Model Question Paper-I (CBCS Scheme)

Third Semester B.E Degree Examination

MATHEMATICS FOR CS ENGINEERING STREAM (BCS301)

TIME: 03 Hours

Max.Marks:100

Note: (i) Answer any FIVE full questions, choosing at least ONE question from each MODULE.

(ii) Statistical tables and Mathematics Formula handbooks are allowed.

Model Paper – 1 Solutions

Module – 1

1.

a) A shipment of 8 similar microcomputers to retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives. Find the mean and variance of this distribution.

Sol^n:

Given 8-3=5 working micro computers

So , there are $8_{C2} = 28$ ways to select 2 computers.

There are $3_{C0} \times 5_{C2} = 10$ ways to select zero defective and 2 working computers.

The probability that to select zero defective computers $P(0) = \frac{10}{28} = 0.3571$

There are $3_{C1} \times 5_{C1} = 15$ ways to select ONE defective and ONE working computers.

The probability that to select one defective computers $P(1) = \frac{15}{28} = 0.5357$

There are $3_{C2} \times 5_{C0} = 3$ ways to select 2 defective and zero working computers.

The probability that to select two defective computers $P(2) = \frac{3}{28} = 0.1071$

No. of defective computers	0	1	2
Probability P(x)	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
0	0.3571	0	0	0
1	0.5357	0.5357	1	0.5357
2	0.1071	0.2142	4	0.4284
$\sum$	-	0.75	-	0.9641

Mean $\mu = E(X) = \sum xP(x) = 0.75$

Variance $\sigma^2 = E(X^2) - \mu^2 = \sum x^2 P(x) - \mu^2 = 0.9641 - (0.75)^2$

$\Rightarrow \sigma^2 = 0.9641 - 0.5625 == 0.4016$

$\Rightarrow S.D. = \sigma = \sqrt{0.4016} = 0.6337$

b) 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 packet 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 defectives 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 defectives out of 10000 packets.

c) If the mileage (in thousands of miles) of a certain radial tyre is a random variable with exponential distribution with mean 40000 miles. Determine the probability that the tyre will last
i) At least 20000 km
ii) At most 30000 km

Sol^n:

Given X be a continuous random variable of an exponential distribution is,

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

and given the mean of exponential distribution is 40000.

$\Rightarrow \mu=40000\Rightarrow\frac{1}{\alpha}=40000\Rightarrow\alpha=\frac{1}{40000}$

$$\therefore f(x)=\begin{cases}\frac{1}{40000}e^{-\frac{x}{40000}} & ,x\geq 0\\ 0 & ,otherwise\end{cases}$$

i) The probability that the radial tyre will last at least 20000 km is,

$$P(x\geq 20000)=\int_{20000}^{\infty}f(x)dx=\int_{20000}^{\infty}\frac{1}{40000}e^{-\frac{x}{40000}}dx$$
$$=\frac{1}{40000}\int_{20000}^{\infty}e^{-\frac{x}{40000}}dx$$

$$= -\left[e^{-\frac{x}{40000}}\right]_{20000}^{\infty} = -\left[0 - e^{-\frac{20000}{40000}}\right] = \frac{1}{\sqrt{e}}$$

ii) The probability that the radial tyre will last at most 30000 km is,

$$P(x \le 30000) = \int_0^{30000} f(x)dx = \int_0^{30000} \frac{1}{40000} e^{-\frac{x}{40000}} dx$$

$$= \frac{1}{40000}\int_0^{30000} e^{-\frac{x}{40000}} dx$$

$$= -\left[e^{-\frac{x}{40000}}\right]_0^{30000} = -\left[e^{-\frac{30000}{40000}} - e^0\right] = 1 - \frac{1}{e^{\frac{3}{4}}}$$

2.

a) The density function of a random variable X is given by

$$f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1 \\ 0, & elsewhere \end{cases},$$

i) Find k

ii) Find the cdf F(x) and use it to evaluate $P[0.3 < X < 0.6]$.

