PRACTICAL-1
(DMSTT05)
(MSC - STATISTICS)

ACHARYA NAGARJUNA UNIVERSITY

CENTRE FOR DISTANCE EDUCATION

NAGARJUNA NAGAR,

GUNTUR

ANDHRA PRADESH

Practical No – 1(a):

Fitting of truncated Binomial Distribution:-

Six coins are tossed and the no of heads noted in the experiment is repeated 120 times and the following distribution:

No. of heads (X)	Frequency f(x)
1	06
2	19
3	35
4	30
5	23
6	07
Total	120

Fit a truncated Binomial distribution for the above data and test for its goodness of fit.

Aim:- To fit a truncated Binomial distribution for the given data and also for its goodness of fit.

Procedure:- For the given frequency data (x_i, f_i) i=0, 1, 2, - - - - ∞ the probability mass function of truncated Binomial distribution is given by

$$G(x) = \frac{\binom{h}{x} P^x q^{n-x}}{1 - q^n}$$ From the given data mean can be obtained as follows.

$$\bar{x} = \frac{\sum \text{fi xi}}{\sum \text{fi}}; \quad \Sigma\text{fi} = N$$

Here, we know the value of N and not known the value of P, p can be obtained by using M.L.E method the mean of the truncated Binomial distribution is given by

$$\bar{x} = \frac{n\,p}{1 - q^n}$$

Let f(P) = $\bar{x}(1 - q^n)$ – np where q_i = 1- P_i

Here we have to estimate the value p, by using formula (namely it has $\hat{P}$)

i.e., $$P^{(i)} = P^{(i-1)} - \frac{f\left(P^{(i-1)}\right)}{f'\left(P^{(i-1)}\right)}$$

Here $f'(p) = n\bar{x}\, q^{n-1} - n$

Now we have to calculate the expected probability by using of

$$P^{(x+1)} = \frac{n - x}{x + 1}\,\frac{\hat{P}}{1 - \hat{P}}\,P(x)$$

$$P(x) = \frac{n\,\hat{P}(1 - \hat{P})^{n-1}}{1 - (1 - \hat{P})^n}$$

Now using there expected probability. We calculate the expected frequency by using the relation E_i = NP (i)

χ^2 – Test for goodness of fit:-

χ^2 calculated value = $\frac{\sum (0_i - e_i)^2}{e_i} \sim \chi^2_{n-1}$ degrees of freedom. Now we have to compare the χ^2 calculate value with χ^2 tabulated value at $\propto$% los for the given dof.

If χ^2 – calculated value is less than χ^2 tabulated value then we accept the null hypothesis i.e., Hence we conclude that it is not good fit for the given data.

Calculation:-

x	f	fi xi
1	6	6
2	19	38
3	35	105
4	30	120
5	23	115
6	07	42
	fi = 120	Σfi xi = 426

We know that the mean of Binomial distribution is $\bar{x}$ = np

Here n = 6, p = $\frac{\bar{x}}{n} = \frac{3.55}{6} = 0.5917$

Let 'P' be P_0 = 0.5917

q_0= 0.4083

We have $P^{(i)} = P^{(i-1)} - \frac{f(P^{(i-1)})}{f'(P^{(i-1)})}$

$$P^{(1)} = P^{(0)} - \frac{f(P^{(0)})}{f'(P^{(0)})}$$

f(p) = $\bar{x}$(1 - q^n) – n p

f(p^0) = 3.55(1-$(0.4083)^6$) – 6(0.5917)

= 3.55 (1 – 0.0046) – 3.5502

= 3.5537 – 3.5502

= - 0.0165

$f'(p) = n\bar{x}\,q^{n-1} - n$

= 6(3.55). $(0.4083)^5$ – 6

= 21.3(0.0113) – 6

= - 5.7583

$P^{(1)}$ = 0.5917 - $\frac{(0.0165)}{5.7583}$

= 0.5888

$q^{(1)}$ = 1 – P(1)

= 1 – 0.5888

= 0.4112.

$$P^{(2)} = P' - \frac{f(p')}{f'(P')}$$

$$f(p') = \bar{x}\left(1 - (q^1)^n\right) - np'$$

= 3.55(0.9952) – 3.5328

= 0.0002

$f'(p') = n\bar{x}\left((q^1)^{n-1}\right) - n$

= 6(3.55)(0.0118) – 6

= 0.2513 - 6

= - 5.7487

$P^{(2)} = P' - \frac{f(p')}{f'(P')}$

= 0.5888 - $\frac{0.0002}{5.7487}$

= 0.5888

$\hat{q} = 1 - \hat{P}$

= 1 – 0.5888

= 0.4112

Now we calculate the probability by using the given formula.

If x = 0 P(1) = $\frac{n\,\hat{p}\,(1-\hat{p})^{n-1}}{1-(1-\hat{p})^{n}}$

= 0.0417

If x = 1 P(2) = $\frac{n-x}{x+1}\frac{\hat{p}}{1-\hat{p}}P(x)$

= $\frac{5}{2}\left(\frac{0.5888}{0.4112}\right)(0.0417)$

= 0.1493

P(3) = $\frac{6-2}{3}\left(\frac{0.5888}{0.4112}\right)(0.1493)$

= 0.2850

P(4) = $\frac{6-3}{4}\left(\frac{0.5888}{0.4112}\right)(0.2850)$

= 0.3061

P(5) = $\frac{6-4}{5}\left(\frac{0.5888}{0.4112}\right)(0.3061)$

= 0.1753

P(6) = $\frac{6-5}{5}\left(\frac{0.5888}{0.4112}\right)(0.1753)$

= 0.0418

= P(1) + P(2) + P(3) + P(4) + P(5) + P(6)

= 0.0417 + 0.1493 + 0.2850 + 0.1753 + 0.0418 + 0.3061 = 0.9992

Now we have to calculate the expected frequency by using the relation

e_i = N(p(i))

e_1 = 120 (P(1)) = 120 (0.0417) = 5.004 $\simeq 5$

e_2 = NP (2) = 120(0.1493) = 17.9160 $\simeq 18$
e_3 = NP (3) = 120(0.2850) = 34.200 $\simeq 34$
e_4 =NP (4) = 120(0.3061) = 36.7320 $\simeq 37$
e_5 = NP (5) =120(0.1753) = 21.0360 $\simeq 21$
e_6 = NP (6) = 120 (0.0418) = 5.0160 $\simeq 5$

x_i	f_i	e_i	f_i. e_i	$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
1	6	5.004	0.9960	0.9920	0.1982
2	19	17.9160	1.0840	1.1751	0.0656
3	35	34.2000	0.8000	0.6400	0.0187
4	30	36.7320	- 6.7320	45.3198	1.2338
5	23	21.0360	1.9640	3.8573	0.1834
6	07	5.0160	1.9840	3.9363	0.7847
					= 2.4844

χ^2 calculated value is 2.4844
χ^2 tabulated Value is 11.07
$\therefore$ χ^2 cal value < χ^2 tale value
2.4844 < 11.07

Inference :-

Hence, for the given data we observe that χ^2 – calculated value is less than χ^2– tabulated value.

Then we accept null hypothesis i.e., we conclude that the truncated Binomial Distribution is good fit for the given data.

Practical No: - 1(b):-

Fitting of truncated Binomial Distribution:-

In 95 litres of mice the number of litres which contains by 1, 2, 3, 4 mices as recorded below

No. of female mice	No. of Litres.
1	32
2	34
3	24
4	05
Total	95

Fit a truncated Binomial distribution for the above data tests for goodness of fit.

Aim:- To fit the truncated Binomial distribution to the given data and also test the goodness of fit.

Procedure:-

For the given frequency data (x_i, f_i) for i=1, 2, - - - - n the mean of the data can be obtained as follows.

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Hence we have given the give value of n & P is unknown. Now we have to estimate P value from the given data by using M.L.E of can be obtained by using solving the equation is

$$E(X) = \overline{X} = \frac{np}{1-q^n}$$

$f(p) = \bar{x}(1-q^n) - np$ Where q_i = 1- P_i

Now M.L.E of P can be obtained by solving the above equation by Newton Repson method Recursive formula for Newton Rapson method is given by

$$P^{(i)} = P^{(i-1)} - \frac{P(P^i - 1)}{f'(P^i - 1)}$$

Here $f'(p) = n\bar{x}\, q^{n-1} - n$

After obtaining the M.L.E estimate of P namely 'P' now we have to find out the expected probability using the recursive formula,

$$P(x+1) = \frac{n-x}{x+1}\ \frac{\hat{P}}{1-\hat{P}} P(x)$$ Where

$$P(x) = \frac{n\,\hat{P}(1-\hat{P})^{n-1}}{1-(1-\hat{P})^{n}}$$; x =1, 2, - - - - n -1

Now By using these expected probability is P(1), P(2) - - - -P(n) we have to calculate the expected frequencies by using the relation e_i =NP(i)

χ^2 – Test for goodness of fit :-

In order to test the goodness of fit we use the following χ^2- test

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(f_i - e_i)^2}{e}\right] \sim \chi^2_{n-1}$$

at specified level of significance, now we compare the χ^2 – calculated value with χ^2- tabulated value at $\propto$% los for the given dot

If χ^2 – Calculated value is less than χ^2 – tabulated then we accept the null hypothesis i.e., we conclude that the Binomial distribution is good fit for the given data otherwise we reject the null hypothesis and conclude that is not good fit for the given data

Calculation: -

No. of female mices x_i	No. of Litres (f_i)	f_i x_i
1	32	32
2	34	68
3	24	72
4	5	20
Total	Σf_i =95	$\Sigma f_i x_i$ = 192

Now $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{192}{95}$ = 2.0211

We know that $\bar{x}$ = n p

P = $\frac{\bar{x}}{n} = \frac{2.0211}{4}$ = 0.5053

q = 1 – p = 0.4947

Given n =4,   N = 95

$$P^{(i)} = P^{(i-1)} - \frac{f(P^{(i-1)})}{f'(P^{(i-1)})}$$

$$P' = P^0 - \frac{f(P^0)}{f'(P^0)}$$

$$f(p^0) = \bar{x}(1 - q^n) - np_0$$

= 2.0211 [1 – $(0.4947)^4$] – 4[0.5053]

= - 0.1211.

$$f'(p^0) = n\bar{x}.q^{n-1} - n$$

= 4(2.0211) $(0.4947)^3$ – 4

= - 3.0212.

$$P^1 = P^0 - \frac{f(P^0)}{f'(P^0)}$$

P' = 0.4652

q' = 0.5348

$$f(p^1) = \bar{x}(1 - (q^1)^n) - np^1$$

= 2.0211 (1 – $(0.5348)^4$) – 4(0.4652)

= - 0.0050

$$P^2 = P^1 - \frac{f(P^1)}{f'(P^1)}$$

= 0.4652 - $\frac{(0.0050)}{-2.7634}$ = 0.46384

P^2 = 0.46384

1 – P^2 = q^2 = 0.5366

$$f(p^2) = \bar{x}(1 - (q^2)^n) - np^2$$

= 2.0211 (1 – $(0.5366)^4$) – 4(0.4634)

$F(p^{(2)})$ = - 0.00006

$F(p^{(2)})$ = 0.

$$f^1(p^{(2)}) = n\bar{x}(q^{n-1}) - n$$

= - 2.75089

$f^1(p^{(2)})$ = - 2.7059

$$P^3 = P^{(2)} - \frac{f(P^2)}{f'(P^{(2)})}$$

= 0.4634 - $\frac{0}{(-2.7504)}$

P^3 = 0.4634

q^3 = 0.5366

$\therefore \hat{P}$ = 0.4634; $\hat{q}$ = 0.5366

Now we have to calculate the probability by using the given formula

$$P(x+1) = \frac{n-x}{x+1} \frac{\hat{P}}{1-\hat{P}} P(x)$$

$$P(x) = \frac{n\,\hat{P}(1-\hat{P})^{n-1}}{1-(1-\hat{P})^{n}}$$

$$P^{(1)} = \frac{4(0.4634)(1-0.4634)^{4}}{1-(1-0.4634)^{4}}$$

= 0.3123.

$P^{(2)} = \frac{4-1}{2}\,\frac{(0.4634)}{0.5366}\,(0.3123)$

= 0.4045

$P^{(3)} = \frac{4-2}{3}\,\frac{0.4634}{0.5366}\,(0.4045)$

= 0.2329

$P^{(4)} = \frac{4-3}{4}\,\frac{(0.4634)}{0.5366}\,(0.2329)$

= 0.0503

$P^{(x)} = P^{(1)} + P^{(2)} + P^{(3)} + P^{(4)} = 1$

Now we calculate the expected frequency by using the given formula

$e_i = NP_i$

e_1 = 95 (0.3123) = 29.6685 $\simeq 30$

e_2 = 95 (0.4045) = 38.4275 $\simeq 38$

e_3 = 95 (0.2329) = 22.1255 $\simeq 22$

e_4 = 95(0.0503) = 4.7785 $\simeq 5$

x_i	f_i	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
1	32	2.3315	5.4359	0.1832
2	34	- 4.4275	19.6028	0.5101
3	24	1.8745	3.5138	0.1588
4	05	0.2215	0.0491	0.0103
Total				= 0.8624

$\therefore \chi^2$ calculated value is 0.8624

χ^2 tabulated Value is 7.81 at 5%. Level of Significance.

$\therefore \chi^2$ calculated value < χ^2 tabulated value i.e., 0.8624 < 7.81.

Inference:-

Hence from the given data. We observe that χ^2 calculated value < χ^2 tabulated value. Hence we accept null hypothesis i.e., we conclude that the truncated Binomial distribution is good fit for the given data.

Practical – 2(a)

Fitting a truncated Poission distribution:-

To fit a truncated Poission distribution to the following data with respect to the real blood of corpuscular (x) per cell.

x	No. of cells
1	148
2	64
3	27
4	05
5	01
Total	250

And also tests the goodness of fit.

Aim:- To fit a truncated Poission distribution for the given data and also test for its goodness of fit.

Procedure:- The Probability of truncated Poission distribution is

$$G(x) = \frac{e^{-\lambda}\, \lambda^{x}}{x!} \Big/ 1 - e^{-\lambda}$$

Mean of the data can be obtained as follows

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Mean of the truncated Poission distribution is $\bar{x} = \frac{\lambda}{1-e^{-\lambda}}$, here λ is unknown value. Now, we have to estimate λ from given data by using M.L.E method.

In truncated Poission distribution M.L.E of λ can be obtained by the iterative procedure.

The recursive formula for a Newton Rapson method is given by

$$P(\lambda^{i}) = P(\lambda^{(i-1)}) - \frac{f(\lambda^{(i-1)})}{f^{1}(\lambda^{1(-1)})}$$

Where $f(\lambda) = \overline{X}(1-e^{-\lambda}) - \lambda$

$$f'(\lambda) = \overline{X}\, e^{-\lambda} - 1$$

After obtaining the M. L E of λ namely λ. We have to find out the expected Probability and the expected Probability are obtained as follows.

$$P(x+1) = \frac{\hat{\lambda}}{x+1} P(x) \quad = \frac{\hat{\lambda}e - \hat{\lambda}}{1-e^{-\lambda}}$$

Now we have to find out the expected frequencies using the relation

$e_i = N.\, P_i$ where $N = \Sigma f_i$

Test for goodness of fit:-

In order to test the goodness of fit. We use the χ^2 – test the test statistic is $\chi^2 = \frac{\sum (f_i - e_i)^2}{e_i} \sim \chi^2_{(n-1)} df$ At specified level of significance. We will compare. The χ^2 – calculated value with χ^2- tabulated value at ∝ %. Los for the given df.

If χ^2 cal value is less than χ^2 tab value we accept the null hypothesis. We conclude that the test is good fit otherwise we reject the null hypothesis and conclude that the test is not suitable for the given data.

Calculation:-

x_i	f_i	$f_i x_i$
1	148	148
2	69	138
3	27	81
4	5	20
5	1	05
Total	Σf_i =250	= 392

$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$ = $\frac{392}{250}$ = 1.5680

Let $\lambda = \lambda_0 = P_0$ = 1.5680

$f(\lambda) = \overline{X}\left(1 - e^{-\lambda}\right) - \lambda$

=1.5680 (1 – $e^{-1.5680}$) – 1.5680

= - 0. 3269

$f'(\lambda)$= (1.5680) $e^{-1.5680}$ -1

= 0.6731

$P^{(1)} = P^{(0)} - \frac{f(\lambda)}{f'(\lambda)}$= 1.5680 - $\left(\frac{-0.3269}{-0.6731}\right)$= 1.0823

$f(\lambda) = \overline{X}\left(1 - e^{-\lambda_1}\right) - \lambda_1$

= 1.5680 (1- $e^{-1.0823}$) – 1.0823

= - 0.0458

$f'(\lambda_1)$= 1.5680 ($e^{-1.0823}$) -1

= - 0.4688

P(2) = 1.0823 -$\frac{(-0.0458)}{-0.4688}$

λ_2 = 0.9846

f(λ_2) = 1.5680 (1 – $e^{-0.9846}$) – 0.9846

= - 0.0024

$f'(\lambda_2)$= 1.5680 ($e^{-0.9846}$) -1

= 0.4142

$P^{(3)} = P^{(2)} - \frac{f(\lambda_2)}{f'(\lambda_2)}$

= 0.9846 - $\frac{0.0024}{-0.4142}$ - 0.9788

f(λ_3) = 1.5680 (1 - $e^{-0.9788}$) – 0.9788

= 0.000005588

$f'(\lambda_3)$ = 1.5680 ($e^{-0.9788}$) -1

= - 0.4108

$$P^{(4)} = P^{(3)} - \frac{f(\lambda_3)}{f'(\lambda_3)}$$

$= 0.9788 - \frac{0.00000}{0.4108}$

$= 0.9788$

$\hat{\lambda} = 0.9788$

Now we have to calculated the probability by using the given formula

$$P(x+1) = \frac{\hat{\lambda}}{x+1} \frac{e^{-\lambda}\lambda^x}{x!}$$

$P(1) = \frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}} = \frac{0.9788(e^{-0.9788})}{1-e^{-0.9788}} = 0.5892$

$P(2) = \frac{0.9788}{2}(0.5892) = 0.2884$

$P(3) = \frac{0.9788}{2+1}(0.2884) = 0.0941$

$P(4) = \frac{0.9788}{2+2}(0.0941) = 0.0230$

$P(5) = \frac{0.9788}{5}(0.0230) = 0.0045$

Now we have to calculate the expected frequency by using the relation $e_i = NP_i$

$e_1 = 250\ (0.5892) = 147.3000 \simeq 147$

$e_2 = 250\ (0.2884) = 72.100 \simeq 72$

$e_3 = 250\ (0.0941) = 23.525 \simeq 24$

$e_4 = 250\ (0.0230) = 5.7500 \simeq 6$

$e_5 = 250\ (0.0045) = 1.1250 \simeq 1$

Now calculate χ^2 values for testing the goodness of fit for the following data.

x_i	f_i	e_i	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
1	148	147	1	1	0.0668
2	69	72	-3	9	0.1250
3	27	24	3	9	0.3750
4	5	6	1	1	0.1669
5	1	1	0	0	0
				= 20	= 0.6735

From the above table we have

χ^2 – Calculated value = 0.6735

χ^2 –Tabulated value at 5% los = 9.49

0.6735 < 9.49

i.e., $\chi^2 \le \chi^2$ tab

Inference:-

Hence, from the given data by using fitting of truncated poission distribution. We observe that $\chi^2 \leq \chi^2$ tab i.e., 0.6735 < 9.49 at 5% los. Hence we accept the null hypotheses is and we conclude that the truncated poission distribution is good fit for the given data.

Practical No: - 2(b):-

Fitting of truncated Poission distribution:-

In a city of 200 diabetics effected family are taken and the following data is the distribution of the families with respect to the no. of diabetics patients in each family by using truncated Poission distribution

X	No. of family
1	113
2	51
3	24
4	09
5	03
Total	200

And also tests the goodness of fit

Aim:- To fit the truncated Poission distribution and also test the goodness of fit.

Procedure:- The probability mass function of truncated Poission distribution. Truncated at origin is given by

$$G(x) = \frac{1}{1-e^{-\lambda}} \frac{e^{-\lambda}\lambda^x}{x!} \text{; x =1, 2, - - - - } \infty$$

$$0 \quad \text{; otherwise}$$

With mean = $\frac{\lambda}{1-e^{-\lambda}}$ from the given frequency data (x_i, f_i) I =1, 2,3 - - - - ∞ the mean of the data can be obtained as follows where $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \lambda$

Here λ is the unknown quantity we have to estimate λ from the given by using M. L. E method In truncated poission distribution M. L. E of λ can be obtained by solving the following by using iteration formula.

$$f(\lambda) = \overline{X}\left(1-e^{-\lambda}\right) - \lambda$$

$$f'(\lambda) = x\,e^{-\lambda} - 1$$

M.L.E of λ can be obtained by solving the equation (1) by using Newton rapson's method the recursive formula for Newton Raphson method.

$$P^{(i)} = P^{(i-1)} = \frac{f\left(\lambda^{((i)-1)}\right)}{f'\left(\lambda^{(i-1)}\right)}$$

$$P^{(i)} = P^{(i-1)} = \frac{f\left(\lambda^{i-1}\right)}{f'\left(\lambda^{i-1}\right)}$$

Here i = 0, 1,2, - - -- , then we get $\hat{\lambda}$ when which is equal to $\bar{x}$ often obtaining the M.L.E of λ namely λ we have to find out the expected probabilities and expected frequencies. The expected probabilities are obtained as follows

$$P(x+1) = \frac{\hat{\lambda}}{x+1} P(x)$$

P(1) = $\dfrac{\hat{\lambda}\, e^{-\lambda}}{1 - e^{-\lambda}}$

After obtaining P(1), P(2), - - - we have to obtain expected $e_i = Np_i$; N $=\Sigma f_i$ frequencies by using the relation $E_i = Np_i$ where N = Σf_i.

