSR - SSC - SST$$

$$\Rightarrow SSE = 9465 - 1097.4 - 986.6 - 6156.2$$

$$\Rightarrow SSE = 1224.8$$

Sources variation	d.f.	SS	MSS	F Ratio	Conclusion
Rows	5-1=4	SSR=1097.4	$MSR = \frac{1097.4}{4} = 274.3$	$F_r = \frac{274.3}{102.66} = 2.672$	$F_r < F(4,12)$ H_0-Accepted
Columns	5-1=4	SSC=986.6	$MSC = \frac{986.6}{4} = 246.65$	$F_c = \frac{246.65}{102.66} = 2.4026$	$F_c < F(4,12)$ H_0-Accepted
Treatments	5-1=4	SST=6156.2	$MST = \frac{6156.2}{4} = 1539.05$	$F_T = \frac{1539.05}{102.66} = 15$	$F_T > F(4,12)$ H_0-Rejected
Error	4x3=12	SSE=1224.8	$MSE = \frac{1224.8}{12}$=102.66		
Total	25-1=24	-	-	-	

6. Present your conclusions after doing analysis of variance to the following results of the Latin-square design experiment conducted in respect of five fertilizers which were used on plots of different fertility.

A 16	B 10	C 11	D 9	E 9
E 10	C 9	A 14	B 12	D 11
B 15	D 8	E 8	C 10	A 18
D 12	E 6	B 13	A 13	C 12
C 13	A 11	D 10	E 7	B 14

Sol. Given observations are

A 16	B 10	C 11	D 9	E 9
E 10	C 9	A 14	B 12	D 11
B 15	D 8	E 8	C 10	A 18
D 12	E 6	B 13	A 13	C 12
C 13	A 11	D 10	E 7	B 14

Null hypothesis H_o: There is no significant difference between rows, columns and treatment,
Code the data by subtracting 10 from each value. We get,

						T	T^2
	A 6	B 0	C 1	D -1	E -1	5	25
	E 0	C -1	A 4	B 2	D 1	6	36
	B 5	D -2	E -2	C 0	A 8	9	81
	D 2	E -4	B 3	A 3	C 2	6	36
	C 3	A 1	D 0	E -3	B 4	5	25
P	16	-6	6	1	14	= 31	
P^2	256	36	36	1	196	-	-

The squares are as follows:

A 36	B 0	C 1	D 1	E 1	
E 0	C 1	A 16	B 4	D 1	
B 25	D 4	E 4	C 0	A 64	
D 4	E 16	B 9	A 9	C 4	
C 9	A 1	D 0	E 9	B 16	
74	22	30	23	86	$\sum_i \sum_j x_{ij}^2 = 235$

Correction Factor $CF = \frac{T^2}{N} = \frac{(31)^2}{25} = \frac{961}{25} = 38.44$

Therefore Total sum of squares $TSS = \sum_i \sum_j x_{ij}^2 - CF$

$$\Rightarrow TSS = 235 - 38.44$$
$$\Rightarrow TSS = 196.56$$

Sum of the row squares $SSR = \sum_i \frac{T_i^2}{n_i} - CF$

$$SSR = \frac{25}{5} + \frac{36}{5} + \frac{81}{5} + \frac{36}{5} + \frac{25}{5} - 38.44$$
$$\Rightarrow SSR = 5 + 7.2 + 16.2 + 7.2 + 5 - 38.44$$
$$\Rightarrow SSR = 40.60 - 38.44$$
$$\Rightarrow SSR = 2.16$$

Sum of the column squares $SSC = \sum_i \frac{{P_i}^2}{n_i} - CF$

$$SSC = \frac{256}{5} + \frac{36}{5} + \frac{36}{5} + \frac{1}{5} + \frac{196}{5} - 38.44$$

$$\Rightarrow SSC = 51.2 + 7.2 + 7.2 + 0.2 + 39.2 - 38.44$$

$$\Rightarrow SSC = 105 - 38.44 \Rightarrow SSC = 66.56$$

To find the sum of the treatments

	Observations					$Q = \sum(Observations)$	Q^2
A	6	4	8	3	1	22	484
B	0	2	5	3	4	14	196
C	1	-1	0	2	3	5	25
D	-1	1	-2	2	0	0	0
E	-1	0	-2	-4	-3	-10	100