Sol^n:

Given probability function, $f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1 \\ 0, & elsewhere \end{cases}$

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

$$\Rightarrow i) \int_{-\infty}^{0} f(x)dx + \int_0^1 f(x)dx + \int_1^{\infty} f(x)dx = 1$$

$$\Rightarrow 0 + \int_0^1 k\sqrt{x}\,dx + 0 = 1$$

$$\Rightarrow k\int_0^1 \sqrt{x}\,dx = 1$$

$$\Rightarrow \frac{2k}{3}\left[x^{\frac{3}{2}}\right]_0^1 = 1$$

$$\Rightarrow \frac{2k}{3} = 1$$

$$\Rightarrow 2k = 3$$

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

WKT, the cdf of the pdf is,

$$F(x) = \int_{-\infty}^{x} f(u)du$$

$$= \int_{-\infty}^{0} f(u)du + \int_0^x f(u)du$$

$$= 0 + \int_0^x k\sqrt{u}\,du$$

$$= \frac{3}{2} \times \frac{2}{3}\left[u^{\frac{3}{2}}\right]_0^x$$

$$= x^{\frac{3}{2}} = x\sqrt{x}$$

$$P[0.3 < X < 0.6] = F(0.6) - F(0.3)$$

$$= 0.6(\sqrt{0.6}) - 0.3(\sqrt{0.3})$$

$$= 0.4647 - 0.1643$$

$$= 0.3004$$

b) Find the mean and variance of Binomial Distribution.

Sol^n:

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,

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

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:

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

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

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

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

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

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

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

$$= np(1)$$

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

i) Variance:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

c) 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^n:

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:

$$P(x > 2150) = P(z > 1.83)$$
$$= 0.5 - A(1.83)$$
$$= 0.5 - 0.4664 = 0.0336$$

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:

$$P(x < 1950) = P(z < -1.5)$$
$$= 0.5 - A(1.5)$$
$$= 0.5 - 0.4332 = 0.0668$$

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

Module – 2

3.

a) 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

X \ Y	-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

b) Find the unique fixed probability vector of P = $\begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

Sol^n:

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

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

$\therefore QP = [x \quad y \quad z] \begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{bmatrix}$

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

WKT

$QP = Q$

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

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

After solving the above equations. we get,

$\Rightarrow 2x + y = 0, x + 6y = 4, 3x + 4y = -4$ where z=1-x-y

On solving above equations. we get,

$\Rightarrow x = \frac{-20}{7}, y = \frac{8}{7}, z = \frac{19}{7}$

$\therefore Q[x \quad y \quad z] = \left[\frac{-20}{7} \quad \frac{8}{7} \quad \frac{19}{7}\right]$

c) 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.

$\therefore$ 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) $\therefore$ The probability to have a Santro car in the year 2002, $p_S^{(2)} = 1/2 = 50\%$

ii) $\therefore$ The probability to have a Maruthi car in the year 2002, $p_M^{(2)} = 0 = 0\%$

iii) $\therefore$ The probability to have an Ambassador car in the year 2003, $p_A^{(3)} = 1/4 = 25\%$

$\therefore$ The probability to have a Santro car in the year 2003, $p_S^{(3)} = 1/2 = 50\%$

4.

a) The joint probability distribution of two random variables X and Y is:

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

i) Are X and Y independent?

ii) Evaluate $P[Y \leq 2]$

iii) Evaluate $P[X + Y \leq 2]$

Sol^n:

Given,

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

And the probabilities are

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

Given the joint probability distribution is follows as

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

The marginal distribution of X and Y are

x_i	1	3
$f(x_i)$	0.5	0.5

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

$$\mu_X = E(X) = \sum_{i=1}^{2} x_i f(x_i) = (1 \times 0.5) + (3 \times 0.5) = 2$$
$$\mu_Y = E(Y) = \sum_{j=1}^{3} y_j g(y_j) = (-3 \times 0.4) + (2 \times 0.3) + (4 \times 0.3) = 0.6$$