Test for goodness of fit:-

In order to test the goodness of fit. We use χ^2 – test the statistic is $\chi^2 = \dfrac{\sum(f_i x_i)}{e_i} \sim \chi^2_{n-1}$ degrees of freedom.

At specified level of significance. We will compare the χ^2 – calculated values with χ^2 – tabulated value of $\propto$ % los for the given dof.

If χ^2 cal $\le \chi^2$ tab value we accept the null hypothesis is and we conclude that the test is the good fit otherwise we reject null hypothesis and we conclude that the test is not suitable for the given data.

Calculation: -

x_i	f_i	$f_i\ x_i$
1	113	113
2	51	102
3	24	72
4	9	36
5	3	15
Total	= 200	= 338

From the above table

$$\overline{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{338}{200} = 1.6900$$

Let $\lambda = \lambda_0 = P_0 = 1.6900$

$f(\lambda) = \overline{x}\,(1 - e^{-\lambda}) - \lambda_0$

= 1.69000 (1 – $e^{-1.6900}$) -1.6900

= - 0.3118

$f^1(\lambda) = 1.6900\ e^{-1.6900} - 1$

= - 0.6882.

$$P^{(1)} = P^{(0)} - \frac{f(\lambda)}{f^1(\lambda)}$$

= 1.6900 - $\dfrac{0.3118}{0.6881}$ = 1.2369

$f(\lambda_1) = \overline{x}\,(1 - e^{-\lambda_1}) - \lambda_1$

= 1.6900 (1 – $e^{-1.2369}$) - 1.2369

= - 0.0375

$f^1(\lambda_2) = \overline{x}\ e^{-\lambda_1} - 1$

= 1.6900 ($e^{-1.2369}$) -1

= - 0.5094

$$P^{(2)} = P^{(1)} - \frac{f(\lambda_1)}{f^1(\lambda_1)}$$

= 1.2369 - $\frac{0.0375}{0.5094}$

=1.1633

f(λ_2) =1.6900($e^{-1.1633}$) – 1.1633

= 0.0013

$f^1(\lambda_2)$ = 1.6900 ($e^{-1.1633}$) -1

= - 0.4720

$P^{(3)}$ = 1.1633 - $\frac{0.0013}{0.4720}$

= 1.1605

f(λ3) = 1.6900 (1 – $e^{-1.1605}$) – 1.1650

= 0.0000

$f^1(\lambda_3)$ = 1.6900($e^{-1.1605}$) – 1

= - 0.4705

$P^{(4)}$ = 1.1605 - $\frac{0.0000}{0.4705}$

$\hat{\lambda}$=1.1605

Now the expected probability are obtained as follows

$$P^{(x+1)} = \frac{\hat{\lambda}}{x+1} P(x)$$

$$= \frac{\hat{\lambda} - e^{-\hat{\lambda}}}{1 - e^{-\hat{\lambda}}}$$

$$= \frac{1.1605.\, e^{-1.1605}}{1 - \mathrm{e}^{-1.1605}} = \frac{0.3636}{0.6867} = 0.5295$$

P(2) = $\frac{1.1605}{2}$ (0.5295) = 0.3072

P(3) = $\frac{1.1605}{3}$ (0.3072) = 0.1189

P(4) = $\frac{1.1605(0.01189)}{4}$ = 0.0345

P(5) = $\frac{1.1605(0.0345)}{5}$

= 0.0080

P(x) = P(1) + P(2) + P(3) + P(4) + P(5)

= 0.998 $\simeq 1$

Now we calculate expected frequencies by using the relation $e_i = Np_i$

$e_1 = 200 \times 0.5295 = 105.9 \simeq 106$
$e_2 = 200 \times 0.3072 = 61.44 \simeq 61$
$e_3 = 200 \times 0.1189 = 23.78 \simeq 24$
$e_4 = 200 \times 0.0345 = 6.900 \simeq 7$
$e_5 = 200 \times 0.0080 = 1.6 \simeq 2$
Now we are testing the goodness of fit

$$\chi^2 = \sum_{i=1}^{5} \frac{(f_i - e_i)^2}{e_i} \sim \chi_4^2$$

x_i	f_i	e_i	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
1	113	106	7	49	0.4623
2	51	61	-10	100	1.6393
3	24	24	0	0	0
4	9	7	2	4	0.5714
5	3	2	1	1	0.5
					=3.1730

χ^2 cal value = 3.1730
χ^2 tab value at 5% Los and 9.49
χ^2 cal value < χ^2 tab value
i.e., 3.1730 < 9.49.

Inference :-

Hence, from the given data by using fitting of truncated Poission distribution we observe that
χ^2 cal value $\leq \chi^2$ tab value
i.e., 3.1730 $\leq$ 9.49 at 5% Los. 30, we accept null hypothesis and we conclude that the truncated Poission distribution is good fit for the given data.

Practical No: - 3

Fitting of Laplace or Double exponential distribution

The distribution of age at the marriage of groups with brides of the following age group

Age group	No. of group
15 - 19	08
19 – 23	25
23 – 27	42
27 – 31	18
31 – 35	07

Fit a Laplace distribution for the given data and also test whether fit is good or not

Aim: - To fit the Laplace distribution and also test the goodness of fit.

Procedure: - The probability density function of a Laplace distribution with location parameter μ; and scale parameter θ is given by

$$f(x_i, \mu, \theta) = \begin{cases} 1 - \frac{1}{2} e^{-\left|\frac{x-\hat{\mu}}{\theta}\right|} & ; x \geq \hat{\mu} \\ \frac{1}{2} e^{-\left|\frac{\hat{\mu}-x}{\hat{\theta}}\right|} & ; x < \mu \end{cases}$$

Here our Problem is to find the minimum likely hood estimates of μ & θ when M.L.E of $\mu = \hat{\mu} = M_d$ of the given frequencies distribution

$$\therefore \ \hat{\mu} = \text{median} = \frac{1 + \frac{N}{2} - c.f}{f} \times C \quad \text{Where}$$

N = Total frequency
L = Lower Limits of the median class
C = Class interval
C. f = Cumulative frequency of the C.I

M.L.E of θ = $\hat{\theta} = \frac{1}{N} \sum f_i / Z_i - m_d /$ where Z_i=Mid value of class interval, Now we have to calculate the value of $f(x_i)$ where x_i is the upper limit of the i^{th} interval

$$f(x_i, \mu, \hat{\theta}) = 1 - \frac{1}{2} e^{-\left|\frac{x-\hat{\mu}}{\theta}\right|} \quad ; x \geq \hat{\mu}$$

$$= \frac{1}{2} e^{-\left|\frac{\hat{\mu}-x}{\hat{\theta}}\right|} \quad ; x < \hat{\mu}$$

The expected frequency of the Laplace distribution are obtained by $e_i = N \, \Delta \, F(x_i)$ where

$\Delta \, F(x_i) = f(x_{i+1}) - f(x_i)$

Test for goodness of fit:-

The null hypothesis is tested here is whether the Laplace distribution is good fit for the given data on not the test statistic is

$$\chi^2 = \frac{\sum (o_i - e_i)^2}{e_i} \sim \chi^2_{n-m-1} \text{ where}$$

M – no. of observation Pooled

If χ^2 cal value < χ^2 tab then we accept the null hypothesis otherwise we reject null hypothesis and we conclude that the Laplace distribution is not suitable for the given data.

Calculation:-

f_i	Frequency	Cumulative frequency	Z_i=Mid value	$\lvert Z - m_d \rvert$	$f_i[z_i - m_d]$
15 – 19	8	8	17	7.6190	60.9520
19 – 23	25	23	21	3.6190	90.4750
23 – 27	42	75	25	0.3610	16.0020
27 – 31	18	93	29	4.3810	78.8580
31 – 35	7	100	33	8.3810	58.6670

$$\text{Median} = l + \frac{\left[\frac{N}{2} - C.F\right]}{f} \times C$$

$$= 23 + \frac{50 - 33}{42} \times C$$

$\hat{\mu}$ = 24.6190

M.L.E of θ = $\hat{\theta}.\frac{1}{N}\sum f_i[z_i - m_d]$

$= \frac{1}{100} \times 304.9540$

$= 3.0495$

Then we have to find out the values

$f(x_i, \mu, \hat{\theta}) = 1 - \frac{1}{2} e^{-\left|\frac{x-\hat{\mu}}{\theta}\right|} \quad ; x \geq \hat{\mu}$

$= \frac{1}{2} e^{-\left|\frac{\hat{\mu}-x}{\hat{\theta}}\right|} \quad ; x < \hat{\mu}$

Then we have that value of x, are

15, 19, 23, 27, 31, 35, 50

x = 15

$f(x_i) = \frac{1}{2} e^{-\left|\frac{24.6190-15}{3.0495}\right|}$

$= 0.0213$

x = 19

$f(x_i) = \frac{1}{2} e^{-\left|\frac{24.6190-19}{3.0495}\right|}$

$= 0.0792$

x = 23

$f(x_i) = \frac{1}{2} e^{-\left|\frac{24.6190-23}{3.0495}\right|}$

$= 0.2940$

x = 27

$f(x_i) = 1 - \frac{1}{2} e^{-\left|\frac{-24.6190+27}{3.0495}\right|}$

$= 0.7710$

X=31

$f(x_i) = 1 - \frac{1}{2} e^{-\left|\frac{-24.6190-31}{3.0495}\right|}$

$= 0.9383$

X = 35

$f(x_i) = 1 - \frac{1}{2} e^{-\left|\frac{35-24.6190}{3.0495}\right|}$

$= 0.9834$

$f(\infty) = 1 - \frac{1}{2} e^{-\left|\frac{x-\hat{\mu}}{\hat{\theta}}\right|}$

$= 1 - e^{-\infty} = 1$

Now we have to find out the Δ x $f(x_i)$ values

Age group	f_i	Upper limit	F(xi)	$\Delta f(x_i) = f_{(xi+1)} - f_{(xi)}$	$e_i = N \Delta f(x)$
- 2 – 15	-	15	0.0213	-	-
15 – 19	80	19	0.0792	0.0579	$\simeq 6$
19 – 23	25	23	0.2940	0.2148	$\simeq 21$
23 – 27	42	27	0.7710	0.4770	$\simeq 48$
27 – 31	18	31	0.9383	0.1673	$\simeq 17$
31 – 35	7	35	0.9834	0.0451	$\simeq 5$
35 - ∞	-	∞	1.0000	0.0166	$\simeq 2$

Pooled observation is 1 i.e., m =1

The χ^2 table is

O_i	e_i	$(O_i - e_i)^2$	$\frac{(o_i - e_i)^2}{e_i}$
08	6	4	0.6667
25	21	16	0.7619
42	48	36	0.75
18	17	1	0.0588
07	7	0	0
			= 2.2374

$$\chi^2 = \frac{\sum (O_i - e_i)^2}{e_i} \sim \chi^2_{\text{n - m -1}} \text{ d. f}$$

$$\sim \chi^2_{5\text{-}1\text{-}1} = \chi^2_3 \text{ d. f}$$

χ^2 cal value = 2.2374

χ^2 tab value at 5% los is = 7.81

χ^2 cal value < χ^2 tab value

Inference: -

Hence, from the given data by using fitting of Laplace distribution we observe that

χ^2 cal value < χ^2 tab value i.e.,

2.2379 < 7.81 at 5% ls. So, we accept the null hypothesis at 3 dof and we conclude that the Laplace distribution is good fit for the given data.

Practical No:- 4

Fitting of logistic distribution:-

Fit a logistics distribution to the following data and obtain respected logistic frequencies

Class Interval	Frequencies
11 – 13	08
13 – 15	24
15 – 17	42
17 – 19	05

19 – 21	36
21 – 23	16
23 – 25	09

And also test for goodness of fit.

Aim: - To fit a logistic distribution for the given data and also test the goodness of fit.

Procedure:- The probability density function of a logistic distribution with parameter ∝ and β is given by

$$f(x) = \frac{e^{-x}}{\left(1+e^{-x}\right)^2}$$

$$f(x,\alpha,\beta) = \frac{1}{\beta}\frac{e^{-\left(\frac{x-\alpha}{\beta}\right)}}{1+\beta^{-\left(\frac{x-\alpha}{\beta}\right)^2}}$$

Cumulative distribution function of f(x, ∝ ,β) = $\frac{1}{1+e^{-\left(\frac{x-\alpha}{\beta}\right)}}$ ⇒this is the logistic distribution. Here our problem is to find out the M.L.E of ∝,β now we have to estimate the parameter of ∝, β is given as follows.

$\hat{\alpha}$ = mean of the given frequency distribution

$\hat{\alpha} = \frac{\sum f_i x_i}{\sum f_i}$ Where Z_i= mid value of frequency distribution

$\hat{\beta}$ = Standard deviation of the given data.

$$\hat{\beta} = \sqrt{\frac{1}{N}\left[\sum f_i z_i^2 - N\hat{\alpha}^2\right]}$$

By substitution of $\hat{\alpha}$, $\hat{\beta}$ in P. d. f we get the logistic distribution for the given data obtain the expected frequencies first we have to conclude to compute

$$f(x_i) = \frac{1}{1+e^{-\left(\frac{x-\hat{\alpha}}{\hat{\beta}}\right)}}$$

Where $\hat{x}_i$ is the upper limit of the class interval then expected frequencies are obtained by using this selection

$e_i = N \Delta f(x_i)$ $\quad \Delta f(x_i) = f_{(xi+1)} + f(x_i)$

Goodness of fit :-

If the null hypothesis "H_0" is accepted then we may conclude that the given logistic distribution is good. Fit. Otherwise we reject the null hypotheses and we conclude that it is not good fit for the given data.

The test statistic is $\chi^2 = \frac{\sum (O_i - e_i)^2}{e_i} \sim \chi^2_{n-m-1}$

Where m is the pooled frequency, when χ^2 cal value < χ^2 tab value. We accept the null hypothesis otherwise we reject the null hypothesis.

Calculation:-

x_i	f_i	Mid Value	$d_i = \frac{x_i - A}{n}$	$f_i\ d_i$	$f_i\ d_i^2$
11 – 13	8	12	- 3	- 24	72
13 – 15	24	14	- 2	- 48	96
15 – 17	42	16	- 1	- 42	42
17 – 19	65	18	0	0	0
19 – 21	36	20	1	36	36
21 – 23	16	22	2	32	64
23 – 25	9	24	3	27	87

$$\hat{\alpha} = A + \frac{\sum f_i d_i}{\sum f_i} \times C$$

$$= 18 + \frac{(-19)}{200} \times 2$$

$$\hat{\alpha} = 1781$$

$$\hat{\beta} = \sqrt{\frac{1}{N}\left(\sum f_i d_i\right)^2 - \left(\frac{\sum f_i d_i}{N}\right)^2}$$

$$= \sqrt{\frac{391}{200} - \left(\frac{19}{200}\right)^2}$$

$$\hat{\beta} = 1.3950$$

$$f(x) = \frac{1}{1 + e^{-\left(\frac{x - \hat{\alpha}}{\hat{\beta}}\right)}} = 0.0075$$

$$f(13) = \frac{1}{1 + e^{-\left(\frac{13 - 17.8}{1.3950}\right)}} = 0.307$$

$$f(15) = \frac{1}{1 + e^{-\left(\frac{15 - 17.8}{1.3950}\right)}} = 0.1177$$

$$f(17) = \frac{1}{1 + e^{-\left(\frac{17 - 17.8}{1.3950}\right)}} = 0.3588$$

$$f(19) = \frac{1}{1 + e^{-\left(\frac{19 - 17.8}{1.3950}\right)}} = 0.7012$$

$$f(21) = \frac{1}{1 + e^{-\left(\frac{21 - 17.8}{1.3950}\right)}} = 0.9078$$

$$f(23) = \frac{1}{1 + e^{-\left(\frac{23 - 17.8}{1.3950}\right)}} = 0.9763$$

$$f(25) = \frac{1}{1 + e^{-\left(\frac{25 - 17.8}{1.3950}\right)}} = 0.9943$$

$$f(\infty) = \frac{1}{1 + e^{-\left(\frac{\infty - 17.8}{1.3950}\right)}} = 1.$$

Age group	f_i	Upper limit	$f(x_i)$	$\Delta f(x_i)$	$e_i = N \Delta f(x_i)$
$-\infty$ - 11	-	11	0.0075	0.0233	4.66 $\simeq 5$
11 – 13	8	13	0.0307	0.0869	17.38 $\simeq 17$
13 – 15	24	15	0.1177	0.241	48.22 $\simeq 48$
15 – 17	42	17	0.3588	0.3424	68.5 $\simeq 69$
17 – 19	65	19	0.7012	0.2066	41.32 $\simeq 41$
19 – 21	36	21	0.9078	0.0685	13.7 $\simeq 14$
21 – 23	16	23	0.9763	0.0180	3.6 $\simeq 4$
23 – 25	9	25	0.9943	0.0057	1.14 $\simeq 1$
25 - ∞	-	∞	1		= 199

Now we are fitting goodness of fit for the given data

O_i	e_i	$O_i - e_i$	$(O_i - e_i)$	$\frac{(O_i - e_i)^2}{e_i}$
8	5	3	9	1.8
24	17	7	49	2.8824
42	48	- 6	36	0.75
65	69	- 4	16	0.2319
36	41	- 5	25	0.6098
16	14	2	4	0.2857
9	5	4	16	3.2
				= 9.7598

Here Polled observation is 1. i.e., m =1

$\chi^2_{n - m - 1} = \chi^2_{7 - 1 - 1} = \chi^2_{51} = 11.07$

χ^2 tab value at 5% los at 5 df is 11.07

χ^2 tab value is 9.7598. χ^2 cal < χ^2 tab value then we accept null hypothesis H_0.

Inference:-

Hence from the given data by fitting of logistic distribution we observe that χ^2 cal < χ^2 tab i.e., 9.7598 < 11.07. Then we accept the null hypothesis H_0. And hence the given logistic distribution is good fit for the given data.

Practical No:- 5(a)

Fitting of multinational distribution

In a Biology experiment making of two red – eyed fruit flies produced x = 432 off spring, among which 253 were red – eyed. 69 were brown – eyed. 87 were scarlet – eyed and 23 were white – eyed, using $\propto$ = 0.05 test the hypothesis that the ratio among the offspring follows that the ratio 9:3:3:1(known as Mendals hird law).

Aim:- to fit the multinational distribution for the given data and also test the (Mendel's third law goodness of fit)

Procedure:- Suppose we have observes (x_1, x_2, - - - - -x_k) as the outcomes of multinational experiment consists of n trails (i.e., x_1t_1, x_2t_2, - - - +x_k) and the probability distribution of (x_1, x_2, - - -x_n) is $\mu_k(n_1, p_1, p_2, - - - p_k)$ our goal is to test

N.H:- H_0: (P_1 - - - -P_2) = (P_{10}, P_{20} - - - - P_{k0}) against

A. H: -H_1: (P_1 - - - P_k) ≠ (P_{10}, P_{20}, - - - - P_{k0}); where (P_{10}, P_{20}, - - - - P_{k0}) is a given set of probability of it possibilities outcomes in a single trail (Such that P_{10} + - - - -+P_{k0} =1). The test statistic to check if the data (x_1, x_2, - - - -x_k) really comes from an Δ GFT =

$$\frac{\left(X_1 - np_{10}\right)^2}{np_{10}} + - - - - - - + \frac{\left(X_k - np_{k0}\right)^2}{np_{k0}}$$

$$= \frac{\sum_{i=1}^{k}\left[(\text{No. of times i}^{\text{th}}\text{ outcomes appears) - expected no. of i}^{\text{th}}\text{ outcomes if } H_0 \text{ is time true})\right]}{\text{Expected no.of i}^{\text{th}}\text{ out come if } H_0 \text{ is true}}$$

The subscript GFT in $^{\Delta}$GFT stands for goodness of fit test – the probability distribution of $^{\Delta}$GFT. If H_0 is true for can be approximate by the $\chi^2_{(k-1)}$- cure thus if $\Delta_{GFT} \le \chi^2_{(R)}$ then reject H_0 (i.e., accept H_A)

$^{\Delta}$GFT ≤ $\chi^2_{(k-1),\alpha)}$ then accept H_0 (i.e., reject H_A where ∝ is significance level)

Calculation:-

Define the experiment as observation the eye color of 432 fruit files,

Note that:-

1. N =no. of trails = 432(observing each off spring)
2. The trails are independent (Assuming that all of 6 spring inherit the eyes – closed independently) and identical
3. If an off spring is Choosen randomly. Then its eye – colosed could be either red (R) or brown (B) or Scoulet (S), or while (w) and
4. The probability are P_1 =P(R); P_2 =P(B), P_3=P(S) and P_4 =P(W) the experiment is an

$\mu_4 = m_4$ = (432, P_1, P_2, P_3, P_4) experiment of Mendel's law holds then

$P_1 = \frac{9}{16}$, $P_2 = \frac{3}{16}$, $P_3 = \frac{3}{16}$, $P_4 = \frac{1}{16}$, thus

We test

H_0: (P_1, P_2, P_3, P_4) = $(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16})$ the test static is computer through the following table.

Computation of ΔGFT for the data in above given problem.