Sum of the squares of treatments $SST = \sum_i \frac{{Q_i}^2}{n_i} - CF$

$$SST = \frac{484}{5} + \frac{196}{5} + \frac{25}{5} + \frac{0}{5} + \frac{100}{5} - 38.44$$

$$\Rightarrow SST = 96.8 + 39.2 + 5 + 0 + 20 - 38.44$$

$$\Rightarrow SST = 161 - 38.44$$

$$\Rightarrow SST = 122.56$$

$$\therefore SSE = TSS - SSR - SSC - SST$$

$$\Rightarrow SSE = 196.56 - 2.16 - 66.56 - 122.56$$

$$\Rightarrow SSE = 5.28$$

Sources variation	**d.f.**	**SS**	**MSS**	**F Ratio**	**Conclusion**
Rows	5-1=4	SSR=2.16	$MSR = \frac{2.16}{4} = 0.54$	$F_r = \frac{0.54}{0.44} = 1.227$	$F_r < F(4,12)$ H_0-Accepted
Columns	5-1=4	SSC=66.56	$MSC = \frac{66.56}{4} = 16.64$	$F_c = \frac{16.64}{0.44} = 37.81$	$F_c > F(4,12)$ H_0-Rejected
Treatments	5-1=4	SST=122.56	$MST = \frac{122.56}{4} = 30.64$	$F_T = \frac{30.64}{0.44} = 69.63$	$F_T > F(4,12)$ H_0-Rejected
Error	4x3=12	SSE=5.28	$MSE = \frac{5.28}{12}=0.44$	-	-
Total	25-1=24	-	-	-	-

7. Set up ANOVA table for the following information relating to three drugs testing to judge the effectiveness in reducing blood pressure for three different groups of people:

Group of people	Drug		
	X	Y	Z
A	14 15	10 9	11 11
B	12 11	7 8	10 11
C	10 11	11 11	8 7

Do the drugs act differently? Are the different groups of people affected differently? Is the interaction term significant? Answer the above questions taking a significant level of 5%.

Sol.

Given observations from different people (A, B, C) to the different drugs (X, Y, Z) are as

Group of people	Drug			T	T^2
	X	Y	Z		
A	14 15	10 9	11 11	70	4900
B	12 11	7 8	10 11	59	3481
C	10 11	11 11	8 7	58	3364
P	73	56	58	=187	-
P^2	5329	3136	3364	-	-

Where N=6+6+6=18

Correction Factor $CF = \frac{T^2}{N} = \frac{(187)^2}{18} = \frac{34969}{18} = 1942.722$

The squares are as follows

Group of people	Drug			Sum of Squares	
	X	Y	Z		
A	196 225	100 81	121 121	844	
B	144 121	49 64	100 121	599	
C	100 121	121 121	64 49	576	$\sum_i \sum_j x_{ij}^2$ =2019

Therefore, Total sum of squares $TSS = \sum_i \sum_j x_{ij}^2 - CF$

$$\Rightarrow TSS = 2019 - 1942.722$$
$$\Rightarrow TSS = 76.28$$

Sum of the row squares $SSR = \sum_i \frac{{T_i}^2}{n_i} - CF$

$$SSR = \frac{4900}{6} + \frac{3481}{6} + \frac{3364}{6} - 1942.722$$
$$\Rightarrow SSR = 816.67 + 580.16 + 560.67 - 1942.722$$
$$\Rightarrow SSR = 14.78$$

Sum of the column squares $SSC = \sum_i \frac{{P_i}^2}{n_i} - CF$

$$SSC = \frac{5329}{6} + \frac{3136}{6} + \frac{3364}{6} - 1942.722$$
$$\Rightarrow SSC = 888.16 + 522.66 + 560.67 - 1942.722$$
$$\Rightarrow SSC = 28.77$$

SS within samples (SST)= $(14 - 14.5)^2 + (15 - 14.5)^2 + (10 - 9.5)^2 + (9 - 9.5)^2 + (11 - 11)^2 + (11 - 11)^2 + (12 - 11.5)^2 + (11 - 11.5)^2 + (7 - 7.5)^2 + (8 - 7.5)^2 + (10 - 10.5)^2 + (11 - 10.5)^2 + (10 - 10.5)^2 + (11 - 10.5)^2 + (11 - 11)^2 + (11 - 11)^2 + (8 - 7.5)^2 + (7 - 7.5)^2$

SST=3.50

Therefore,

SSE=TSS-SSR-SSC-SST

$$\Rightarrow \text{SSE} = 76.28 - 14.78 - 28.77 - 3.5$$
$$\Rightarrow \text{SSE} = 29.23$$

We have $F_{(2,9)}$=4.26 , F(4,9)=3.63

Sources variation	d.f.	SS	MSS	F Ratio	Conclusion
Rows	3-1=2	SSR=14.78	$MSR = \frac{14.78}{2} = 7.39$	$F_r = \frac{7.39}{0.389} = 19$	$F_r > F(2,9)$ H_0-Rejected
Columns	3-1=2	SSC=28.77	$MSC = \frac{28.77}{2} = 14.385$	$F_c = \frac{14.385}{0.389} = 37$	$F_r > F(2,9)$ H_0-Rejected
Treatments	9	SST=3.5	$MST = \frac{3.5}{9} = 0.389$	$F_T = \frac{7.33}{0.389} = 18.84$	$F_T > F(4,9)$ H_0-Rejected
Error	4	SSE=29.33	$MSE = \frac{29.33}{4}$=7.33	-	-