$$E(XY) = \sum_{i=1}^{2}\sum_{j=1}^{3} x_i y_j f(x_i, y_j)$$
$$= (1 \times (-3) \times 0.1) + (1 \times 2 \times 0.2) + (1 \times 4 \times 0.2) + (3 \times (-3) \times 0.3) + (3 \times 2 \times 0.1) + (3 \times 4 \times 0.1)$$
$$= -0.3 + 0.4 + 0.8 - 2.7 + 0.6 + 1.2 = 0$$

$$COV(X,Y) = E(XY) - \mu_X \mu_Y = 0 - (2)(0.6) = 0 - 1.2 = -1.2$$

Hence the given random variables are not independent

$$\therefore P(Y \le 2) = p_{11} + p_{12} + p_{21} + p_{22} = 0.1 + 0.2 + 0.3 + 0.1 = 0.7$$

$$\therefore P(X + Y \le 2) = p_{11} + p_{21} = 0.1 + 0.3 = 0.4$$

b) Define probability vectors, Stochastic matrices, Regular Stochastic matrix, Stationary Distribution and Absorbing state of Markov Chain.

Sol^n:

<u>Probability Vector</u>

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…

<u>Stochastic Matrix</u>

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

<u>Regular Stochastic Matrix</u>

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.

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

c) 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

$\therefore QP = Q$

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

$\Rightarrow \left[\frac{2y}{3} + \frac{2z}{3} \quad x + \frac{z}{3} \quad \frac{y}{3}\right] = [x \quad y \quad z]$

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

$\Rightarrow 3x - 2y - 2z = 0 \ , \ 3x - 3y + z = 0$

$\Rightarrow 3x - 2y - 2(1 - x - y) = 0 \ , \ 3x - 3y + (1 - x - y) = 0$

$\Rightarrow 3x - 2y - 2 + 2x + 2y = 0 \ , \ 3x - 3y + 1 - x - y = 0$

$\Rightarrow 5x = 2 \ , \ 2x - 4y = -1$

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

$\Rightarrow 4y = \frac{9}{5} \Rightarrow y = \frac{9}{20}$

$\Rightarrow z = 1 - x - y \Rightarrow z = 1 - \frac{2}{5} - \frac{9}{20} \Rightarrow z = \frac{3}{20}$

$\therefore Q = [x \quad y \quad z] = \left[\frac{2}{5} \quad \frac{9}{20} \quad \frac{3}{20}\right]$

Thus, the salesman in the long run sells,

$\frac{2}{5}\ in\ city\ A = 40\%$, $\frac{9}{20}\ in\ city\ B = 45\%$, $\frac{3}{20}\ in\ city\ C = 15\%$

Module – 3

5.

a) Define Null Hypothesis, Significance Level, Critical Region, Type-I and Type-II errors in a statistical test.

Sol^n:

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.

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.

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.

b) A coin is tossed 400 times and it turns up head 216 times. Discuss whether the coin may be regarded as unbiased one.

Sol^n:

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.

c) 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^n:

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 $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{180 + 296}{900 + 1600}$$
$$\Rightarrow P = \frac{476}{2500}$$
$$\Rightarrow P = 0.1904 \Rightarrow Q = 1 - P = 1 - 0.1904 = 0.8096$$
$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$\Rightarrow Z = \frac{0.2 - 0.185}{\sqrt{\left(0.1904 \times 0.8096\right)\left(\frac{1}{900} + \frac{1}{1600}\right)}}$$
$$\Rightarrow Z = \frac{0.015}{\sqrt{(0.1541)(0.00173)}}$$
$$\Rightarrow Z = \frac{0.015}{\sqrt{0.00026}}$$
$$\Rightarrow Z = \frac{0.015}{0.01612}$$
$$\Rightarrow Z = 0.9305$$

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.

a) Explain the following terms:

i) Standard Error

ii) Statistical Hypothesis

iii) Critical Region of a statistical test

iv) Test of Significance

Sol^n:

Standard Error:

The standard deviation of the sampling distribution of a statistic is Known as Standard Error (S.E.).