Categories	O	e	$(o – e)^2/e$
R	O_1 =253	e_1 = 243	0.4115
B	O_2 =69	e_2 = 81	1.7772
S	O_3 =87	e_3 = 81	0.444
W	O_4 =23	e_4 = 27	0.5926
Total	n = 432	n = 432	ΔGFT =3.2263

$E_1 = nP_1$

$= 432(\frac{9}{16}) = 243$

$E_2 = nP_2 = 432(\frac{3}{16}) = 81$

$E_3 = nP_3 = 432(\frac{3}{16}) = 81$

$E_4 = nP_4 = 432(\frac{1}{16}) = 27$

The test statistic value ΔGFT is compared with

$\chi^2_{(k-1),\alpha} = \chi^2_{(R)}$ = 7.81 =30.05

Since $\Delta_{GFT} < 7.81$, we accept H_0

(i.e., reject H_A)

Inference:-

Hence, from the given data by using fitting of multinomial distribution we observe that ΔGFT = 3.2263 and since. $\Delta_{GFT} < 7.81$ Then we accept H_0 (i.e., reject HA). There tore the observed data fit (or support) Mender's ratio of 9:3:3:1 for categories R.B.S and W

Practical No:- 5(b)

Fitting of multinomial Distribution:-

A decade ago a city's day time traffic composition of private passenger. Vehicles (PPV) light commercial vehicles (LCV) and Heavy commercial vehicles (HCV) was approximately 40%, 35%, and 25%. Respectively. three independent surveys were conducted by three agencies to study whether this composition is still the same, the survey result are given in the following Drawn a conclusion at 5% los survey data on a city current daytime traffic composition.

Traffic category	Survey 1	Survey – 2	Survey - 3
PPV	436	520	376
LCV	391	401	281
HCL	297	319	191
Total	1124	1240	848

Aim:-

To fit a multinomial distribution for the given data, and also test the goodness of fit

Procedure:-

Suppose we have absolved (x_1, x_2, - - - x_n) as the outcome of multinomial experiment consisting of n – trails (i.e., x_1+x_2+ -- - - - -$+x_k = n$) and the probability distribution of (x_1, x_2, - - - - x_k) is M_n (n, p_1, p_2,- - - -P_k) our good is to test.

Null hypothesis:-

H_0: (P_1, - - - P_k) = (P_{10}, P_{20}, - - - P_{k0}) against alternative hypothesis H_A: (P_1 - - - P_K) ≠ (P_{10}, P_{20}, - - - - P_{k0}) is a given set of probabilities of K possibilities of k outcomes in a single trail (such that $P_{10} + P_{20} +$ - - - -P_{10} =1) the test statistic to check if the data [x_1, x_2, - - -x_k] really comes from an M_k(n, P_{10}, - - - -P_{k0}) distribution is $\Delta GFT = \frac{(X_1 - np_{10})^2}{np_{10}} + - - - - - - + \frac{(X_k - np_{k0})^2}{np_{k0}}$

$$= \frac{\sum_{i=1}^{k}\left[\left(\text{No. of times i}^{\text{th}}\text{ outcomes appears}\right) - \text{expected no. of i}^{\text{th}}\text{ outcomes if } H_0 \text{ is time true}\right)\right]}{\text{Expected no.of i}^{\text{th}}\text{ out come if } H_0 \text{ is true}}$$

The subscript "GFT" in ∆GFT stands for.

Goodness of fit test:-

The probability distribution of ∆GFT if H_0 is true can be approximately by the $\chi^2_{(k-1)}$- curve $\chi^2_{(k-1)}$- Curve with (k -1) d f thus if

$\Delta_{GFT} > \chi^{2(R)}_{(k-1),\alpha}$ then reject H_0

$\Delta_{GFT} > \chi^{2(R)}_{(k-1),\alpha}$ then accept H_0. Whether ∝ is los

Here we have three repeated experiments conducted by three agencies in each experiment

1. nj- no. of trails (each trail is observing the type of each vechical)
   n = 1124. n_2= 1240 n_3 =848
2. All trails with in an experiment are independent and identical. Also experiments are identical (or similar)
3. In each trail for each experiment there are possible outcomes PPV, LCV, and HCV
4. In each trail, the probabilities of 3 possible are

P_1 = P(PPV)

P_2 = P(LCV)

P_3 = P(HCV)

Thus, the three surveys are

$M_3(1124, P_1, P_2, P_3)$, M_3 $(1240, P_1, P_2, P_3)$, and $M_3(848, P_1, P_2, P_3)$ respectively. we want to test

H_0:(P_1, P_2, P_3) = (0.40, 0.35, 0.25) against

H_A:$(P_1, P_2, P_3) \neq$ (0.40, 0.35, 0.25)

If H_0 is accepted, then the three survey indicate that the traffic composition has remained same on the other hand. If H_0 is rejected then the other hand if H_0 is rejected then it shows that the traffic composition by changed in the following we

Take 16% to incorporate the multiple composition of Δ_{GFT} for the data in problem

P_1 = 0.40

P_2 = 0.35

P_3 = 0.25

The test statistic value is

$\Delta^{\text{Pooled}}_{\text{GFT}} = \Delta^{(1)}_{\text{GFT}} + \Delta^{(2)}_{\text{GFT}} + \Delta^{(3)}_{\text{GFT}}$

= 1.337 +3.9318 +6.9138

= 12.1826

Which is Row compared with

$\chi^{2(R)}_{(k-1),\alpha} = \chi^{2(R)}_{(6.005)}$ = 12.592.

Categ ories	$O^{(1)}$	$O^{(2)}$	$O^{(3)}$	$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$\frac{(O^{(1)}-E^{(1)})^2}{E^{(1)}}$	$\frac{(O^{(2)}-E^{(2)})^2}{E^{(2)}}$	$\frac{(O^{(3)}-E^{(3)})^2}{E^{(3)}}$
PPV	436	520	376	4449.6	496	339.2	0.4114	1.1013	3.9925
LCV	391	401	281	393.4	434	296.8	0.0146	2.5092	0.8411
HCV	297	319	191	281	310	212	0.9110	0.2613	2.0802
Total	1124	1240	848	1124	1240	848	1.337	3.9318	6.9138

PPV $\Rightarrow$ $E^{(1)}_{PPV}$ = n_1P_1 = 1124(0.40)= 449.6 $\quad$ $E^{(2)}_{PPV}$ = n_2P_1 = 1240(0.40)= 496 $\quad$ $E^{(3)}_{PPV}$ = n_3P_1 = 848(0.40)= 339.6

LCV $\Rightarrow$ $E^{(1)}_{LCV}$ = n_1P_2 = 1124(0.35)= 393.4 $\quad$ $E^{(2)}_{LCV}$ = n_2P_2 = 1240(0.35)= 434 $\quad$ $E^{(3)}_{LCV}$ = n_3P_2 = 848(0.35)= 296.8

HCV $\Rightarrow$ $E^{(1)}_{HCV}$ = n_1P_3 = 1124(0.25)= 281 $\quad$ $E^{(2)}_{HCV}$ = n_2P_3 = 1240(0.25)=310 $\quad$ $E^{(3)}_{HCV}$ = n_3P_3 = 848(0.25)= 212

Since $\Delta^{Pooled}_{GFT} < \chi^{2(R)}_{l(k-1)1\alpha}$ we accept H0 at level of 0.05

Inference:-

Here Δ^{Pooled}_{GFT} = 12.1826

$\chi^{2(R)}_{l(k-1)1\alpha}$ = 12.592

Since $\Delta^{Pooled}_{GFT} < \chi^{2(R)}_{l(k-1)1\alpha}$ we accept H_0 at level 0.05 this means that the survey result support the null hypothesis i.e., the traffic composition has remained the same.

Application of two dimensional Random variable selecting a committee

ADHOC committee 3 is selected randomly from pool of 10 students consisting of 3 seniors and 3 juniors 2 hostlers , 2 day/scholar. Let x be the no. of seniors and y be the no. of juniors selected let us compute marginal functions. Then find

(i) P((0 < x ≤ 2), y =3) (ii) P(0 < x ≤ 2, y =1) (iii) P(x ≥ 1)

(iv) P(x = 2, 1≤ Y≤3) (v) P(x =3, 2 ≤ Y < 3)

Aim:-

To compute marginal probabilities for the given data

Procedure:

Clearly, these are ${}^{10}C_3$ = 120 ways such a committee and each is assigned the same probability $= \frac{1}{120}$

New P(x =i, y = j) = $\frac{n(i,j)}{120};\ \begin{matrix} i = 0,1,2,3 \\ j = 0,1,2,3 \end{matrix}$

Where n(i, j) is the no. of ways of choosing 3 seniors (0ut of 3) j juniors (out of 3) 3 – i – j day scholars as Hostels (out of 4)

$$n(i,j) = \binom{3}{i}\binom{3}{j}\binom{4}{3-i-j}$$

Calculation:-

If the following that the joint probability function of (x, y). the contingency table is

X / y	0	1	2	3
0	$\frac{4}{120}$	$\frac{18}{120}$	$\frac{12}{120}$	$\frac{1}{120}$
1	$\frac{18}{120}$	$\frac{36}{120}$	$\frac{9}{120}$	0
2	$\frac{12}{120}$	$\frac{9}{120}$	0	0
3	$\frac{1}{120}$	0	0	0

It is easy to write joint distribution function of (x, y)

P(0< x ≤ 2, Y=3) = $\sum_{x} P(X = x, y = 3)$

= P(x =0, Y=3) +P(x=1, y=3) + P(x = 2, y = 3)

= $\frac{1}{120}$ + 0 + 0 = $\frac{1}{120}$

P(0< x ≤ 2, Y=1) = $\sum_{x=0}^{2} P(X = x, y = 1)$

= P(x =0, Y=1) +P(x=1, y=1) + P(x = 2, y = 1)

= $\frac{18}{120}$ + $\frac{36}{120}$ + $\frac{9}{120}$ = $\frac{63}{120}$

P(x ≥ 1) = 1 – P(x <1)

= 1 –P(x=0)

= 1 – 4/120

= $\frac{116}{120}$ = $\frac{85}{120}$

P(x=2, 1< y≤3) = $\sum_{x} P(x = 2, Y = y)$

= P(x=2, Y=1) + P(x=2, y=2) +P(x=2, y=3)

= $\frac{9}{120}$ + 0 + 0 = $\frac{9}{120}$

P(x = 3, 2 ≤ y < 3) = $\sum_{y} P(x = 3, Y = y)$

= P(x= 3, Y= 2)

Conclusion:-

The marginal function of x and y is

x	0	1	2	3
P(X = x)	$\frac{35}{120}$	$\frac{63}{120}$	$\frac{21}{120}$	$\frac{1}{120}$
Y	0	1	2	3
P(Y=y)	$\frac{35}{120}$	$\frac{63}{120}$	$\frac{21}{120}$	$\frac{1}{120}$

And the joint distributions are

P((0< x ≤ 2) y= 3) =0

P(0 ≤ x ≤ 2, y=1) = $\frac{45}{120}$

P(x ≥1) = $\frac{85}{120}$

P(x=2, 1≤Y≤3) = $\frac{9}{120}$

P(x =3, 2 ≤ Y<3) = 0

Additive in Gassoline:-

Let X and Y be the proportion of two different additive in sample taken from a certain brand of gasoline suppose joint density of (x, y) is given by

f(x, y) =2; $\begin{matrix} 0 \le x \le 1 \\ 0 \le y \le 1 \end{matrix}; 0 \le x + y \le 1$

= 0 otherwise

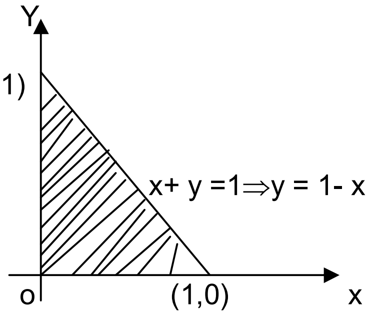
Then find the P ($\frac{1}{2} \le y \le \frac{7}{8} \Big/ x = \frac{1}{3}$)

Aim:-

To find the joint and marginal, conditional distribution of given data.

Procedure:-

By symmetry X and Y marginal densities and comman density is obtained by integrating f over the shaded triangle in the following figure indeed for 0 ≤ x≤1



$f_1(x) = f_2(x) = \int_{-\infty}^{\infty} f(x,y)dy$

$= \int_0^{1-x} 2dy$

= 2(1-x)

And f (x) = 0 elsewhere

Calculations:-

The conditional density of Y given X = x given by

$$g(y/x) = \frac{g(y/x)}{g(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x}; \; 0 \le y \le 1-x$$

= 0 ; otherwise

For 0 ≤ x ≤1 in particular if x = 1/3 then

$$g(Y/1/3) = \frac{2}{2\left(1-\frac{1}{3}\right)} = \frac{3}{2}; \; 0 \le Y \le 2/3$$

= 0 ; otherwise

and $P\left(\frac{1}{2} \le Y \le 7/8/x - 1/3\right) = \int_{1/2}^{7/8} g(1/1/3)dy$

$$= \frac{3}{2}\int_{1/2}^{7/8} dy = \frac{3}{2}\int_{1/2}^{2/3} dy$$

$$= \frac{3}{2}\left[\frac{7}{8} - \frac{1}{2}\right] = \frac{1}{4}$$

f(x, y) =2; 0< x < 1

= 0; otherwise

Find marginal & conditional density function

1. f(y) = $\int_x f(x,y)dy$

$= \int_y^1 2dx$ = 2 [1- y]

2.f(x) = $\int_y f(x,y)dy$

= $\int_0^x 2dy$ = 2x

The conditional probability x/y is

f(x / y) = $\frac{f(x,y)}{f(y)} = \frac{2}{2(1-y)} = \frac{1}{1-y}$; 0 < y < x <1

Similarly

f(y / x) = $\frac{f(x,y)}{f(x)} = \frac{2}{2x} = \frac{1}{x}$; 0 < y < x <1

Conclusion:-

The conditional distribution of

$$P(\frac{1}{2} \le Y \le 7/8 \Big/ x = 1/3) = \frac{1}{4}$$

Practical No: 6

2^3 Factorial Experiments

The following table gives the layout and the request of 2^3 factorial designs laid out in 4replications. The purpose of the experiment is to determine the effect of different kinds of fertilizer Nitrogen (N), potash (K) and phosphorous (P) on potato crop yield.

2^3 factorial experiment laid out in 4 (blocks)

Rep – I	291	391	312	373	101	265	106	450
Rep – II	407	324	272	306	89	449	338	106
Rep – III	323	87	321	423	334	279	128	471
Rep – IV	361	272	103	324	302	131	437	445

Obtain the main effect and interaction effect. Also analyse the data and draw conclusions.

Aim:-

For the given factorial 2^3 experiment, obtain the main effect and interaction effect. Analyse the data and draw conclusion.

Procedure:-

To get main effect and interaction effect by using YATE's (method) or (YATE's algorithm).

*We take first column as treatment in standard order.

*We take next column denoting it as 0^{th} column (c_o) by filling the first column with the corresponding treatment totals sum over replications

* We fill up the next column denoting it as column 1 (c_1) by filling the first half by pair wise addition of 0^{th} column and next half with pair wise subtraction and similarly for 2^{nd}, 3^{rd} columns and so on.

*We continue this procedure upto column c^3 (i.e., up to no. of factors)

*The last column is filled up by contrast sum of squares by using

Formula = $\frac{c_3^2}{2^3.r}$

Replicate sum of square = $\frac{\sum R_i^2}{2^3}$ – CF

CF = $\frac{G^2}{2^3.r}$

Total sum of squares T = $\sum\sum y_{ij}^2$ – CF

The null hypothesis

H_{01} : Main effect of N is not significant

H_{02} : Main effect of K is not significant

H_{03} : Main effect of P is not significant

H_{04} : Interaction effect of NK is not significant

H_{05} : Interaction effect of NP is not significant

H_{06} : Interaction effect of KP is not significant

H_{07} : Interaction effect of NPK is not significant

H_{08} : Replication effect is not significant

The ANOVA Table:-

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F_{cal}	F_{Ta}
Replications	(r – 1)	R = $\frac{R_i^2}{2^3}$ - CF	R′ = R / (r – 1)	R′ / E′	$F_{(r-1)}$
Treatment	7				
Main effect N	1	N = $[N]^2/2^3$	N′ = N / 1	N′ / E′	$F_{(1, *)}$
Main effect K	1	K = $[K]^2/2^3$	K′ = K / 1	K′ / E′	$F_{(1, *)}$
Main effect P	1	P = $[P]^2/2^3$	P′ = P / 1	P′ / E′	$F_{(1, *)}$
Interaction effect NK	1	NK = $[NK]^2/2^3$	(NK)′ = NK / 1	(NK)′/E′	$F_{(1, *)}$
Interaction effect NP	1	NP = $[NP]^2/2^3$	(NP)′ = NP / 1	(NP)′/E′	$F_{(1, *)}$
Interaction effect KP	1	KP = $[KP]^2/2^3$	(KP)′ = KP / 1	(KP)′/E′	$F_{(1, *)}$
Interaction effect NKP	1	NKP = $[NKP]^2/2^3$	(NKP)′=NKP/1	(NKP)′/E′	$F_{(1, *)}$
Error	*	**	E′ = $\frac{**}{*}$		
Total	$(2^3r – 1)$	TSS			

* = $(2^3 r - 1) - (r - 1) - 7$

** = TSS – R – N – K – P – NK – NP – KP – NKP

Conclusion:-

If F – calculated value < The F – Table value we accept the null hypothesis otherwise we reject H_0.

Standard order	C_0	C_1	C_2	C_3	Contrast sum of squares
1	425	851	3172	9324	2716780.5
n	426	2321	6152	-340	3612.5
K	1118	2679	-86	-2264	160178.0
nK	1203	3473	-254	112	392
P	1283	-1	-1470	-2980	277512.5
nP	1396	-85	-794	168	882
kP	1666	-113	84	-676	14280.5
nKP	1807	-141	28	56	98

Rep – I	291	391	312	373	101	265	106	450	2289
Rep – II	407	324	272	306	89	449	338	106	2291
Rep – III	323	87	321	423	334	279	128	471	2369
Rep – IV	361	272	103	324	302	131	437	445	2375

Correction factor CF = $\frac{G^2}{2^3.r}$

$= \frac{(9324)^2}{2^3.4}$ = 2716780.5

Replicate sum of squares $= \frac{\sum R_i^2}{2^3} - CF$

$= \frac{(2289)^2 + (2291)^2 + (2369)^2 + (2375)^2}{8}$ = 843

Total sum of square

TSS $= \sum\sum y_{ij}^2 - CF$

= 465337.5

ANOVA Table:

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F_{cal}	F_{Ta}
Replications	3	843	281	0.7827	3.02
Treatment	7				
Main effect N	1	3612.5	3612.5	10.0627	4.32
Main effect K	1	160178	160178	446.1783	4.32
Main effect P	1	277512.5	277512.5	773.0153	4.32
Interaction effect NK	1	392	392	1.0919	4.32
Interaction effect NP	1	882	882	2.4568	4.32
Interaction effect KP	1	14280.5	14280.5	39.7786	4.32
Interaction effect NKP	1	98	98	0.2730	4.32
Error	21	7539	359		
Total	31	465337.5			

Inference:

1. The calculated value of F for N is greater than the table value of F for N i.e., the main effect of N is highly significant.

2. The calculated value of F for K is greater than the table value of F for K i.e., the main effect of K is highly significant.

3. The calculated value of F for P is greater than the table value of F for P i.e., the main effect of K is highly significant.

4. The calculated value of F for NK is less than the table value of F for NK i.e., the main effect of NK is highly significant.

5. The calculated value of F for NP is less than the table value of F for NP i.e., the main effect of NK is highly significant.

6. The calculated value of F for KP is less than the table value of F for KP i.e., the main effect of NK is highly significant.

7. The calculated value of F for NPK is less than the table value of F for NPK i.e., the main effect of NK is highly significant.

8. For the replication, the calculated value of F is less than the table value of F i.e., replication effect is not significant.

Practical: - 7

3^2 Factorial Design (partial confounding)

A 3^2 factorial design experiment was conducted blocks of 3 plots in 4 replicates then the following data is obtained.

(00) 64	(10) 69	(20) 81
(11) 67	(11) 70	(01) 82
(22) 69	(02) 75	(12) 76

(02) 69	(12) 72	(22) 64
(11) 81	(21) 67	(01) 83
(20) 72	(00) 69	(10) 61

(12) 74	(22) 61	(02) 69
(21) 65	(01) 82	(20) 76
(00) 70	(10) 61	(11) 82

(01) 85	(21) 72	(22) 70
(12) 75	(02) 75	(11) 70
(20) 80	(10) 73	(00) 65

Identify the confounded interactions and analyse the data.

Aim:-

To identify the confounded interactions and analyse the data.

Procedure:-

By using the appropriate Galois field equation. We can identify the confounded interaction. First we divide NP I_{NP}, J_{NP} components and find in which I_{NP} is confounded and J_{NP} is confounded.

If the principle block in a replicate statistics Galois field equation. $x_1 + 2x_2=0.(mod3)$ –(1) for x_1 = 0, 1, 2 and x_2 = 0, 1, 2 then we say that the component I_{NP} is component.

2. If the Galois field equation $x_1 + x_2=0.(mod3)$ for x_1 = 0, 1, 2 and x_2 = 0, 1, 2 then the component is J_{NP} confounded. Now verify the replicate in which the I_{NP} or J_{NP} is confounded and the replicate in which I_{NP} (or) J_{NP} is not confounded.

Note:-

If different interactions are confounded in which replicate then there are called partially confounded.

Null hypothesis:-

H_{01}: There is no significant effect in blocks.

H_{02}: Main effect N is not significant.

H_{03}: Main effect N is not significant.

H_{04} : Interaction effect INP is not significant

H_{05} : Interaction effect INP is not significant

Now, to test the above null hypothesis, we he to formulated the two- way table of treatment totals scmed over r-replications

		P			
		0	1	2	Total
	0	(0,0)	(01)	(02)	N0
N	1	(10)	(11)	(12)	N1
	2	(20)	(21)	(22)	N2
	Total	P_0	P_1	P_2	

C.F = $\frac{G^2}{3^2 r}$

Where r – No . of replications

Sum of squares due to M.E.N

$$N^* = \frac{N_o^2 + N_1^2 + N_2^2}{3^{2-1} r} - CF$$

Sum of square due to M.E.P

$$P^* = \frac{P_0^2 + P_1^2 + P_2^2}{3^{2-1} r} - CF$$

Total sum of squares TSS = $\sum_i \sum_j y_{ij}^2 - CF$

Block sum of squares B = $\sum_{i=1}^{12} \frac{B_i^2}{k} CF$

Where k = size of the block = 3 = 3 (or)

No. of plots in the block.

$$I_{NP}^* = \frac{I_0^2 + I_1^2 + I_2^2}{3^{2-1} r} - CF$$

r – no .of replicates I_{NP} not confounded

$$J^*_{NP} = \frac{J_0^2 + J_1^2 + J_2^2}{3^{2-1} r} - CF$$

r – no of replicates. J_{NP} is not confounded

$CF_2 = \frac{G_1^2}{3^2 r}$

G_1 – Grand total in which INP is not confounded similarly $CF_2 = \frac{G_2^2}{3^2 r}$

G_2 – Grad total in which JNP is not confounded

Source of variation	Degrees of freedom	Som of squares	Mean sum of squares	F_{cal}	F_{tab}
Blocks main effect	(3r-1)	B^*	$B = B^*/3r-1$	B/E	$F_{(3r-1)}$
N	2	N^*	$N = N^*/2$	N/E	$F_{(2,*)}$
P	2	P^*	$P = P^*/2$	P/E	$F_{(2,*)}$
Interaction effect					
I' NP	2	INP^*	$INP = I^*NP/2$	INP/E	$F_{(2,*)}$
J' NP	2	JNP^*	$JNP = J^*NP/2$	JNP/E	$F_{(2,*)}$
Error	$*$	E^*	$E = E^{*/*}$		
Total	(3^2r-1)	TSS			

$* = (3^2 r-1) - (3r-1) - 2 – 2 – 2 – 2$

E^{-*} TSS - B^* - N^* - P^* - INP^* - JNP^*

CONCLUSION:-

If F calculated value is less then F – Table value at 5 % level of significance. Then we accept H_0 otherwise reject H_0.

Calculations:-

The principle block in Replication I

1. $X_1 + 2 x_2 \equiv 0 \pmod 3$ 2. $X_1 + x_2 \equiv 0 \pmod 3$

$0 + 2 (0) \equiv 0$ (mod3) $0 + 0 \equiv$ (mod3) = 0

$1 + 2 (1) \equiv 0$ (mod3) $1 + 1 \equiv 0$ (mod3) $\neq 0$

$2 + 2 (2) \equiv 0$ (mod3)

$\therefore$ the principle block is (0 0), (1 1), (2 2).

Here INP confounded and

JNP not confounded.

$x_1 + 2 x_2 \equiv 0$ (mod3) $\rightarrow B_1$

$x_1 + 2 x_2 \equiv 0$ (mod3) $\rightarrow B_2$

$x_1 + 2 x_2 \equiv 0$ (mod3) $\rightarrow B_3$.

The principle block in Replication II.

1. $x_1 + 2 x_2 \equiv 0$ (mod3) 2. $x_1 + x_2 \equiv 0$ (mod3)

$1 + 2 (2) \equiv 0$ (mod3) $\neq 0$ $1 + 2 \equiv 0$ (mod3) = 0

$2 + 1 \equiv 0$ (mod3) = 0

$0 + 0 \equiv 0$ (mod3) = 0

Here the principle block is (1 2), (2 1), (0 , 0)

here INP not confounded.

JNP confounded

$x_1 + x_2 \equiv 0$ (mod3) $\rightarrow B_1$

$x_1 + x_2 \equiv 1 \pmod 3 \rightarrow B_3$

$x_1 + 2x_2 \equiv 2 \pmod 3 \rightarrow B_3$.

The principle block in Replication III.

1. $x_1 + 2x_2 \equiv 0 \pmod 3$ 2. $x_1 + x_2 \equiv 0 \pmod 3$

$1 + 4 \equiv 0 \pmod 3 \neq 0$ $1 + 2 \equiv 0 \pmod 3$

$2 + 1 \equiv 0 \pmod 3$

$0 + 0 \equiv 0 \pmod 3$

Here the principle block is (1 2), (2 1), (0 , 0)

here I_{NP} not confounded.

J_{NP} confounded.

$x_1 + x_2 \equiv 0 \pmod 3 \rightarrow B_1$

$x_1 + x_2 \equiv 1 \pmod 3 \rightarrow B_2$

$x_1 + 2x_2 \equiv 2 \pmod 3 \rightarrow B_3$

The principle block in Replication III.

1. $x_1 + 2x_2 \equiv 0 \pmod 3$ 2. $x_1 + x_2 \equiv 0 \pmod 3$

$2 + 4 \equiv 0 \pmod 3$ $2 + 2 \equiv 0 \pmod 3 \neq 0$

$1 + 2 \equiv 0 \pmod 3$

$0 + 0 \equiv 0 \pmod 3$

Here the principle block is B_3 i.e., (2 2) (1 1) , (0 0)

Here $I_{NP}$is confounded.

J_{NP} is not confounded.

$x_1 + x_2 \equiv 0 \pmod 3$ -------- B_3

$x_1 + x_2 \equiv 1 \pmod 3$ -------- B_2

$x_1 + 2x_2 \equiv 2 \pmod 3$ ------ B_1

The two – way table of treatment totals summed over replication

	0	1	2	Totals
0	268	332	288	888
1	264	300	297	861
2	269	274	264	897
Totals	841	906	849	2596

Corrections factor CF = $\frac{G^2}{3^2 r}$

$= \frac{(2596)^2}{9.4}$ = 187200.4444

Total sum of squares TSS = $\sum\sum y_{ij}^2$ - CF

= 188780 – 187200.4444

= 1579.5556

Block sum of squares B = $\sum_{i=1}^{12} \frac{B_i^2}{k}$ - CF

$$= \frac{(200)^2 + (214)^2 + (239)^2 + (222)^2 + (208)^2 + (208)^2 + (209)^2 + (2)^2 + (240)^2 + (220)^2 + (205)^2}{3} - CF$$

= 187860 – 187200.4444

= 659.5556

Sum of squares due to main effect N

$$= \frac{N_0^2 + N_1^2 + N_2^2}{3r} - CF$$

$$= \frac{(888)2 + (861)2 + (847)2}{3(3+1)} - CF$$

= 187272-8333 – 187200.4444 = 72.3889

Sum of squares due to main effect p

$$= \frac{P_0^2 + P_1^2 + P_2^2}{3r} - CF$$

$$= \frac{(841)^2 + (906)^2 + (849)^2}{3.4} - CF$$

= 187409.8333 – 187200.4444

= 209.3889.

Sum of square due to I_{NP}

$$= \frac{I_0^2 + I_1^2 + I_2^3}{3r} - CF$$

$$CF_1 = 90738$$

$$CF_2 = 96506.8889$$

I_0 = 427 I_1 = 392 $I 2$ = 459

$$INP = \frac{(427)^2 + (392)^2 + (459)^2}{6} - 90738$$

= 374.3333

Sum of squares due to I_{NP}:

$$= \frac{J_0^2 + J_1^2 + J_2^3}{3r} - CF_2$$

J_0 = 422 J_1 = 448 J_2 = 448

$$j_{NP} = \frac{(422)^2 + (448)^2 + (448)^2}{6} - 96506.8889$$

= 75.1111

ANOVA TABLE:-

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F_{cal}	F_{Tab}
Blocks main effect	(11)	659.5556	59.9596	5.0819	2.45
N	2	72.3889	36.19445	3.0677	3.63
P	2	209.3998	209.3889	8.8735	3.63
Interaction effect					
I_{NP}	2	374.3333	374.3333	15.8635	3.63
J_{NP}	2	75.1111	75.1111	3.1830	3.63
Error	(16)	188.7778	188.7778		
Total	(35)	1579.5556	1579.5556		

Inference:-

1. If F calculated value > the F – Tabulated value for Replicates. Hence we reject the null hypothesis.
2. If F calculated value < the F – Tabulated value tosmain effect N. hence we accept the null hypothesis i.e., the main effect due to N is not significant.
3. The F calculated value > the F – Tabulated value for main effect p. hence we reject the null hypothesis
4. The F – calculated value > the F – Tabulated value for Interaction effect I_{NP} Hence we reject the null hypothesis.
5. The F – Tabulated value < the F – Tabulated value for Interaction effect J_{NP} Hence we accept the null hypothesis i.e., Interaction effect J_{NP} is not significant.

Practical:- 8

Balanced Incomplete Block Design

The following table was obtained in an experiment conclusion in a BIBD with 9 treatments in 18 Blocks, 4 plots each with r and d = 3 (v = 9, b= 18, r= 8, k= 4, d = 3) the treatments are denote by a, b,c,d,e,f,g,h,i.

Blocks Data on yields of Plots

B_1	f 2.6	d.9.7	c 5.4	e 6.9
B_2	f 5.9	g 2.6	i 5.9	b.6.3
B_3	a .70	f 4.6	i5.9	c 3.3
B_4	i 2.4	d 4.0	g 3.0	f 2.4
B_5	i 5.0	b 7.4	e 10.3	c 9.4
B_6	d 10.1	a 9.7	f. 5.7	b.7.5
B_7	b 8.9	d 4.1	e 6.4	i 6.3
B_8	b 4.0	f 6.1	g4.4	c 3.3
B_9	b 2.8	f 2.6	e2.9	b 3.3
B_{10}	b.5.7	h.9.3	c 5.4	i 6.1
B_{11}	b.4.7	g 6.6	a 5.5	b 5.3
B_{12}	a 3.0	h 1.4	i 4.2	d 2.8
B_{13}	c.75	g 2.2	e 2.6	a 4.4
B_{14}	c 3.7	a 5.2	d 2.4	b 2.4
B_{15}	i 3.0	g 2.6	a 4.7	e 2.4
B_{16}	d 4.5	b 6.0	g 4.6	c 3.3
B_{17}	g.2.6	e 4.9	d 6.0	b 4.6
B_{18}	b7.3	e 5.4	f 5.7	a4.4

Analyse the data and draw conclusions.

Aim:-

To analyse the data and draw conclusions.

Procedure:-

First we find from the given data the two – way table between treatment and blocks taking blocks as records a treatments as columns. Form this table we find the unadjusted sum of squares due to blocks and treatment sum of squares due to blocks (unadjusted) is obtained by

B = $\sum_{i=1}^{18} \frac{B_i^2}{k} - cf$; where B_i is the i^{th} Block total in two a table; k – block size in the given two – way table

Correction factor CF = $\frac{G^2}{bk}$ or $\frac{G^2}{vr}$

Sum of squares due to treatments (unadjusted) is obtained

$\frac{\sum_{j=1}^{9} v_j^2}{r} - CF$; where v_j is the j^{th} treatment total in two – way table. And r is no. of replications

Sum of squares due to treatment (adjusted) is obtained b.

$\frac{\sum Q_i^2}{rE}$; where Q_i is detained by Q_i = $v_j \frac{-T_j}{k}$ and , T_j – Total of Blocks in which j^{th} treatment appear.

Sum of squares due to Blocks (adjusted) = Treatment sum of square (adjusted) t sum of square due to Block

(Unadjusted) – Treatment sum squares (unadjusted)

Total sum of squares = $\sum_i \sum_j y_{ij}^2 - CF$

Null hypo

H_{01}:- Treatment effects is not significant

H_{02}:- Block effects is not significant

Calculations:-

In this design

(u = 9, b=18, r=8, k=4, λ=3)

New to formulate two – way table between blocks & treatment

ANOUATABLE:-

S.V	d.f	ss	mss	F_{cal}	F_{cal}	mss	ss	d.f	s.v
Adjusted Block	(b-1)	B^1	B^{1*}	B1/E	-	-	B	(b-1)	Unadjusted Block
Unadjusted Treatment	(t-1)	T	-	-	T^{1*}/E	T1	T1	(t-1)	Adjusted treatment
Error	*	E*	E=E*/*			E=E*/*	E*	*	Error
Total	(n..-1)	TSS					TSS	(n..-1)	Total

= (n-1) – (b -1) – (t-1)

Conclusion:-

If f_{cal} value less than F- tabulated value at ∝% less of significance. Then we accept the null hypothesis others we reject the null hypothesis.

	a	b	c	d	e	f	g	h	i	Total
B_1			5.4	9.7	6.9	2.6				24.6
B_2		6.3				5.9	2.6		5.9	20.7
B_3	7.0		3.3			4.6			5.9	20.8
B_4				4.0		2.4	3.0		2.4	11.8
B_5			9.4		10.3			7.4	5.0	32.1
B_6	9.7			10.1		5.7		7.5		33
B_7		8.9		4.1	6.4				6.3	25.7

B_8			3.3			6.1	4.4	4.0		17.8
B_9		2.8			2.8	2.6		3.3		11.8
B_{10}		5.7	5.4					9.3	6.1	26.8
B_{11}	5.5	4.7					6.6	5.3		22.1
B_{12}	3.0			2.8				1.4	4.2	11.4
B_{13}	4.4		7.5		2.6		2.2			16.7
B_{14}	5.2	2.4	3.7	2.4						13.7
B_{15}	4.7				2.4		2.6		3.0	12.7
B_{16}		60	3.3	4.5			4.6			18
B_{17}				6.0	4.9		2.6	4.6		18.1
B_{18}	4.4	7.3			5.4	5.7				22.8
Total	43.9	44.1	41.3	43.6	41.7	35.6	28.6	42.8	38.8	360.4

Correction factor CF = $\frac{G^2}{VR} = \frac{(360.4)^2}{9.8}$ = 1804.0022

Sum of squares due to Blocks (un adjusted) is.

$$= \sum_{i=1}^{18} \frac{B_1^2}{K} - CF$$

$$= \frac{(24.6)^2 + (20.7)^2 + (20.8)^2 + (11.8)^2 + (32.1)^2 + (33)^2 + (25.7)^2 + (11.5)^2 + (26.5)^2 + (22-1)^2 + (11.4)^2 + (16.7)^2 + (13.7)^2 + (12.7)^2 + (18.1)^2 + (18.1)^2 + (22.8)^2}{4} - 1804.002$$

$$= \frac{7969.02}{4} - 1804.0022 = 188.2528$$

Sum of squares due to treatments (unadjusted) is

$$= \sum_{j=1}^{9} \frac{V_j^2}{r} - CF$$

$$= \frac{(43.9)^2 + (44.1)^2 + (41.3)^2 + (43.6)^2 + (41.7)^2 + (35.6)^2 + (28.6)^2 + (42.8)^2 + (38.8)^2}{8} - CF$$

$$= \frac{14640.16}{8} - 1840.0022 \quad = 1830.02 - 1840.0022 \quad = 26.0178$$

Treatments	T_j	V_j	T_j/k	Q_i	Q_i^2
a	153.2	43.9	38.3	5.6	31.36
b	161.4	44.1	40.35	3.75	14.0625
c	170.6	41.3	42.65	-1.35	1.8225
d	156.7	43.6	39.175	4.425	19.5806
e	164.2	41.7	41.05	-5.15	26.5225
f	163	35.6	40.75	0.65	0.4225
g	138.3	28.6	34.575	-5.975	35.7006
h	172.5	42.8	43.125	-0.325	0.105625
i	161.7	38.8	40.425	-1.625	2.64062
					132.21706

Treatment sum of squares (adjusted)

$$= \frac{Q_i^2}{rE} = \frac{(132.21706)}{8\left(\frac{27}{32}\right)} = 19.7434$$

Block sum of squares (adjusted) = 19.7434 +188.2528 – 26.0178
= 181.9244

Total sum of squares = $\sum_i \sum_j Y_{ij}^2 - CF$

= 2137.96 – 1804.0022 = 333.9578

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares	Fcal	Fcal	Mean sum of squares	Sum of squares	Degrees of Freedom	Source of variation	F tab
Adjusted Block	(17)	10.7014	10.7014	3.9063	-		188.2528	(18-1)=17	Unadjusted block	1.89
Unadjusted treatment	(8)	26.0178	-	-	0.9012	2.4679	19.7434	8	Adjusted treatment	
Error	(46)	126.0156	2.7395			2.7383	125.9616	46	Error	2.16
total	(71)	333.9578					333.9578	(71)	Total	

Inference

If F – calculated value for adjusted Block > the F tabulated value, hence. We reject the null hypothesis

If F- calculated value for Adjusted Treatment < the f tabulated value hence we accept the null hypothesis.

Practical:- 9
Graeco Latin Square Design

In Graeco Latin Square design the data is given below

	1	2	3	4	5
1	Aα -1	Bγ -5	Cθ -6	Dβ -1	Eδ -1
2	Bβ -8	Cδ -6	Dα 5	Eγ 2	Aθ 11
3	Cγ -7	Dθ 13	Eβ 1	Aδ 2	Bα -4
4	Dδ 1	Eα 6	Aγ 1	Bθ -2	Cβ -3
5	Eθ -3	Aβ 5	Bδ -5	Cα 4	Dγ 6

Analyze the data and draw conclusions

Aim

To analyse the data and draw conclusions. For Graew Latin square design

Procedure

The mathematical for the Graew Latin square design is

$Y_{ij}(h) = \mu + r_i + c_j + g_k + t_1 + E_{ij}(k)$

There is no significant difference between Rows
There is no significant difference between columns
There is no significant difference between Greek Letter
There is no significant difference between Treatment

Sum of square due to Row R = $\frac{\sum_{i=1}^{t} R_i^2}{t} - CF$

Sum of square due to column C = $\frac{\sum_{j=1}^{t} C_j^2}{t} - CF$

Sum of square due to Greek letter G = $\frac{\sum_{k=1}^{t} G_k^2}{t} - CF$

Sum of Square due to Treatment T = $\sum_{l=1}^{t} \frac{t_l^2}{t} - CF$

CF = $\frac{G^2}{T^2}$; G – Grand Total

Total sum of squares T = $\sum_{i}^{t}\sum_{j}^{t}\sum_{k}^{t}\sum_{l}^{t} y_{ijkl}^2 - CF$

Correction Factors CF = $\frac{G^2}{T^2}$ = $\frac{(10)^2}{(5)^2}$ = 4

Total sum of squares TSS = $\sum_{i}\sum_{j}\sum_{k}\sum_{l} {y_{ijkl}}^2 - CF$

= 680 – 4 = 676

Row sum of squares R = $\sum_{i=1}^{5} \frac{R_i^2}{t} - CF$

= $\frac{(-14)^2 + 9^2 + 5^2 + 3^2 + 7^2}{5} - 4$

= $\frac{360}{5} - 4$ = 68

Column Sum of squares C = $\sum_{j=1}^{5} \frac{C_j^2}{t} - CF$

$$= \frac{(-18)^2 + (18)^2 + 4^2 + 5^2 + 9^2}{5} - CF$$

$$= \frac{770}{5} - 4 = 154 - 4 = 150$$

ANOVA TABLE:-

Source of variation	Degree of freedom	Sum of squares	Mean sum of square	F_{cal}	F_{tab}
Rows	(t -1)	R	$R^1 = R(t-1)$	R^1/E^1	F(t -1)
Columns	(t -1)	C	$C^1 = C/(t-1)$	C^1/E^1	F(t - 2)
Greek Letters	(t - 1)	G	$G^1 = G/(t-1)$	G^1/E^1	F(t - 3)
Treatment(Latin letter)	(t -1)	T	$T^1 = T/(t-1)$	T^1/E^1	F(t - 4)
Error	*	**	$E1 = \frac{**}{*}$		
Total	(t^2-1)	TSS			

Where * = $(t^2 -1) - 4 (t -1)$

* * = TSS – R – C – G – T

Conclusion:-

If F calculated Value less than the F table value then w accept the null hypothesis otherwise reject the null hypotheses

Calculations:-

	1	2	3	4	5	Totals
1	-1	-5	-6	-1	-1	-14
2	-8	-1	5	2	11	9
3	-7	13	1	2	-4	5
4	1	6	1	-2	-3	3
5	-3	5	-5	4	6	7
	-18	18	-4	5	9	10

	α	β	$\Im$	δ	θ	Total
A	-1	5	1	2	11	18
B	-4	-8	-5	-5	-2	-24
C	4	-3	-7	-1	-6	-13
D	5	-1	6	1	B	24
E	6	1	2	-1	-3	5
Total	10	-6	-3	-4	B	10

Sum of square due to freek letters

$$G = \sum_{k=1}^{5} \frac{g_k^2}{t} - CF$$

$$= \frac{(10)^2 + (16)^2 + (-3)^2 + (-4)^2 + (13)^2}{5} - 4$$

$$= \frac{330}{5} - 4$$

= 66 – 4 = 62

Sum of square due to treatment

$$F = \sum_{l=1}^{5} \frac{t_j^2}{5} - CF$$

$$= \frac{(18)^2 + -(24)^2 + -(13)^2 + (24)^2 + 5^2}{5} - CF$$

= 334 – 4

= 330

Ource of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F_{cal}	F_{Tab}
Rows	4	68	17	2.0606	13.84
Columns	4	150	37.5	4.5454	3.84
Greek letters	4	62	15.5	1.8787	3.84
Treatments	4	330	82.5	10	3.84
Error	8	66	8.25		
total	24	676			

Inference:-

1. The F calculated value for Row is < the F Table value then we accept the null hypothesis i.e., there is no significant difference between rows.
2. The F calculated value for column is > the F – Tab leveled then we reject the null hypothesis i. e., there is a significant difference between columns.
3. The F calculated value for greek letter is < the F – Table then we accept the null hypothesis i.e., there is no significant difference between greek letters.
4. The F – calculated value for Treatments is > the F – Table value then we reject the null hypothesis i.e., there is a significant difference between Treatments.