Null Hypothesis (or) Statistical 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.

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.

Test of Significance (or) 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

$$Z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$$

b) 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^n:

The probability of getting 3 or 4 in a single through is $p = {}^2/_6 = {}^1/_3$

And $q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

$\therefore$ Expected number of success $= \frac{1}{3} \times 9000 = 3000$

$\therefore$ 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.

c) 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^n:

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 $P = \frac{x_1+x_2}{n_1+n_2}$

$$\Rightarrow P = \frac{450+450}{600+900}$$

$$\Rightarrow P = \frac{900}{1500} = 0.6$$

$$\Rightarrow P = 0.6 \Rightarrow Q = 1 - P = 1 - 0.6 = 0.4$$

$$\therefore Z = \frac{p_1-p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\Rightarrow Z = \frac{0.75-0.5}{\sqrt{(0.6\times0.4)\left(\frac{1}{600}+\frac{1}{900}\right)}}$$

$$\Rightarrow Z = \frac{0.25}{\sqrt{(0.24)(0.00277)}}$$

$$\Rightarrow Z = \frac{0.25}{\sqrt{0.0006648}}$$

$$\Rightarrow Z = \frac{0.25}{0.02578}$$

$$\Rightarrow Z = 9.69$$

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.

Module – 4

7.

a) 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^n:

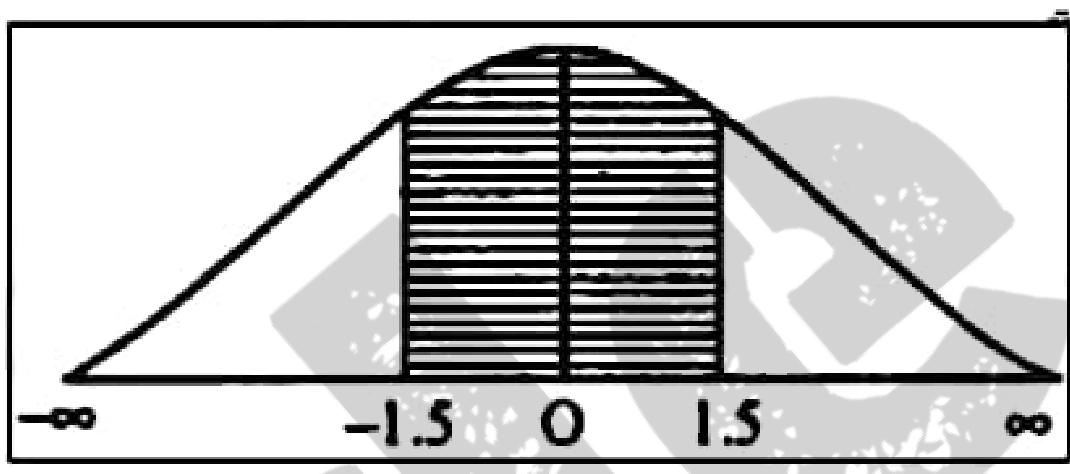
The central limit theorem states that the sample mean $\bar{x}$ follows approximately the normal distribution with mean $\mu$and standard deviation $\frac{\sigma}{\sqrt{n}}$ (is also called Standard error), i.e., $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, where μ , $\sigma$are mean and standard deviation of the population from where the sample.

Given,
Sample size n=100
Mean of the population $\mu = 53$
Variance of the population $\sigma^2 = 400 \Rightarrow \sigma = \sqrt{400} = 20$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\Rightarrow \bar{X} \sim N\left(53, \frac{20}{\sqrt{100}}\right)$$
$$\Rightarrow \bar{X} \sim N(53,2)$$
$$\therefore \text{ we know that } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