Practical No: 10

Split – Plot Design

With a view of study the relative utility of Nitrogen and phosphorous combinations with different rates at increasing of surfer can crop an experiment was conducted. A split plot design in 2 replicates, consisting of 3 whole plots with irrigation treatments and 4 sufple with NP combinational treatments was adopted for stuo further details are given below

Sub plot 1/40 per an acre

Whole plot 1/10 per an acre

N_0 – Nitrogen at o^{th} level

N_1 – Nitrogen at 10 lbs
P_0 – $p_2\, 0_5$ at 0^{th} level
P_1 – $p_2\, 0_5$ at 72 lbs
I – Irrigation type
S – Subplot

Replication I

	I_1	I_2	I_3
$n_o p_o$	16	20	22
$n_o p_1$	11	20	26
$n_1 p_o$	13	19	22
$n_1 p_1$	19	19	21

Replication II

	I_1	I_2	I_3
$n_o p_o$	10	19	24
$n_o p_1$	11	18	25
$n_1 p_o$	10	18	21
$n_1 p_1$	12	20	22

Analyse the data and draw conclusions.

Aim :-
To analyse the data and draw conclusion for the given split – plot Design
Procedure:-
We have to test the following hypothesis 1
H_{01}: the effect due to replicates is not significe
H_{02}: the effect due to irrigation levels is not significe
H_{03}: the effect due to NP levels is not significant.
H_{04}: the effects due to (sx I) interaction is not significant.
(r x t) table.
Where, whole plot treatments

		I_0	I_1	I_2	Total
Replicates	R_1				$\sum R_1$
	R_2				$\sum R_2$
	Total	$\sum I_0$	$\sum I_1$	$\sum I_2$	G

(r x t) table.

	I_0	I_1	I_2	Total
$n_o p_o$				$\sum n_o p_o = s_1$
$n_o p_1$				$\sum n_o p_1 = s_2$
$n_1 p_o$				$\sum n_1 p_o = s_3$
$n_1 p_1$				$\sum n_1 p_1 = s_4$
Total	$\sum I_0$	$\sum I_1$	$\sum I_2$	G

Where r is the no. of replicates and s is no. of split plots and t is no. of whole plot treatments.

The corrections factor CF = $\frac{G^2}{rts}$

Sum of squares due to replicates = R = $\frac{\sum R_i^2}{st} - CF$

$$= \frac{R_1^2 + R_2^2}{st} - CF$$

Sum of squares due to split plots (or)
Sum of squares due to Irrigation level

$$I = \frac{I_0^2 + I_1^2 + I_2^2}{rs} - CF$$

Sum of squares due to subplots (or)
Sub of squares due to N_p levels is

$$S = \frac{S_1^2 + S_2^2 + S_3^2 + S_4^2}{r\,t} - CF$$

Total sum of squares due to (s x t) table is

$$= \frac{\sum (st)^2}{r} - CF$$

= Individual sum of squares due to (s x t) table divided by r – CF

Indirection sum of squares due (s x I) is

S_i = Total sum of squares due to (s × t) – sum squares due to Np levels – sum of squares due split plots.

Total sum of squares due to (r x t) table is

$$P = \frac{\sum (rt)^2}{3} - CF$$

= Individual sum of squares due to

$$\frac{(rxt)table}{\&} - CF$$

Total sum of squares due to (s x I) is

Q = Individual sum of squares due to replication I and Replicate II – CF

ANNOVA Table:-

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F_{cal}	F_{Tab}
Replicates Irrigation Whole plot error	(r-1) (t-1) ∗ : (rt-1) (r-1) – (t-1)	R I E=P-R-I	$R^1 = R/(r-1)$ $I^1 = I/(t-1)$ $E^1 = E/*$	R^1 / E^1 I^1 / E^1	FC FC
TSS of (rxt)	(r t -1) = x (say)	P			
Split plot treatment (SXI) Interaction Split plot error	(s-1) = y (say) (s-1)(t-1)= z (say) ∗ = (rst-1)-x-y- (r-1)-(t-1)	S SI E_1=Q-R-S-SI-I-E	$S^1 = s/y$ $(SI)^1$-SI/Z $E_1^1 = E_1 / x^1$	$S^1 = E_1^1$ $(SI)^1 / E_1^1$	FC FC
Tssot (SXI)	(rst-1)	Q			

$*^1$: (rst-1) – z – y - ∗ - (t-1)- (r-1)

Conclusion:-

If F – calculated value less than the F – Tabulated value then we accept the null hypothesis other reject H_0

Calculations:-

Construct (rxt) table:

Whole plot treatments

		I_{01}	I_2	I_3	Total
Replicates	R_1	59	78	91	228
	R_2	43	75	92	210
	Total	102	153	183	438

Construct (sxt) table.

	I_0	I_1	I_2	Total
n_0p_0	26	39	46	111
n_op_1	22	38	51	111
n_1p_o	23	37	43	103
$n_1 p_1$	31	39	43	113
Total	102	153	183	438

Correction factor CF = $\frac{G^2}{rts}$

= $\frac{(438)^2}{2(3)(4)} = 7993.5$

Sum of squares due to replicates

R = $\frac{\sum R_i^2}{st} - CF$

= $\frac{(228^2)+(210)^2}{(4)(3)}$ - 7993.5 = 8007 – 7993 = 13.5

Sum of squares due to Irrigation levels

I = $\frac{I_1^2 + I_2^2 + I_3^2}{rs} - CF$

= $\frac{(102)^2 + (153)^2 + (183)^2}{2(4)}$ - 7993.5

= 8412.75 – 7993.5

= 419.25

Sum of squares due to Np levels is

S = $\frac{S_1^2 + S_2^2 + S_3^2 + S_4^2}{rt} - CF$

= $\frac{(111)^2 + (111)^2 + (103)^2 + (113)^2}{2(3)}$ - 7993.5

= 8003.3333-7993.5 = 9.8333.

Total sum of squares due to (s x t) table is

= $\frac{(26)^2 + (39)^2 + (46)^2 + \ldots\ldots\ldots\ldots + (43)^2}{r} - 7993.5$

= 8460 – 7993.5 = 466.5

Interaction sum of squares due to (SXI) is

S_i = 466.5 – 9.8333 – 419.25

= 37.4167

Total sum of squares due to (s x t) table is

= $\frac{(59)^2 + (78)^2 + (91)^2 + .(43)^2 + (75)^2 + (92)^2}{4} - 7993.5$

= $\frac{33784}{4} - 7993.5$

= 8446 – 7993.5

= 452.5

Total sum of squares due to (SXI) table

$Q = (16)^2 + (20)^2 + (22)^2 + (22)^2 - CF$

$= 8514 – 7993.5 \quad = 520.5$

ANNOVA Table:-

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F_{cal}	F_{Tab}
Replicates	1	13.5	13.5	1.3671	18.51
Irrigation	2	419.25	209.625	21.1178	19.00
Whole plot error	2	19.75	9.875		
TSS of (rxt)	5	452.5			
Split plot treatment	3	9.8333	3.2777	1.4216	3.86
(SXI) Interaction	6	37.4167	6.2361	2.7047	3.37
Split plot error	9	20.75	2.3056		
Total (SXI)	23	520.5			

Inference:-

1. F calculated value < the F – Tabulated value for replicates then we accept the null hypothesis i.e., the effect due to replicates is not significant
2. F- calculated value > the F – Tabulated value for Irrigation then we reject the null hypothesis.
3. F - calculated value < the F – Tabulated value for split plot then we accept the null hypothesis i. e., the effect due to NP levels is not significant.
4. F – calculated value < the F – Tabulated value for Interaction. Then we accept the null hypothesis, i.e., the effect due to Interaction is not significant.

Practical No.: 11
Lattice Design

Experimental layout and observations of a simple square Lattice design is given below:

Block	Block Contents				Total
	A	B	C	D	
1	5	4	4	3	16
	E	F	G	H	
2	5	6	3	4	18
	I	J	K	L	
3	3	5	6	6	20
	M	N	O	P	
4	4	4	3	4	15
	A	E	I	M	
5	6	3	5	4	18
	B	F	J	N	
6	4	6	5	6	21
	C	G	K	O	
7	3	2	4	5	14
	D	H	L	P	
8	5	4	2	3	14

Analyse the data and draw conclusions.

Aim:- To analyse the data and to draw conclusions for the gi Lattice Design.

Procedure:- Mathematical model

$$Y_{ij(m)} = \mu + \beta_i + \tau_m + \varepsilon_{ijc(m)}$$

Null hypothesis:-

H_{01}: The effect due to treatments is not significant.

Let 'G' be grand total

B_i – is i^{th} block total

T_m – m^{th} treatment total

Q_m – adjusted treatment of the m^{th} treatment obtained by subtracting the sum of the block mean in which m^{th} treatment occurs to T_m.

$S(Q_m)$ = sum of adjusted treatment total of (2k – treatment which occurs in the same row and same column as the m^{th} treatment.

Estimated treatment effect $\hat{\tau}_m$ of τ_m is given by

$$\hat{\tau}_m = \frac{1}{2k}[(k+2)Q_m + S(Q_m)]$$

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F_{cal}	F_{tab}
Blocks (ignoring treatments)	(2k – 1)	$B = \frac{\sum_{i=1}^{2k} B_i^2}{K} - \frac{G^2}{2K^2}$	-	-	
Treatments (eliminating blocks)	$(K^2 – 1)$	$T = \sum_{m=1}^{k} \hat{\tau}_m Q_m$	$M_{st} = T/(K^2 – 1)$	$\frac{M_{st}}{M_{S_E}}$	$F(K^2 – 1, *)$
Error	*	**	$M_{S_E} = \frac{**}{*}$		
Total	$(2K^2 – 1)$	$\sum_{i=1}^{2k}\sum_{j=1}^{k^2} y_{ij}^2 - \frac{G^2}{2K^2}$			

where * = $(2k^2 – 1) – (2k – 1) – (k^2 – 1)$

** = TSS – B - T

Conclusion:-

If F – calculated value less than the F – tabulated value, then we accept the null hypothesis otherwise reject the null hypothesis.

Calculations:-

Block					Total
1	5	4	4	3	16
2	5	6	3	4	18
3	3	5	6	6	20
4	4	4	3	4	15
5	6	3	5	4	18
6	4	6	5	6	21
7	3	2	4	5	14
8	5	4	2	3	14

Correction factor CF $= \frac{G^2}{2K^2}$ $= \frac{(136)^2}{2(4)^2}$

$= \frac{(136)^2}{32}$ $= 578$

Total sum of squares TSS $= \sum\sum y_{ij}^2 - CF$

$= 5^2 + 4^2 + \ldots..+2^2 + 3^2 - CF$

$= 622 - 578$ $= 44$

Block sum of squares B $= \frac{\sum B_i^2}{k} - CF$

$= \frac{(16)^2 + (18)^2 + \ldots..(14)^2}{4} - CF$

$= \frac{2362}{4} - 578$ $= 590.5 - 578$ $= 12.5$

Treatments	Treatment Total	Sum of block mean in which the treatment occur	Q_m	$S(Q_m)$	$\hat{\tau}_m = [6Q_m + S(Q_m)]$	$\hat{\tau}_m Q_m$
A	11	8.5	2.5	-4	1.375	3.4375
B	8	9.25	-1.25	5.5	-0.25	0.3125
C	7	7.5	-0.5	1	-0.25	0.125
D	8	7.5	0.5	0	0.375	0.1875
E	8	9	-1	0	-0.75	0.75
F	12	9.75	2.25	-4.5	1.125	2.5312
G	5	8	-3	3	-1.875	5.625
H	8	8	0	-2	-0.25	0
I	8	9.5	-1.5	2	-0.875	1.3125
J	10	10.25	-0.25	1.5	0	0
K	10	8.5	1.5	-5	0.5	0.75
L	8	8.5	-0.5	0	-0.375	0.1875
M	8	8.25	-0.25	1.5	0	0
N	10	9	1	1	0.875	0.875
O	8	7.25	0.75	-1.5	0.375	0.2812
P	7	7.25	-0.25	1.5	0	0
Total	136	136	0	0	0	16.3749

Treatment sum of squares $= \sum_{m=1}^{k} \tau_m Q_m$ $= 16.3749$

ANOVA Table:-

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F_{cal}	F_{tab}
Blocks (ignoring treatments)	7	12.5	-	-	
Treatments (eliminating blocks)	15	16.3749	1.09166	0.64958	3.02
Error	9	15.1251	1.68056		
Total	31	44			

Inference:-
If F – calculated value less than the F – Tabulated value for treatment. Hence we accept the null hypothesis i.e., there is no significant difference due to treatments.

Inference:-
The F – calculated value for testing the significance of treatments effects is less than one and such F – values are interpreted as non – significant. In such cases the model has not will. Accounted the possible sources of variation and extreme case has to be exercised for future experiments using that material.

Practical No.: 12
M.L.ESTIMATION IN ZERO TRUNCATED POISSON DISTRIBUTION

For the following truncated Poisson date (truncated at zero), estimate the parameter by the method of Maximum likelihood Method:

x	1	2	3	4	5	6	7	8	9
f	22	18	18	11	3	6	3	0	1

* * * * *

AIM:- To estimate the parameter by ML method for the given zero truncated poisson data.

FORMULA:- the density function of ZTPD is p(x) = $\frac{e^{-\lambda}\lambda^{x}}{(1-e^{-\lambda})x!}$ x = 1,2,…..

The likelihood function is

L = $\pi_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^{x_{i!}}}{(1-e^{-\lambda})x_i!}$.

Log L = -n λ + n $\bar{x}$ -log λ - n log (1-$e^{-\lambda}$) - ∑log x_i!

ML equation to be solved is

$$\frac{\partial \log L}{\partial \lambda} = 0$$

$$\Rightarrow -n\left(1+\frac{e^{-\lambda}}{1-e^{-\lambda}}\right)+\frac{n\bar{x}}{\lambda}=0$$

$\Rightarrow \lambda = \bar{x}\,(1 - e^{-\lambda})$

Using method of iteration we have,

$\lambda_{i+1} = \bar{x}\,(1-e^{-\lambda_i})$. i = 0, 1, 2, …………

Take $\lambda_0 = \bar{x}$.

When, 1$|\lambda_{i+1}-\lambda_i| < .01$, stop the iteration and take λ_{i+1} as the estimate ofλ. Otherwise continue iteration procedure.

CALCULATIONS:-

For the given data mean = $\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$

$= \frac{237}{82}$ = 2.89

λ_I	1- $e^{-\lambda_i}$	$\bar{x}\left(1-e^{-\lambda_i}\right)$
2.89	.9444	2.7293
2.7293	.9347	2.7014
2.7012	.9329	2.6961
2.6961	. 9325	2.6950
2.6950	.9325	2.6950

$\lambda_3 = \lambda_4$ = 2.6950.

INFERENCE:-

The maximum likelihood estimator of the parameter λ is 2.6950.

Practical No.: 13
MAXIMUM LIKELIHOOD ESTIMATION IN WEIBULL

A random sample of 25 observations are generated from the wiebull distribution with c = 2 and b = 4. Obtain ML estimation of c and b.

1.8487	0.3761	0.7500	3.0530	1.3545	1.8802	1.5700
1.7708	1.3592	3.0464	1.7961	1.5319	0.5903	0.6288
0.6461	1.6560	1.7172	1.9310	1.0509	1.6173	1.3162
0.7705	1.8889	1.8889	4.1505			

====== * * * ======

AIM:- to obtain the ml estimators in weibull distribution.

FORMULA:-

The ML equations are

$$\left[\frac{\sum_{i=1}^{n} x_i^c \log x_i}{\sum_{i=1}^{n} x_i^c} - \frac{1}{c}\right] - \frac{1}{n}\sum_{i=1}^{n} \log x_i - 0 \quad i = 1, 2, \ldots\ldots\ldots N$$

$$b = \frac{1}{n}\sum_{i=1}^{n} x_i^{c^*}$$

where c^* is the solution of the above equation. To get the c value we use iteration procedure

$c_{k+1} = c_{k+nk}$

where $h_k = -\dfrac{f(c_k)}{f^1(c_k)}$

$$f(c) = \frac{\sum_i^n x_i^c \log x_i}{\sum_1^n x_i^c} - \frac{1}{c} - \frac{1}{n}\sum_1^n \log x_i$$

$$f^1(c) = \frac{\left[\sum_i^n x_i^c (\log x_i)^2\right]\sum_i^n x_i^c - \left(\sum_i^n x_i^c \log x_i\right)^2}{\left[\sum x_i^c\right]^2} + \frac{1}{c^2}$$

Initial value of c, c_0 = 1.9

we stop the iterations when two successive c's values are equal up to 4 decimal places. The last iterated value is the ml estimator this c in b, we get the ml estimator of b.

CALUCULATIO:-

Initial value of c = 1.9.

x_i	x_i^c	$\log x_i$	$x_i^c \log x_i$	$x_i^c (\log x_i)^2$
1.8487	3.215	.6146	1.9759	1.2144
1.8802	3.318	.6315	2.0953	1.3232
1.7961	3.05	.5870	2.0574	1.2077
1.6960	2.068	.5046	1.3155	.6640
1.3162	1.685	.2475	.4623	.1269
.3761	.1560	T.9779	T.8175	.1480
1.5700	2.356	.4516	1.0639	.4804
1.5319	2.250	.4268	.9603	.4098
1.7172	2.792	.5405	1.5090	.8156
.7705	.6905	T.7393	T.841	.0414
.7500	.579	T.7124	T.70041	.0862
1.7708	2.962	.5716	1.6931	.9677
.5903	.3674	T.4728	T.4589	.1080
1.9310	3.491	.6582	2.2977	1.3123
1.8889	3.350	.6363	2.1316	1.3563
3.0530	8.335	1.116	9.3043	10.3864
1.3592	1.792	.3068	.5494	.1683
.6288	.4143	.4143	-.1920	.0891
1.0509	.4104	.3038	.1247	.379
1.889	3.350	.0497	.1665	.0082
1.3545	.739	06363	1132	.072
3.0466	8.319	1.1149	9.2651	10.3279
.6491	.4351	.4351	T.5031	.0832
1.6173	2.491	2.491	.4804	.5748

$$F(c) = \left[\frac{60.5599}{751} - \frac{1}{1.9}\right] - \frac{8.8712}{125} = -.081$$

$$f^1(c) = \frac{63.65.98 \times 75.741 - (60.65 \leq 99)^2}{(75.741)^2} + \frac{1}{(19)^2} = .48$$

$c_1 = c_0 + n_1 = 1.9 + \frac{.081}{.48} = 2.068$

$c_1 = 2.068$

x_i	x_i^c	$\log x_i$	$x_i^c \log x_i$	$x_i^c (\log x_i)^2$
1.8487	3.56	.6142	2.1866	1.343
1.8802	3.69	.6315	2.330	1.4715
1.7961	3.358	5.5859	1.9674	1.1569
1.6960	2.837	.5045	1.4325	.7228
13162	1.764	.2745	.4842	.1329
.3761	.1323	-.9780	-.1294	.1205
1.5700	2.542	.4511	1.4465	.5174
1.5319	2.416	.4262	1.6310	.4399
1.7705	3.058	.5405	.6528	.8933
.7500	.5834	-.2607	-.1587	.0456
.5903	3.260	.5716	1.8634	1.0651
1.931	.3359	-.527	-772	.0935
1.8839	3.729	.6319	2.3715	1.5090
3.053	10.06	1.1162	11.2296	12.5352
1.3592	1.886	.3067	.5785	.1774
.6288	.8882	-.4638	-.1777	.0824
1.0509	1.109	.0497	.0351	.0027
1.8889	3.727	.6359	2.3715	1.5090
1.3545	1.875	.3037	.5695	.1736
3.6466	16.002	1.1144	11.1462	12.4218
.6461	.4071	-.4368	-.1769	.0768
1.6173	2.7010	.4804	1.2975	.6233
4.1505	18.98	1.4235	27.0117	38.4797
	86.8467	8.3710	72.3275	77.310

$f(c_1) = \frac{72.3265}{86.8467} - \frac{1}{2068} - \frac{8.3710}{25}$

$f^1(c_1) = \frac{77.310 \times 86.8467 - (72.3265)^2}{(86.8467)^2} + \frac{1}{(2.068)^2}$

=.43043.

$h_i = -\frac{f(c_1)}{f^1(c_1)} = .0332$.

C_2 = 2.068 - .0332 – 2.025.

$b = \frac{1}{n}\sum x_i^c$.

$= \frac{1}{25}(3.4771.)$

= 2.7351

INFERENCE:- maximum likelihood estimators of weibull to the given data or
c = 2.025
b = 2.7331.

Practical No.: 14
MINIMUM CHI-SQUARE AND MODIFIED CHI-SQUARE TEST

TYPE	f	prod.
Long and Purple	296	$(2+\theta)/4$
Long and Red	27	$(1 - \theta)/4$
Round and purple	19	$(1 - \theta)/4$
Red and purple	85	$\theta/4$

a) Estimate θ using minimum Chi-Square
b) Estimate θ using modified minimum Chi-Square and test the goodness fo fit.

*** *** ***

AIM:- to estimate θ by the method of modified minimum χ^2 and to test for goodness of fit.

PROCEDURE: - The mL equation is $\chi^{1^2} = \sum (np_i - f_i)^2 / f_i$

The mL equation to be solved for estimating θ is $\frac{\partial \chi^{1^2}}{\partial \theta} = 0$

If $\hat{\theta}$ is the ml estimator of θ, then find the expected values of θ. Then calculate

$$\chi^2 = \frac{\sum o_i^2}{e_i} - n$$

If calculated χ^2- value is less than or equal to table χ^2-value accept null hypothis. Otherwise reject null hypothesis.

CALCULATIONS:-

$$\chi^{1^2} = \frac{\left[427\left(\frac{2+\theta}{4}\right) - 296\right]^2}{296} + \frac{\left[427\left(\frac{1-\theta}{4}\right) - 27\right]^2}{27} + \frac{\left[427\left(\frac{1-\theta}{4}\right) - 19\right]^2}{19} + \frac{\left[427\left(\frac{\theta}{4}\right) - 85\right]^2}{85}$$

$$\frac{O\chi^{1^2}}{O\theta} = \frac{(427)^2}{2\times 296}\left[\frac{2+\theta}{4} - .6932\right] + \frac{(427)^2}{27\times 2}\left(\frac{1-\theta}{4} - .0632\right)(-1) + \frac{(427)^2}{19\times 2}\left[\frac{1-\theta}{4} - .0445\right](-1) + \frac{(427)^2}{85\times 2}\left(\frac{\theta}{4} - .19\right.$$

=307.9882

$$\left(\frac{\theta}{4} - .1922\right) + 3376.463\left(\frac{\theta}{4} - .1868\right) + 4798.1316\left(\frac{\theta}{4} - .2055\right) + 1072.5235\left(\frac{\theta}{4} - .1991\right)$$

θ = .7910

Observed fre q. o_i	Theoretical probabilities	Expected frequency e_i : n.p_i	o_i^2/e_i
294 28 19 86	(2+θ)/4 (1+θ)/4 (1+θ)/4 θ/4	298 23 22 84	290. ≤ 37 34.0870 16.4091 88.0476
n :427			428.5974

$$\chi^2 = \sum \frac{o_i^2}{e_i} - N = 1.5974$$

Table χ^2 – value at 5 % level = 5.99.

Accept H_0 at 5 % level.

CONCLUSION:-

Estimated θ – value = .79 10.

The fit is good.

Practical No.: 15

DECISION PROBLEM – MINIMAX APPROACH

In a decision problem, -^- = {0, 0, 0} whether 0 = 0.1, 0= 0.2, 0 = 0.3, x is Binomial (3,0) 0 -^. The action space A = {a , a , a} . The loss function L (a / 0) is described in the following table:

	01	02	03
a	0	0	0
a	10	-40	-40
a	15	-35	-85

Let D = {d , d, d, d, d , d, d, d,} , where di's are as follows.

X	d(x)	d(x)	d(x)	d(x)	d(x)	d(x)	d(x)	d(x)
0	a	a	a	a	a	a	a	a
1	a	a	a	a	a	a	a	a
2	a	a	a	a	a	a	a	a
3	a	a	a	a	a	a	a	a

(i) Evaluate the risk functions R (d / 0) for i = 1,2,………..,8.

(ii) Find out the minimax decision rule in D.

= = = ∗ ∗ ∗ = = =

AIM:- To evaluate the risk functions R (θ_i, d_j) i = 1, 2, 3 j = 1, 2, ……….8. to determine the admissible in D,J it exists and to find the minimax decision rule and test it is admissible or not.

FORULA:- The risk function R (θ_i, d_j) = d_j (0) p_i + d_j (1) p_i +_ d_j (2) p_i + d_j (3) p_j j = 1, 2, ………8.

The decision rule d^x is that rule whose loss is less than any then rule.

CALCULATIONS:-

Risk functions for i = 1, 2, 3 ; J = 1, 2, ……..8 are

$E_{\theta_1} L(\theta_1, d_1)$ = 0 + 0 + 0 + = 0

$E_{\theta_1} L(\theta_1, d_2)$= 10 (.1) + 10 (.1) + 10 (.1) 10 (.1) = 4

$E_{\theta_1} L(\theta_1, d_3)$ = 15 (.1) + 15 (.1) +10 (.5) +15 (.1) = 6

$E_{\theta_1} L(\theta_1, d_4)$ = 0 + 0 + 10 (.1) + .15 (.1) = 2.5

$E_{\theta_1} L(\theta_1, d_5)$ = 0 + 10 (.1) + 10 (.1) – 15 (.1) = 3.5

$E_{\theta_1} L(\theta_1, d_6)$ = 0 + 10 (.1) + 15 (.1) + 15 (.1) = 4

$E_{\theta_1} L(\theta_1, d_7)$ = 0 + 0 + 0 + 10 (.1) = 1

$E_{\theta_1} L(\theta_1, d_8)$ = 10 (.1) + 10 (.1) + 10 (.1) + 15 (.1) = 4.5

$E_{\theta_1} L(\theta_2, d_1)$ = 0 + 0 + 0 + 0 = 0

$E_{\theta_1} L(\theta_2, d_2)$ = - 40 (.2) -40 (.2) – 40 (.2) – 40 (.2) = - 3.2

$E_{\theta_1} L(\theta_2, d_3)$ = - 35 (.2) – 35 (.2) – 35 (.2) – 35 (.2) = - 28

$E_{\theta_1} L(\theta_2, d_4)$ = 0 + 0 – 40 (.2) – 35 (.2) = - 15

$E_{\theta_1} L(\theta_2, d_5)$ = 0 + 40 (.2) – 40 (.2) – 35 (.2) = - 23

$E_{\theta_1} L(\theta_2, d_6)$ = 0 – 40 (.2) – 35 (.2) - 35 (.2) = -22

$E_{\theta_1} L(\theta_2, d_7)$ = 0 + 0 + 0 – 40 (.2) = -8

$E_{\theta_1} L(\theta_2, d_8)$ = - 40 (.2) – 40 (.2) – 40 (.2) – 35 (.2) = - 31

$E_{\theta_1} L(\theta_3, d_1)$= 0 + 0 + 0 + 0 = 0

$E_{\theta_1} L(\theta_3, d_2)$= - 40(.3) – 40 (.3) – 40 (.3) – 40 (.3) = - 42

$E_{\theta_1} L(\theta_3, d_4)$ = - 85 (.3) – 85 (.3) – 85 (.3) = - 102

$E_{\theta_1} L(\theta_3, d_5)$ = 0 – 40 (.3) – 40 (.3) – 85 (.3) = - 49.5

$E_{\theta_1} L(\theta_3, d_6)$ = 0 – 40 (.3) – 85 (.3) – 85 (.3) = - 63

$E_{\theta_1} L(\theta_3, d_7)$ = 0 + 0 + 0 – 40 (.3) = -12

$E_{\theta_1} L(\theta_3, d_8)$= - 40 (.3) – 40 (.3) – 40 (.3) – 85 (.3) = -61.5

x / d	a	1	2	3	.1	.2	.3	Max.Loss
d_1	a_1	a_1	a_1	a_1	0	0	0	0
d_2	a_2	a_2	a_2	a_2	4	-32	-48	4
d_3	a_3	a_3	a_3	a_3	6	-28	-102	6
d_4	a_1	a_1	a_2	a_3	.25	-15	-37.5	2.5
d_5	a_1	a_2	a_2	a_3	3.5	-23	149.5	3.5
d_6	a_1	a_2	a_3	a_3	4	-22	-63	4

d_7	a_1	a_1	a_1	a_2	1	-8	-12	1
d_8	a_2	a_2	a_2	a_3	-61.5	4.5	-31	4.5

CONCLUSION:-
Minimum of maximum loss = 0.
d_1 is the correct decision.
Hence, admissible rule does't exist.
a_1 is the minimax decision rule.

Practical No.: 16
BAYE'S DECISION RULE

A drug company would like to introduce a drug to reduce acid indigestion. It is desirable to estimate θ, the proportion of the market share that this drug will capture. If in the past new drgs tend to capture a proporation between say 0.05 and 0.15 of the market and if all values in between are assumed equally likely. Then Θ has uniform distribution on {0.05.0.15}.

Obtain posteriori distribution and Baye*s rule. Assuming x = 15 and n = 90

*** *** ***

AIM:- To obtain posteriori distribution and Baye's rule of the given data.

FORMULA:- posteriori distribution is given by

$$n(\theta/x) = \frac{f(x,\theta)}{g(x)}$$

Where f (x,Θ) is the joint distribution of x and Θ. g (x) is the marginal distribution.

f(x,Θ) = ∏ (Θ) f (x/Θ).

Baye's rule d^* is given by

$d^* = \int \theta n(x/\theta) dx$

CALCULATIONS:-

$$\pi(\theta) = \frac{1}{.15-.05} = 10$$

$$f(x,\theta) = 10\binom{n}{x} B^x (1-\theta)^{n-x}$$

$$g(x) = \int_{.05}^{.15} (10)\binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

$$\therefore h(x/\theta) = \frac{f(x,\theta)}{g(x)}$$

$$= \frac{10\binom{n}{x}\theta^x(1-\theta)^{n-x}}{\int_{.05}^{.15} 10.\binom{n}{x}\theta^x(1-\theta)^{n-x}}$$

$$= \frac{\theta^x (1-\theta)^{n-x}}{\int_{.05}^{.15} \theta^x (1-\theta)^{n-x}}$$

Baye's decision rule is,

d* = $\int_{.05}^{.15} \theta h(x/\theta) dx.$

$$= \frac{\int_{.05}^{.15} \theta.\ \theta^x (1-\theta)^{n-x} d\theta}{\int_{.05}^{.15} \theta^x (1-\theta)^{n-x} d\theta}$$

When x = 15, n = 90,

$$\int_{.05}^{.15} \theta^{x+1} (1-\theta)^{n-x} d\theta$$

$$= \int_{.05}^{.15} \theta^{16} (1-\theta)^{75} d\theta$$

= β (16,76) [p ($x_1 \le .15$) – p ($x_2 \le .05$)]

Where x ~ β (16, 76)

= $[\beta(17,76)[p(u_1 \ge 17) - p(u_2 \ge 17)]]$

$$\beta(17,76)\left[\sum_{k\ge17}^{92} \binom{92}{k} (.15)^k (.85)^{92-k} - \sum \binom{92}{k} (.05)^k (.95)^{n-k}\right]$$

$$= \frac{\pi(.17)\pi(76)}{\pi(93)}.1762$$

$$= \int_{.05}^{.15} \theta^{15} (1-\theta)^{75} d\theta.$$

= $\beta(16,76)[p(x_1 \le .15) - p(x_2 \le .05)]$

= $\beta(16,76)\left[\sum_{k\ge16} \binom{91}{k} (.15)^k (.85)^{91-k} - \sum_{k\ge16} \binom{91}{k} (.05)^k (.95)^{91-k}\right]$

$$= \frac{\pi(16)\pi(76)}{\pi(92)}[.2611 - 0]$$

d* = $\frac{\pi(17)\pi(76)(.1762)}{\pi(93)} \Big| \frac{\pi(16)\pi(76)}{\pi(92)}(.2611)$

= $.1739 \times \frac{.1762}{.2611}$ =.1173541

Conclusion:-

Posteriori distribution = $\dfrac{\theta^x(1=-\theta)^{n-x}}{\int_{.05}^{.15}\theta^x(1-\theta)^{n-x}}$

$d^* = .1173541$

Practical No.: 17

TEST FOR HOMOGENITY OF SEVERAL VARIABLES

a) The following table given estimates of variances obtained from 8 samples of different sizes:

ni :	130	58	336	76	123	298
si :	36.238	50.908	41.0886	39.4928	30.411	40.3686
ni :	169	138				
si :	43.3968	38.2306				

Can the above sample variances be considered is considered as homogenious?

b) Test for Homogeneity of several correlations.

The following table gives correlations obtained from 10 samples of sizes 10, 14, 16, 20, 25, 28, 32, 35, 39 and 42 are as follows:

Sample:	1	2	3	4	5
r :	0.238	0.106	0.256	0.340	0.116
Sample:	6	7	8	9	10
r :	0.112	0.234	0.207	0.308	0.127

Can the correlations be considered as homogeneous?

Aim:- (a) To test whether the given sample variances can be considered as homogeneous or not.

FORMULA:- were the null hypothesis can be considered as H_o: sample variances are homogeneous.

The test statistic is,

$M^1 = \dfrac{M}{1 + c/3(k-1)} \sim \chi^2_{k-1}$

Where M = $\dfrac{2.3026}{\Lambda}\left[n\left(\log\sum n_i s_i^2 - \log_n^{10}\right) - \sum n_i \log_{10}^{s_i^2}\right]$

$n = \sum n_i$

$c = \sum \dfrac{1}{n_i} - \dfrac{1}{n}$

If $c_1 \leq \chi^2_{k-1} \leq c_2$ are accept the null hypotheses is otherwise reject the null hypothesis where c_1 and c_2 are standard χ^2 – table values, obtained from tables.

CALCULATIONS:-

n_i	s_i	$n_i s_i^2$	$1/n_i$	$\log s_i^2$	$n_i \log s_i^2$
130	36.238	4710.94	.007692	1.5592	202.696
58	50.908	2952.664	017241379	1.7068	98.9944

336	41.0886	13805.7696	.00297619	1.6137	542.2032
76	39.4928	3001.4528	.01315789	1.5965	121.3340
123	30.414	3740.5530	.008130081	1.4830	182.3967
298	40.3686	12029.8428	.0033557	1.6061	478.6178
169	43.3968	7334.0592	.005917159	1.6355	276.3995
138	38.2306	5275.8225	.00724637	1.5824	218.3712
1328		52851.1042	=06571709		2121.0251

M = 2.303 [1328 (4.7230 – 3.1232) – 2121.0251]
= 7.25445.
C = 0.0664 - .00075 = 06571709

$$M^1 = M|1 + c|\,3(k - 1)$$

= 7.25445

$$= \frac{7.25445}{1 + \frac{.06571709}{3 \times 7}}$$

= 7.2318

χ^2 – table values at 7 degrees of freedom are 1.69 and 16.09
1.69 < 7.2318 < 16.01.

INFERENCE:-

Since calculated χ^2 –value in between the table values we accept the null hypothesis.
Hence, the given sample variances are “Homogeneous”.

b.

AIM:- To test whether the given sample correlations are homogeneous or not.
H_o : sample corrections are homogeneous. The test statistic to be use is

$$M = \frac{T_2 - T_1^2}{N}$$

Where $T_2 = \sum (x_i - 3)\, Z_1^2$

$T_1 = \sum (x_i - 3) Z_1$

$N = \sum (x_i - 3)$

$$Z = \frac{1}{2} \log \frac{1 + r_i}{1 - r_i}$$

r_I is the i[th] sample correlation.

CALCULATIONS:-

X_i - 3	Z_i	$(x_i - 3)Z_i$	$(x_i - 3)\, z_i^2$
7	.242736	1.6991534	.412446
11	.106053	1.166583	12.37196
13	.2618511	3.404063	.8913579

17	.3540862	6.0194654	2.1314096
22	.11641665	2/561652	.2981621
25	.112	2.8	.3136
29	.238	6.902	1.642676
32	.2100	6.72	1.4112
36	.31835	11.4606	3.64838
39	1277	4.9803	.6359943
213		47.8307	11.509843

$$M = T_2 - \frac{T_1^2}{N}$$

= 11.509038 - 9.9037916

= 1.6052465

Table χ^2 – values at 5% level of significance are 2.70 and 19.02

INFERNCE:-

Since calculated χ^2 – value lies out side the χ^2 – table values, are reject H_o at 5 % level. Hence, we conclude that sample correlations are not homogeneous.

Practical No.: 17

SEQUENTIAL PROBABILITY RATIO TEST IN BINOMAL

BY SPRT method test (0.02.0.03) for the following data of F^* s and S^* s obtained sequentially from a Binomial population.

FFS FFS FFFF SF SSF SFFFFS FFS

AIM:- To test the given null hypothesis against alternative hypothesis by using S.P.R.T.. Binomial test procedure and to draw o.c. and A.S.N. Cuoves taking at least 5 points.

FORMULA AND PROCEDURE:-

Compute $a_m = h_o + s_m$

And $r_m = h_1 + s_m$. , m = 1, 2, ……… At the n^{th} step of $\sum x_i \geq r_m$ Reject $H_{o.}$ of $a_m < \sum x_i < r_m$. then continue the process by taking one more observation.

Where

$$h_o = \log\left(\frac{1-\beta}{1-\alpha}\right) \Bigg| \left(\log\frac{p_1}{p_0} - \log\frac{1-p_1}{1-p_0}\right)$$

$$h_1 = \log\left(\frac{1-\beta}{1-\alpha}\right) \Bigg| \left(\log\frac{p_1}{p_0} - \log\left(\frac{1-p_1}{1-p_0}\right)\right)$$

$$s = \log\left(\frac{1-p_1}{1-p_0}\right) \Bigg| \left(\log\frac{p_1}{p_0} - \log\frac{1-p_1}{1-p_0}\right)$$

(α,β) is the strength of the test.

O.C.Curve:- the o.c. function is given by

$$L(p) = \left(\frac{1-\beta}{\alpha}\right)^h - 1 \quad | \quad \left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h$$

Where h is determined by

$$P = 1 - \left(\frac{1-p1}{1-p0}\right)^h \quad | \quad \left(\frac{p1}{p0}\right)^h - \left(\frac{1-p1}{1-p0}\right)^h y \quad h \neq 0 \quad p \neq s.$$

When p = s.

$$L(P) = h_1 / |h_1 + |h_0|.$$

o.c. cure is obtained by drawing the graph, taking p on x – axis and L (p) on y – axis.

ASN Curve: the ASN function is given by

$$E(m) = \frac{L(p)\log B + (1 - L(p))\log(A)}{P\log\frac{p_1}{p_0} + (1-p)\log\frac{1-p_1}{1-p_0}}$$ When p $\neq$ s.

$$= \log B \log \ A \quad | \quad \log\frac{p_1}{p_0}\log\frac{1-p_1}{1-p_0}$$ if P = s

ASN curve is obtained on drawing the graph by taking p on X – axis and E (m) on Y – axis

CALUCULATIONS:-

$$h_0 = \frac{\bar{2}.4858}{.4771 - \bar{1}.8908}$$

= - 2.5826

$$h_1 = \frac{1.6857}{.4771 - \bar{1}.8908}$$

= 2.8751

$$s = \frac{.1 - 93}{.3679}$$

= 0.1864.

m	$\sum x_i$	a_m	r_m
1	0	- 2.3964	3.0614
2	0	- 2.21	3.2478
3	1	- 2.0236	3.4342
4	1	- 1.8372	3.6206
5	1	- 1.6508	3.8071
6	2	- 1.4644	3.9886
7	2	- 1.8372	4.1799
8	2	- 1.0916	4.3662
9	2	- 0.7188	4.7391

10	2	- 0.5324	4.9255
11	3	- 0.3460	5.1180
12	3	- .1596	5.2982
13	4	0.0268	5.4846
14	5	0.2132	5.6710
15	5	0.3996	5.6725
16	6	.5970	5.8574

At 16th step, $\sum \mathbf{x_i} > r_m$.

Hence we reject H_0.

$P_1 = .3 = p$

o.c. function.