$\frac{-53}{\sqrt{100}}$



$$\Rightarrow Z = \frac{\bar{X}-53}{2}$$
$$\therefore \text{ At } \bar{X}=50 \Rightarrow Z = \frac{50-53}{2} = -\frac{3}{2} = -1.5 = z_1$$
$$\text{At } \bar{X}=56 \Rightarrow Z = \frac{56-53}{2} = \frac{3}{2} = 1.5 = z_2$$
$$\therefore P(50 < \bar{X} < 56) = P(-1.5 < z < 1.5)$$
$$= 2P(0 < z < 1.5)$$
$$= 2A(1.5)$$
$$= 2 \times 0.4332$$
$$\therefore P(50 < \bar{X} < 56) = 0.8664$$

b) 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^n:

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 $C.I. = \mu = Mean \pm Z(S\tan d\,ardDeviation/\sqrt{SampleSize} = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$

$$\therefore C.I. = \mu = 78.3 \pm \left(2.58 \times \frac{2}{\sqrt{25}}\right)$$
$$\Rightarrow \mu = 78.3 \pm 1.032$$
$$\Rightarrow C.I \Rightarrow (78.3 - 1.032\,, 78.3 + 1.032) = (77.268\,, 79.332)$$

c) 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^n:

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
$E(x) = 320 \times P(x) = 320 \times 5C_x (0.5)^5$

$$\therefore E(0) = 320 \times P(0) = 320 \times 5C_0 (0.5)^5 = 320 \times (0.5)^5 = 10 = E_0$$
$$E(1) = 320 \times P(1) = 320 \times 5C_1 (0.5)^5 = 320 \times 5 \times (0.5)^5 = 50 = E_{\backslash 1}$$
$$E(2) = 320 \times P(2) = 320 \times 5C_2 (0.5)^5 = 320 \times 5C_2 \times (0.5)^5 = 100 = E_{\backslash 2}$$
$$E(3) = 320 \times P(3) = 320 \times 5C_3 (0.5)^5 = 320 \times 5C_3 \times (0.5)^5 = 100 = E_{\backslash 3}$$
$$E(4) = 320 \times P(4) = 320 \times 5C_4 (0.5)^5 = 320 \times 5C_4 \times (0.5)^5 = 50 = E_{\backslash 4}$$
$$E(5) = 320 \times P(5) = 320 \times 5C_5 (0.5)^5 = 320 \times 5C_5 \times (0.5)^5 = 10 = E_{\backslash 5}$$

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 of χ^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^2 value 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.

8.

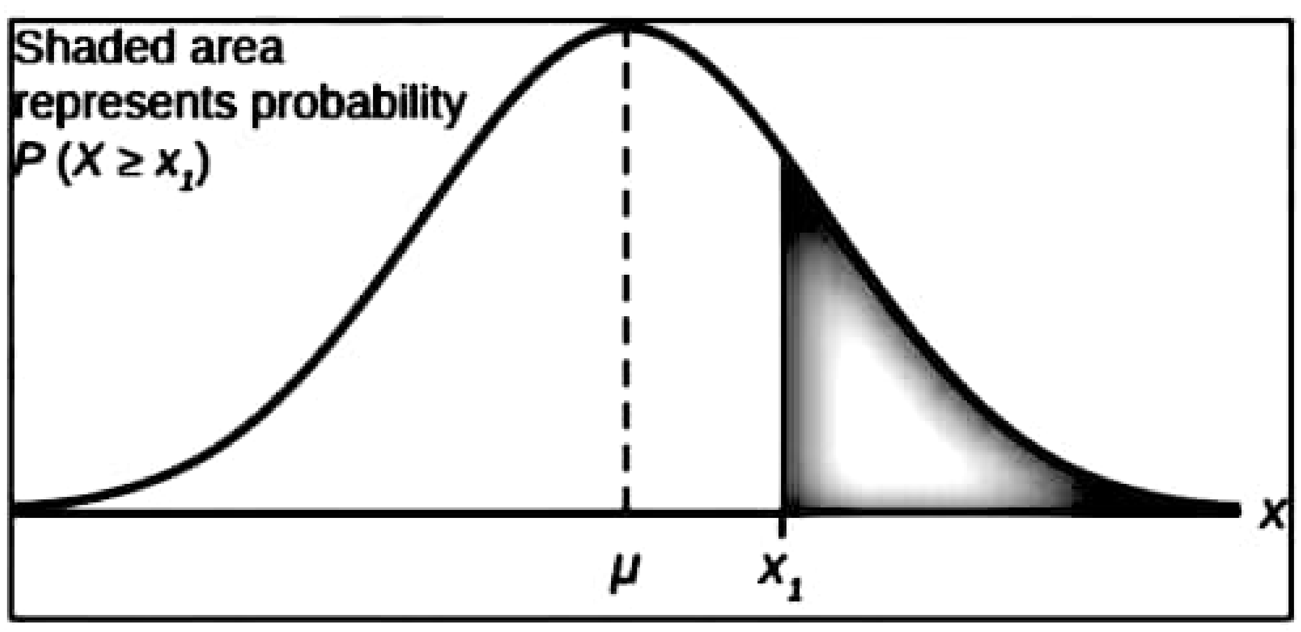
a) A random sample of size 64 is taken from an infinite population having mean 112 and variance 144. Using central limit theorem, find the probability of getting the sample mean $\bar{X}$ greater than 114.5.