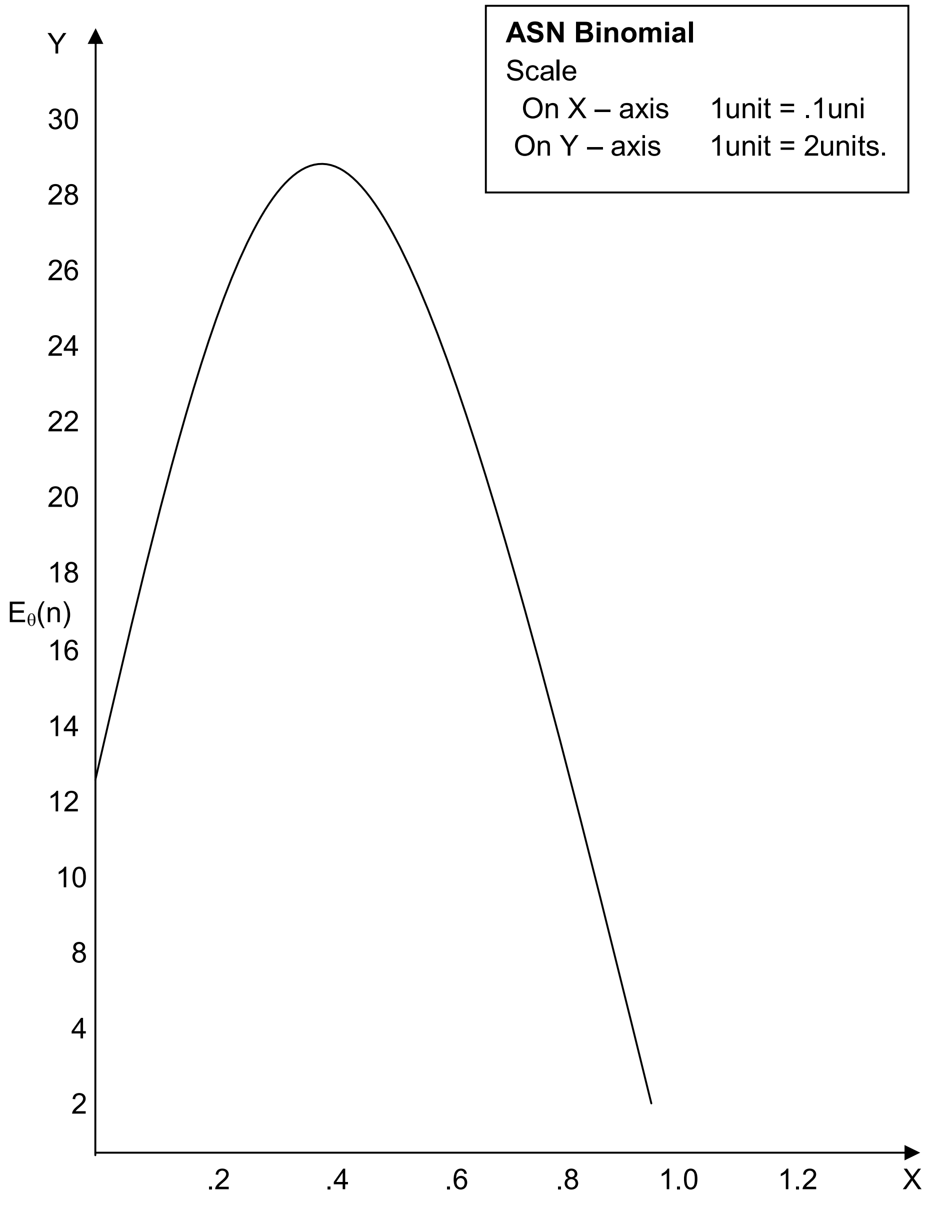
$L(0) = 1$

$L(P_0) = 1 - \alpha = .98$

$L(P_1) = \beta = .03$

$L(s) = \dfrac{h_1}{h_1 + |h_0|} = .52722$

$L(1) = 0.$



ASN Function.

$$E_p(n) \quad \frac{L(p)\log p(1-\alpha)+[1-L(p)]\log\frac{1-p}{\alpha}}{p\log\frac{p_1}{p_0}+(1-p)\log\left(\frac{1-p_1}{1-p_0}\right)}$$

When, L (0) = 1 E_p (n) = 13.8758

L (P_o) = 98 E_p (n) = 28.2800

L (P_1) = .03 E_p (n) = 24.815

L (1) = 0 E_p (n) = 3.533

P = s E_p (n) = 0.4663

INFERENC:-

At 16th step we reject H_o.

o.c and ASN functions are

p	**L (p)**	**E_θ (n)**
0	1	13.8758
.1	.98	28.2800
03	.03	24.8150
1	0	3.5331
.1862	.5272	0.4663

Practical No.: 18
SEQUENTIAL PROBABILITY RATIO TEST – NORMAL

By SPRT for N (0, 25) yest H_0: 0 = 135 Vs H_i 0 = 150 using the following sequential sample data and strength of the test (0,01,0,03).

151 144 121 137 138 136 155 160 144 145

130 120 104 140 125 145 106 125 138 120

108

Draw OC and ASN curves for the tost procedure choosing at least six points.

AIM:- To draw O.C. and ASN curves for the test procedure choosing at teat six points by SPRT N (0, 25) to test H_0: ϴ = 135 us H_1: ϴ = 150.

PROCEDURE:- the acceptance and rejection lines are given by $a_m = h_0 + s_m$

$r_m = h_1 + s_m$

Where $h_0 = \frac{\sigma^2}{\theta_1-\theta_0}\log\frac{\beta}{1-\alpha}$

$h_1 = \frac{\sigma^2}{\theta_1-\theta_0}\log\frac{1-\beta}{\alpha}$

$s = \frac{\theta_1 + \theta_2}{2}$

Conclusions are

If $\sum x_i \le a_m$ accept H_0 and stop the procedure

If $\sum x_i \ge r_m$ reject H_0 and stop the procedure

If $a_m < \sum x_i \le r_m$ continue the process. o.c. function is given by

$$L(\theta) = \frac{e^{2/\sigma^2 (s-\theta)h_1} - 1}{e^{\frac{2}{\sigma^2}(s-\theta)h_1} - e^{\frac{2}{\sigma^2}(s-\theta)h_0}} \qquad \bar{y}\ s \ne \theta$$

$$= \log 1 - \beta \left|\alpha\right| \log \frac{1-\beta}{\alpha} - \log \frac{\beta}{1-\alpha} \qquad \bar{y}\ s = \theta$$

By taking θ an x – axis, L (θ) on Y – axis draw o.c. curve. ASN function is given by

$$E_0(n) = \frac{L(\theta)(h_0 - h_1) + h_1}{\theta - s} \qquad \bar{y}\ s \ne \theta$$

$$= \frac{h_o\ h_1}{\sigma^2} \qquad \bar{y}\ s = \theta$$

By taking θ on X – axis, E_θ (n) on Y – axis draw ASN curve.

CALCULATIONS:-

$h_0 : = \frac{25}{150-135} \log \frac{.03}{.99}$ 2.303 = -5.8291

$h_1 : \frac{25}{150-135} \log \frac{.97}{.01}$ 2.303 = 7.6262.

$s = \frac{150+135}{2} = 142.5$

serial No: m.	$\sum x_i$	a_m	r_m
1	151	136.6709	150.1262

At the first stage $\sum x_i > r_m$.

$\therefore$ Reject H_o.

We have α = .01 β = .03 σ^2 = 25.

$$L(\theta) = \frac{e^{2/\sigma^2}(s-\theta)^{h_1} - 1}{e^{2/\sigma^2}(s-\theta)^{h_1} - e^{2/\sigma^2 (s-\sigma)h_o}} \text{ when } s \ne \theta.$$

We know that L (-∞) = 1 and L (∞) = 0 **.**

$L(\theta_0)$ = 1- α1 = 1.99

L (θ_1) = β = .03

$L(s) = \log \frac{1-\beta}{\alpha} / \log \frac{1-\beta}{\alpha} - \log \frac{\beta}{1-\alpha} = .5668$

L (140) = .83994

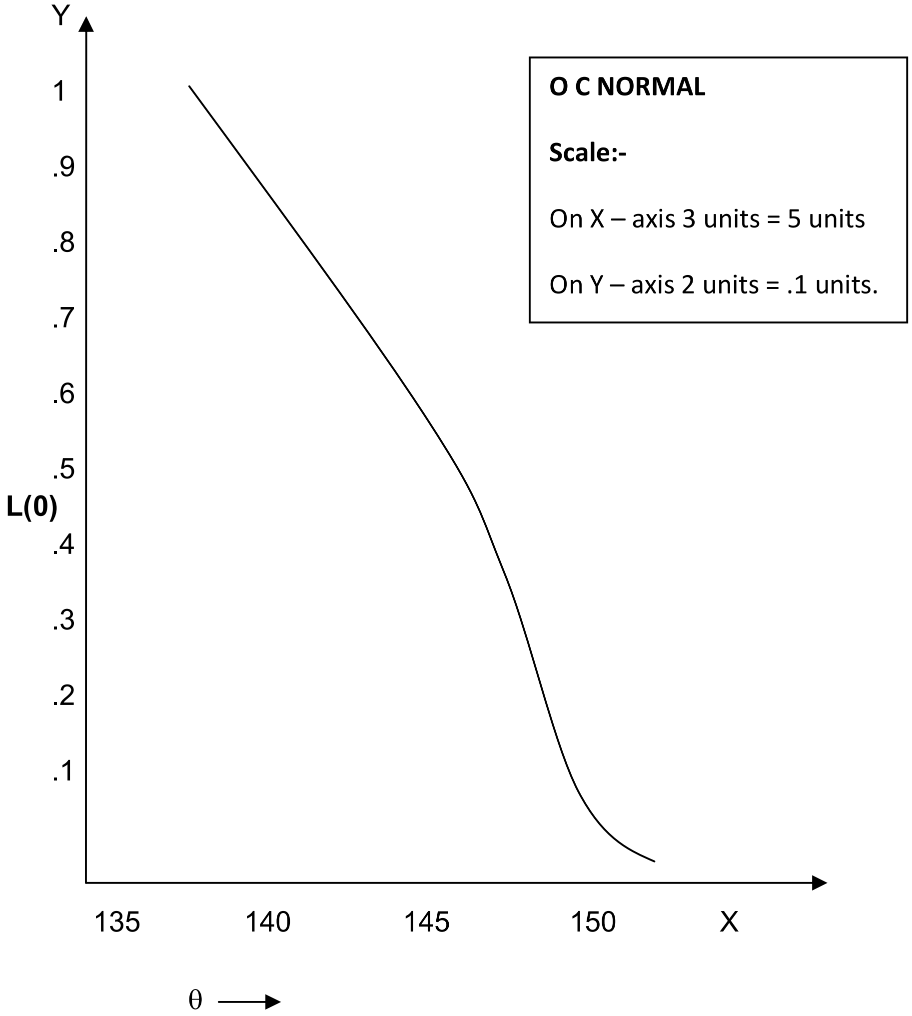
L (144) = .3752

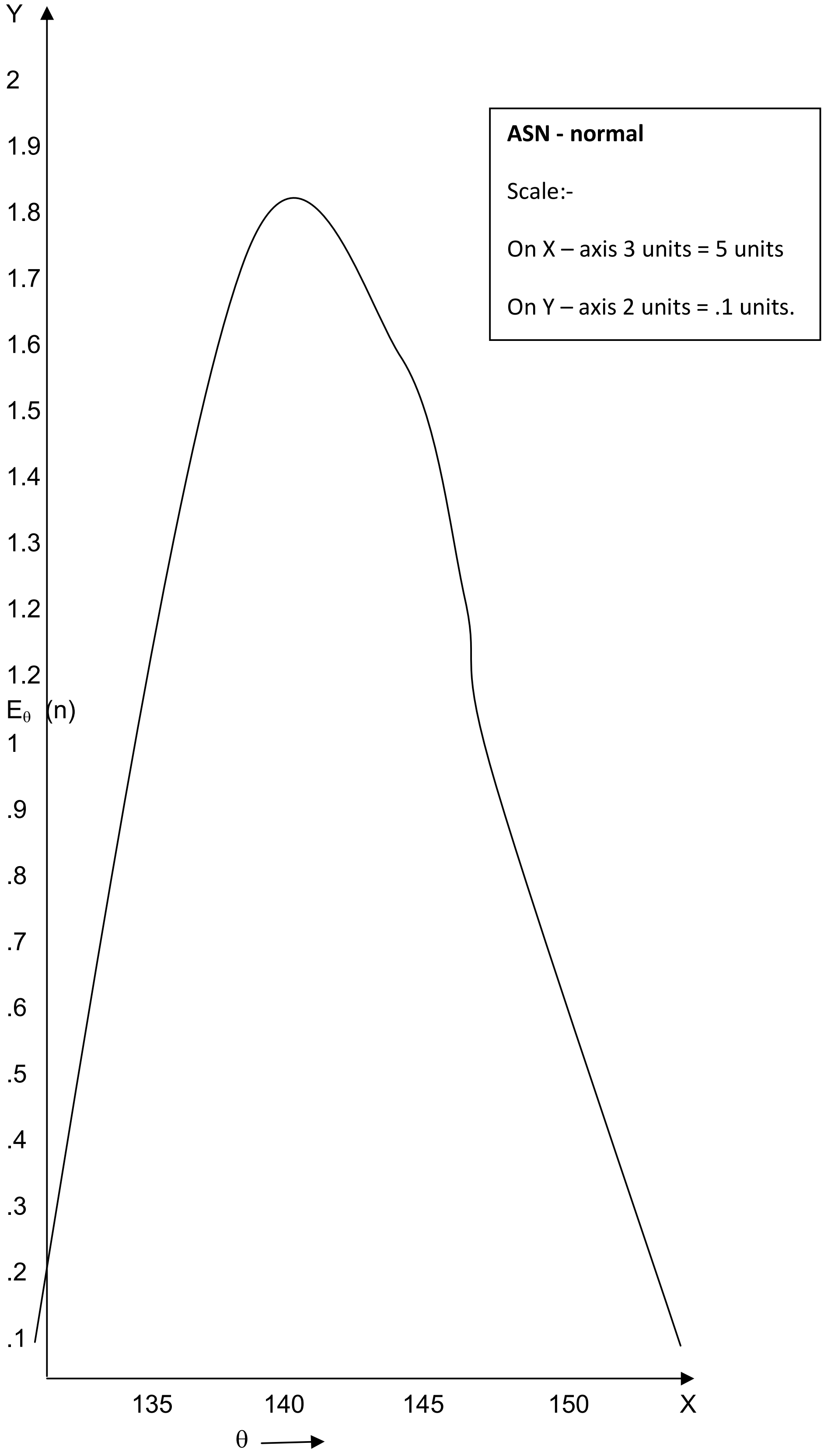
L (146) = .83559.

We have

$$E_{\theta} = (m) = \frac{L(p)\log B + (1 - L(p)\log A)}{P\log\frac{P_1}{P_o} - (1-p)\log\frac{1-p_1}{1-p_o}} \quad s \neq p.$$

$$= \frac{\log B \log p}{\log\frac{p_1}{p_0}\log\frac{1-p_1}{1-p_0}} \quad s = p$$




Y
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.2
E_θ (n)
1
.9
.8
.7
.6
.5
.4
.3
.2
.1
135
140
145
150
X
θ
ASN - normal
Scale:-
On X – axis 3 units = 5 units
On Y – axis 2 units = .1 units.

$\therefore$ E_n (1425) = + 1.7782
E_n (146) = - 1.03340
E_n (144) = 1.7185
E_n (148) = 1.20359
E_n (150) = .963005
E_n (135) = .7592729

INFERENCE:-
ASN and O.C Curve were drawn

Practical No.: 19
POWER CURVES

A) Draw the power curve for the MP test based on the sample size 10.
i) H_o: u = 4 Vs H_i : u > 4
ii) H_o : u = 4 Vs H_i : u < 4
where u is the mean of the Normal population having σ = 2 with level of significance 3 %.
B) Draw the power curves for testing :
i) H_o: 0 = 2 Vs H_i : 0 > 2
ii) H_o :0 = 2 Vs H_i : 0 < 2
in the distribution f (x;0) = 0 exp {-0x} with n = 1 and level of significance 5 %.

AIM:-
To draw the power curve for the test based on a sample of size 10
(a) H_0: μ = 4 Vs H_1 : μ > 4
(b) H_0: μ = 4 Vs H_1 : μ < 4.

PROCEDURE:-

(a) Z = $\frac{\mu - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

Test function for testing H_0 : μ = 4 Vs H_1 : μ > 4 is

$\varphi\ (x) = 1 \quad y \quad \bar{x} > c.$

= 0 other wise.

Where c is, $P\{\bar{x} > c/H_0\}$ = . 05

Power function is $\beta_\varphi(\theta) = P\{\bar{x} > C/H_1\}$

$= 1 - \Phi\left[\frac{c-\mu}{\sigma/\sqrt{n}}\right]$

Where c = μ_0 + 1.64 $\frac{\sigma}{\sqrt{n}}$.

Power curve is a graph obtained by drawing a graph. Taking μ on X – axis, $\beta_\varphi(\theta)$ on Y – axis.

(b) the test for testing H_o μ = 4 Vs H_1 : μ < 4 is

$$\varphi\ (x) = 1 \quad y \quad \bar{x} < c.$$

= 0 other wise

Where c is given by $P\left[\bar{x} < c \middle| H_o\right]$ = .05

Power function p (μ) = $P\left[\bar{x} < c \middle| H_1\right]$.

$$= \Phi\left[\frac{c-\mu}{\sigma/\sqrt{n}}\right]$$

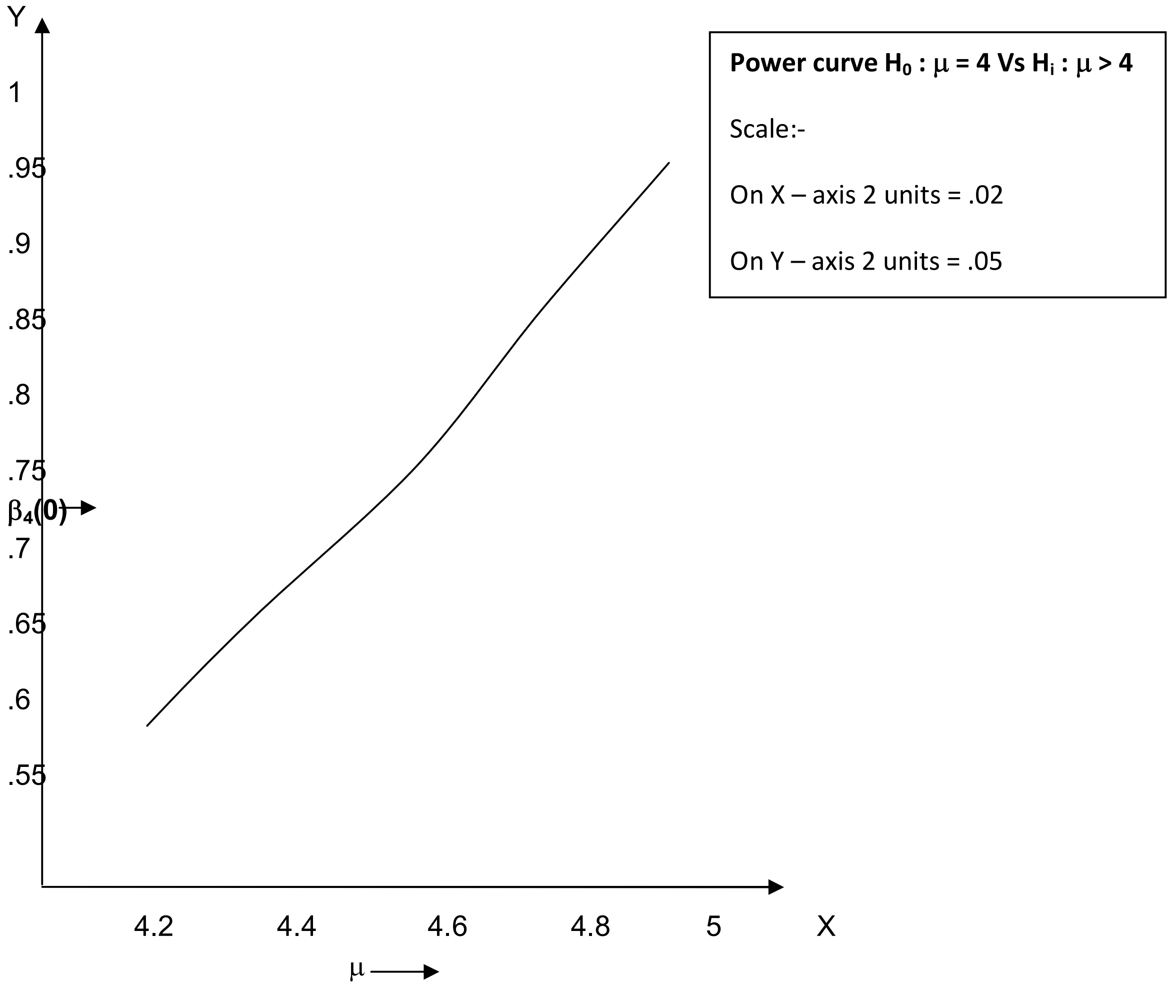
Draw a graph between μ and P (μ).

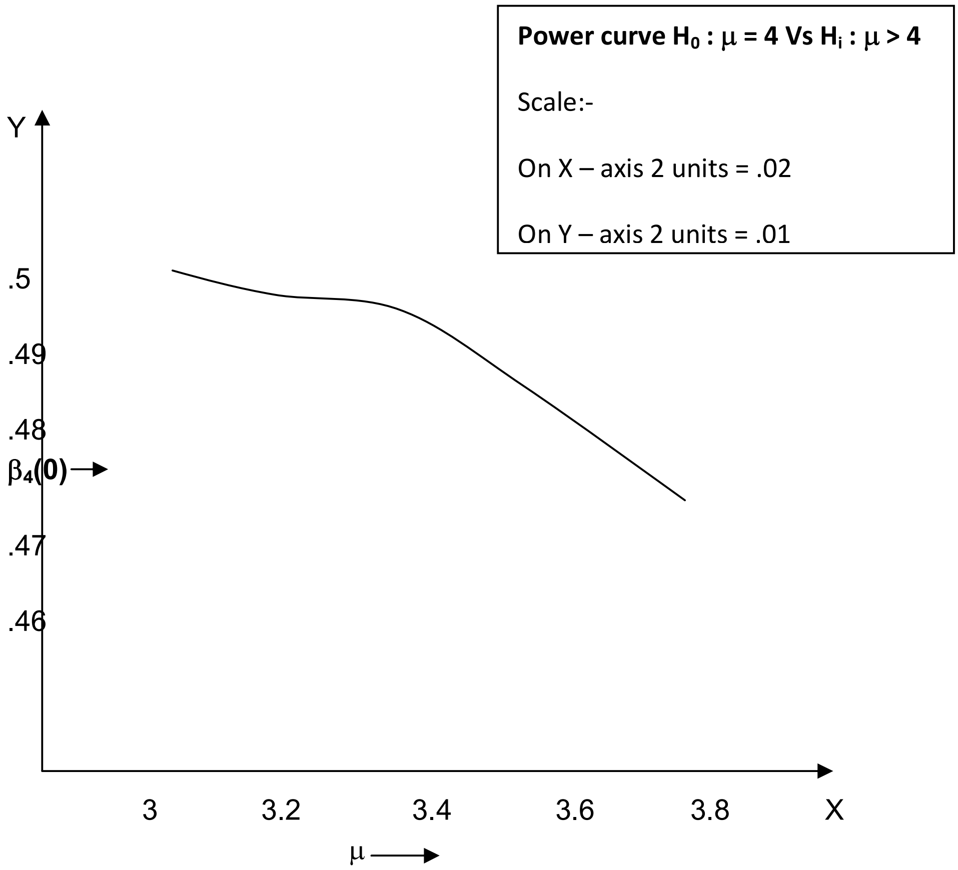
CALCULATIONS:-

(a) c = $\mu_0 + 1.64 \frac{\sigma}{\sqrt{n}}$

$$= 4 + 1.64\ \frac{2}{\sqrt{10}} = 5.\ 0372$$

(b) C = $\mu_0 - 1.64 \frac{2}{\sqrt{10}}$ = 2.9628.