Sol^n:

Given,
Sample size n=64
Mean of the population $\mu = 112$
Variance of the population $\Rightarrow \sigma^2 = 144 \Rightarrow \sigma = 12$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\Rightarrow \bar{X} \sim N\left(112, \frac{12}{\sqrt{64}}\right)$$
$$\Rightarrow \bar{X} \sim N(90, 1.5)$$
$$\therefore \text{ we know that } Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$
$$\Rightarrow Z = \frac{\bar{X} - 112}{12/\sqrt{64}}$$
$$\Rightarrow Z = \frac{\bar{X} - 112}{1.5}$$
$$\therefore \text{At } \bar{X} = 114.5 \Rightarrow z = \frac{114.5 - 112}{1.5} = 1.66$$



$$\therefore P(\bar{X} > 114.5) = P(z > 1.66)$$
$$\Rightarrow P(z > 1.66) = 0.5 - P(0 < z < 1.66)$$
$$= 0.5 - 0.4515$$
$$\Rightarrow P(z > 1.66) = 0.0489$$

b) Let the observed value of the mean $\bar{X}$ of a random sample of size 20 from a normal distribution with *mean* μ and variance $\sigma^2 = 80$ be 81.2. Find a 90% and 95% confidence intervals for μ.

Sol^n:

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%:

$$\therefore C.I. = \mu = 81.2 \pm \left(1.645 \times \frac{8.9442}{\sqrt{20}}\right)$$
$$\Rightarrow \mu = 81.2 \pm 3.29$$
$$\Rightarrow C.I = (81.2 - 3.29, 81.2 + 3.29) = (77.91, 84.49)$$

c) 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^n:

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, $s^2 = \frac{1}{n-1}\sum(x - \bar{x})^2$

$$\Rightarrow s^2 = \frac{1}{8}\left\{\begin{matrix}(45-49.11)^2 + (47-49.11)^2 + (50-49.11)^2 + (52-49.11)^2 + (48-49.11)^2 \\ +(47-49.11)^2 + (49-49.11)^2 + (53-49.11)^2 + (51-49.11)^2\end{matrix}\right\}$$

$$\Rightarrow s^2 = \frac{54.9}{8} = 6.8625 \Rightarrow s = \sqrt{6.8625} = 2.6196$$

$\therefore$Thee Null hypothesis $H_0: \mu = 47.5$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$
$$\Rightarrow t = \frac{49.11 - 47.5}{\left(2.6196/\sqrt{9}\right)}$$
$$\Rightarrow t = \frac{1.61}{0.8732}$$
$$\Rightarrow t = 1.8437$$

$\therefore$ 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.

Module – 5

9.

a) 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^n:

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.

b) 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^n:

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	-	-	-	-

10.

a) 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^n:

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

b) 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^n:

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

$$\Rightarrow SSE = 76.28 - 14.78 - 28.77 - 3.5$$
$$\Rightarrow SSE = 29.23$$

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	-	-