(a)

μ	$\frac{c-\mu}{\sigma/\sqrt{n}}$	$\phi\left[\frac{c-\mu}{\sigma/\sqrt{n}}\right]$	**P (μ)**
4.2	1.3282	.407911	.592089
4.4	1.0120	.344231	.655769
4.6	.6957	.256472	.743528
4.8	.3795	.147656	.852344
5	.0632	.025117	.974883

(b)

μ	$\frac{c-\mu}{\sigma/\sqrt{n}}$	**P (μ)**
3.8	1.9607	.475002
3.6	2.2770	.488607
3.5	2.5932	.495243
3.2	2.9095	.498187
3	3.2258	.499359

INFERENCE:-

The curves are drawn on the graph sheets.

EXPONETIAL DISTRIBUTION

AIM:- To draw the power curves for testig the hypothesis

(a) $H_0 : \theta = 2$ Vs $H_1 : \theta > 2$

(b) $H_0 : \theta = 2$ Vs $H_1 : \theta < 2$.

Over the distribution $f(x, \theta) = e^{-\theta x}$ $\theta > 0$, $0 < x < \infty$.

PROCEDURE:- (a) According to NP lemma the test function to test the hypothesis $H_0 : \theta = 2$ Vs $H_1 : \theta > 2$ is $\phi(x) = 1$ y $x > c_1$

= 0 otherwise

Where c_1 is given by $P[x < c_1 | H_0] = .05$.

Power function is $p(\theta) = 1 - e^{-\theta c_1}$

Taking the values of θ n X- axis, p (θ) on Y – axis we draw a curve. The curve is power curve

(b) To test the hypothesis $H_0 : \theta = 2$ Vs $H_1 : \theta < 2$, the lest function is

$\phi(x) = 1$ y $x < c_2$

= otherwise

Where c_2 is given by $P[x < c_2 | H_0] = .05$

Power function is $e^{-\theta c_1}$

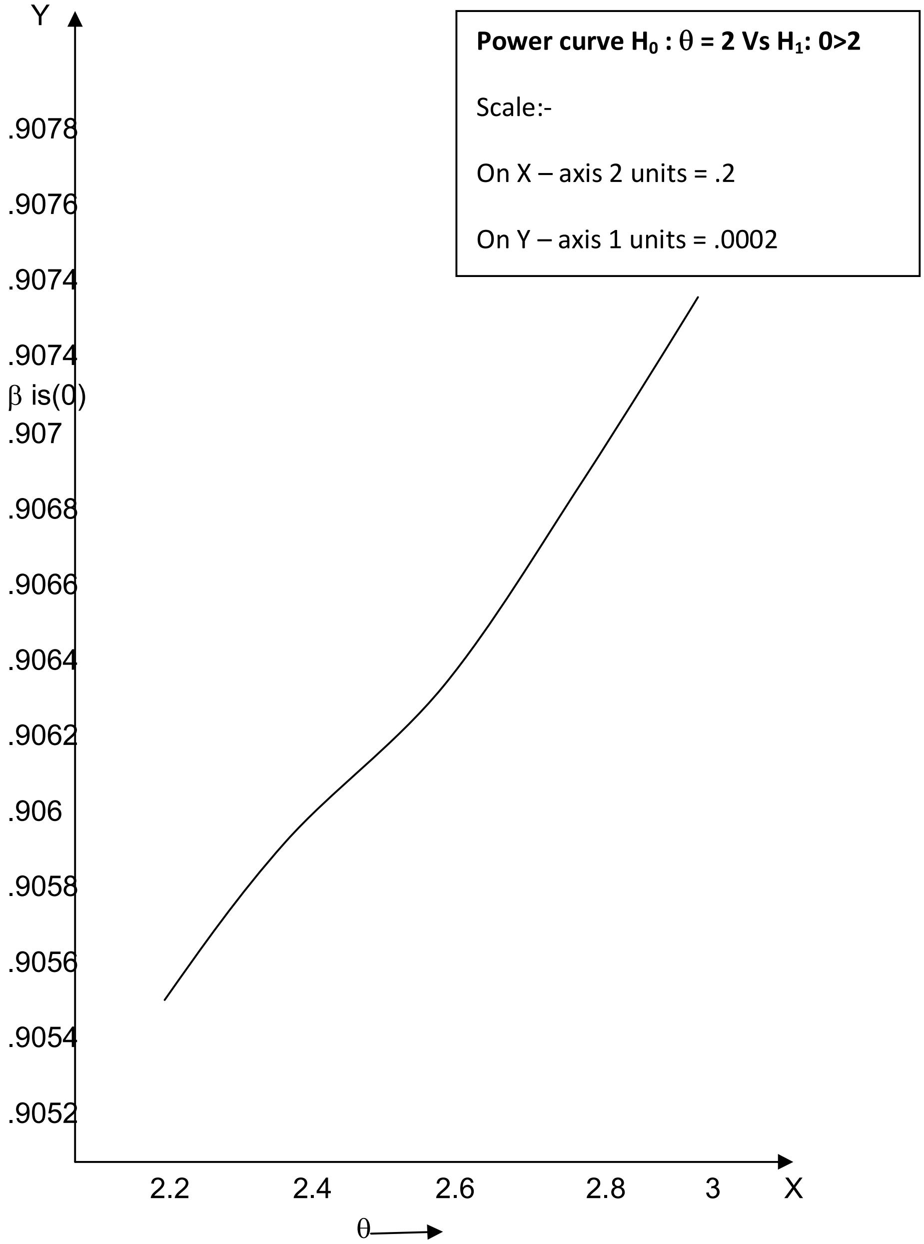
Power was drawn as above.

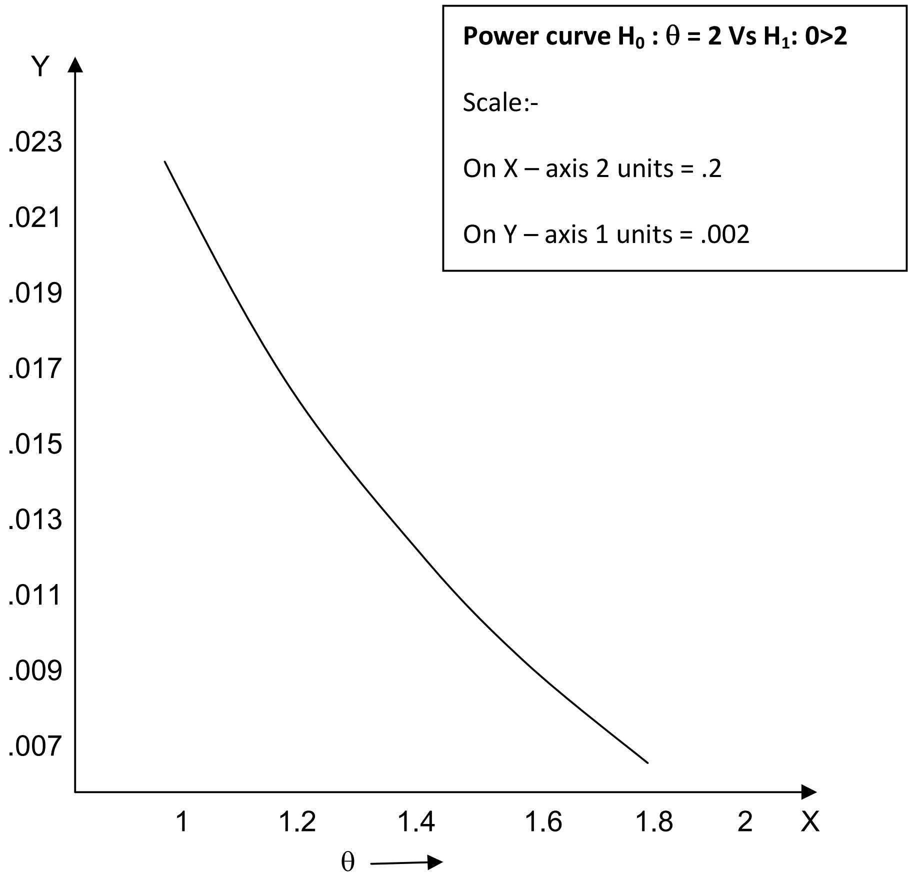
CALCULATIONS:- (a) c_1 is given by

$$\int_0^{c_1} \theta_0 e^{-\theta_0 x} dx = .05$$

$\Rightarrow c_1 = .02567$

θ	$-\theta c_1$.	$e^{-\theta c_1}$.	$1 - e^{-\theta c_1}$
2.2	-.056474	.09450	.90545
2.4	-.061608	.09404	.90596
2.6	- .071876	.09356	.90644
2.8	- .066742	.09307	.90693
3	- .07701	.09260	.90740


Power curve H0 : θ = 2 Vs H1: 0>2
Scale:-
On X – axis 2 units = .2
On Y – axis 1 units = .0002
Y
.9078
.9076
.9074
.9074
β is(0)
.907
.9068
.9066
.9064
.9062
.906
.9058
.9056
.9054
.9052
2.2
2.4
2.6
2.8
3
X
θ



(b) c_2 is given by,

$$p[x > c_2 | H_0] = .05$$

$$\int_{c_2}^{\infty} \theta_0 e^{-\theta_0 x} dx = .05$$

$\Rightarrow c_2 = 1.40781$

θ	$-\theta\, c_2$	$e^{-\theta c_2}$
1.8	- 2. 696058	.006747
1.6	- 2.396496	.009035
1.4	- 2.096934	.01257
1.2	- 1.797372	.01658
1	- 1.49781	.02237

INFERENCE:-

The power curves are drawn on the graph sheets.

Practical No.: 20
Mann- Whitney-Wi1coxon Test

1. Suppose two drugs, A and B are being compared. The minutes until pain relief recorded are given below. Is the number of minutes until pain relief, the same for both drugs at 5 % level?

Drug A :	9	11	15	
Drug B :	6	8	10	13

2. In order to compare the breaking strength of nylon fiber produced by two different manufacturers. 10 measurements on one (say x) and 13 on the other (say y) were taken with the following results.

Fiber x : 1.7 1.9 1.8 1.1 0.7 0.9 2.1 1.6 1.7 & 1.3
Fiber y : 2.1 2.7 1.6 1.8 1.7 1.8 1.6 2.2 2.4 1.3 1.9 & 1.8 2.0

Do the data indicate a significant difference between the breaking strengths?

a.

AIM:- To test where the time until pain relief of two drugs are the same or not.

PROCEDURE:- Here, the Hypothesis can be set as

$H_0 : s = 0$ $H_1 : s > 0$

$H_0 : s = 0$ $H_1^1 : s < 0$

$H_0 : s = 0$ $H_1^{11} : s \neq 0$ where s is the difference of time pain relief time minutes.

To test the hypothesis, pool the observations and rank them at $T = \sum R(x_i)$.

Where R is Rank.

Let $U = T - \frac{n(n+1)}{2}$

Reject H_0 : (H_1) $\bar{y}$ $u \geq a$

Reject H_0 : in favour of H^1 $\bar{y}$ $\mu \leq b$.

Reject H_0 : in favour of H^1 $\bar{y}$ $\mu \geq c$.

Where a, b, c are obtained from theoretical values.

CALCULATIONS:-

A : 9 11 15

B : 6 8 10 13

Pooling A & B.

		A		A		A
6	8	9	10	11	13	15

Ranks :	1	2	3	4	5	6	7

T = 3+ 5 + 7 = 15

U = 15 - $\frac{3 \times 4}{2}$ = 9.

The various combinations of A's and β's are,

B	A	A	B	B	A	B
A	A	B	B	B	A	A
A	B	A	B	A	B	
A	B	B	A	A	B	B
A	B	B	A	A	B	A
A	A	B	B	B	A	B
A	A	B	A	B	B	B
A	A	A	B	B	B	B
B	A	A	A	B	A	B
B	A	B	A	B	A	B
B	A	A	B	A	B	B
B	A	A	B	B	B	A
B	A	B	B	A	A	B
B	A	B	A	B	B	A
B	A	B	B	B	A	A
A	B	A	B	B	B	A

A	B	B	A	B	A	B
A	B	B	A	A	B	B
A	B	B	A	B	B	A
A	A	B	B	B	B	A
A	A	B	B	A	B	B
B	B	A	A	A	B	B
B	B	A	B	A	B	A
B	B	A	B	A	A	B
B	B	A	A	B	A	B
B	B	B	A	A	A	B
B	B	B	A	B	A	A
B	B	B	A	A	B	A
B	B	B	B	A	A	A
A	B	A	B	B	A	B
A	B	A	B	B	A	B
A	B	A	A	B	B	B
B	A	B	A	A	B	B

u	**. +**	**. P (u)**
0	1	1/35
1	1	1/35
2	2	2/35
3	3	3/35
4	4	4/35
5	4	4/35
6	5	5/35
7	4	4/35
8	4	4/35
9	3	3/35
10	2	2/35
11	1	1/35
12	1	1/35

From this tabte,

$$c_1 = \frac{1}{35} = 1028$$

$$c_1 = \frac{1}{35} = 12.08$$

Calculated v value lies between c_1 & c_2.

INFERENCE:-

Pain relief hours of the two drugs are the same.

b. AIM ;- To compare the breaking strength of nylon fiber produced by two different manufactures.

PROCEDURE:- Here the null hypothesis can be set as H_o : no difference in breaking strength. To test the null hypothesis. The test statistic used is, under H_o :

$$Z = \frac{U - E(u)}{\sqrt{V(u)}} \sim N(0,1)$$

Where u = T - $\frac{m(m+1)}{2}$

E (u) = $\frac{mn}{2}$

V (u) = $\frac{mn(m+n+1)}{12}$

T = $\sum$ R (x_i)

Where R is the rank obtained by giving ranks to the combned variables.

M is the sample sige of I sample n is the II sample size.

If calculated Z $\leq$ 1.96 we accept the at 5 % level.

CALCULATIONS:-

Combined sample:- .7 .9 1.1 1.3 1.3 1.6 1.6 1.6 N 1.7
1.7 1.7 1.8 1.8 1.8 1.8 1.9 1.9 2 2.1 2.1 2.1
2.2 2.4 2.7

Ranks: 1 2 3 4.5 4.5 8 8 8 10 10 10
13.5 13.5 13.5 13.5 13.5 16.5 18 19.5 19.5 21 22 23

T = $\sum$ R (x_i)

= 1 + 2 + 3 + 4.5 + 10 + 10 + 13.5 + 16.5 + 19.5 = 87

U = T - $\frac{m(m+1)}{2} = 87 - \frac{10 \times 11}{2} = 32.$

E (u) = $\frac{mn}{2} = \frac{10 \times 13}{2} = 65$

V (u) = $\frac{10 \times 13 \times 24}{12}$ = 260

$$z = \left| \frac{32 - 65}{`16.124515} \right| = 2.04657$$

Since, 2.04657 > 1.96, we reject H_0 at 5% level of significance.

INFERECE:- there is significant difference between the breaking strength of two manufacturers.

Practical No.: 21

WILCOXON SIGNED RANK TEST

a) A certain universities brochure claims that the amount of money needed for boarding and lodging in the hostel for a single student is Rs/..75 per week. A random sample of size 9 students from this University showed the following weekely expenditure.

75 92 80 73 84 60 84 91 78

Is there evidence to suggest that the University's estimate is not currect?

b) In order to determine if children en watching more TV in preteen years, a random sample of 20 children aged 9, 10 or 11 was selected and their daily average TV viewing times were recorded. The same children were the covered by years later and their daily average viewing time was recorded (children's daily TV viewing times in hrs.)

NO:	1	2	3	4	5	6	7	8	9
Pre:	3.5	2.8	4.6	3.7	3.6	4.2	2.2	1.6	3.6
Teen:	4.2	2.2	5.2	2.1	0.5	5.4	7.2	1.0	2.8
No:	10	11	12	13	14	15	16	17	18
Pre:	5.0	3.0	4.8	1.5	2.5	3.2	3.4	1.2	0.5
Teen:	4.6	4.0	2.2	1.3	2.5	3.0	2.6	2.6	2.3
No:	19	20							
Pre:	1.8	3.5							
Teen:	o.5	2.7							

How strong is the evidence that the TV watching habits in preteen and then years is the same?

AIM:- To test whether there is any evidence to suggest that the universities estimate is currect or not.

PROCEDURE:-

Suppose the hypothesis to be tested is

H_o : m = 75 Vs H_1 : m $\neq$ 75.

To test the hypothesis calculate the deviations $X_i - m_o$, give ranks to abselute deviations. Let T be the sum of the ranks of positive deviations.

If this T lies in between the critical values obtained from tables we accept the null hypothesis, otherwise reject.

CALCULATIONS:-

x_I	x_I -75	$\lvert x_i - 75\rvert$	Ranks
75	0	0	1
92	17	17	9
80	5	5	4
84	9	9	5.5
73	-2	2	2
60	-15	15	7
84	9	9	5.5
91	16	16	8
78	3	3	3

T = 36

Critical values are 4 and 32

INFERENCE:-

Because T lies outside the critical values we reject H_o .

So, there is evidence to suggest that the university's is not correct.

b. AIM:-

To test whether there is any difference in TV watching children in preteen tyears and in teen years, or not.

PROCEDURE:-

Here, the hypotheses can be set as $H_o : m_1 - m_2 = 0$ Vs $H_1 : m_1 - m_2 \neq 0$.

To test the hypotheses calculate the deviations $x_1 - y_i$. given ranks to +ve deviations. Let T be the sum of the positive deviations.

$$E(T) = \frac{n(n+1)}{4}$$

$$V(T) = \frac{n(n+1)\,(2n+1)}{24}$$

$$Z = \frac{T - E(T)}{\sqrt{VT}} \sim N(0, 1).$$

If $Z \leq 1.96$ we accept the null hypothesis at 5% level of significance, otherwise reject the null hypotheses.

CALCULATIONS:-

x_i	y_i	X_i - y_i	$\lvert x_i - y_i\rvert$	Ranks
3.5	4.2	-.7	.7	9
2.8	2.2	.6	.6	7
4.6	5.2	-.6	.6	7

3.7	2.1	1.6	1.6	16
3.6	.5	3.1	3.1	19
4.2	5.4	-1.2	1.2	13
2.2	2.2	0	0	1.5
1.6	1	.6	.6	7
3.6	2.8	.8	.8	11
50	4.6	.4	.4	5
3	4	-1	1	12
48	2.2	2.6	2.6	18
1.5	1.3	.2	.2	3.5
2.5	2.5	0	0	1.5
3.2	3	.2	.2	3.5
3.4	2.6	.8	.8	11
1.2	2.6	-1.4	1.4	15
.5	2.3	-1.8	1.8	17
1.8	.5	1.3	1.3	14
3.5	2.7	.8	.8	11

T = 129.

E (T) = $\frac{18 \times 19}{4}$ = 85 .5

V (T) = $\frac{18 \times 19 \times 37}{24}$ = 527.25

$|Z| = \left|\frac{129 - 85.5}{22.9619}\right|$ = 1.8944.

CONCLUSION:-

Since $|Z|$ < 1.96 we accept H_o at 5 % level of significance. Hence, the TV watching habits in preteen and teen years is the same.

Practical No.: 22 `

KOLMOGOROV – SMIRNOV TEST

Test the null hypothesis that the following observations came from

Fo (x) = 0 if x < 0
= x if 0 < x < 1
= 1 if x > 1

0.59	0.72	0.47	0.43	0.31	0.56	0.22	0.90	0.96
0.78	0.66	0.18	0.73	0.43	0.58	0.11		

AIM:-
To test whether the given sample observations come from the given distribution or not.

FORMULA:- Hypotheses to be tested is $H_0 : F = F_0$ Vs $H_1: F \neq F_0$.
To test the hypothesis the following computations are to be made.

$D^+ = \max \left\{ \frac{\Lambda}{n} - F_0(x_i) \right\}$

$D^- = \max \left\{ F_0(x_i) - \frac{x-1}{n} \right\}$.

$D = \max \{D^+, D^-\}$ i = 1, 2, ……..

If this calculated D value is less than or equal to the table value, accept the null hypothesis otherwise reject the null hypothesis at 5% level of significance.

CALCULATIONS:-

x_i	$F_0(x_i)$	$D^+ = \frac{i}{n} - F_o(x_i)$	D^-	Max (D^+, D^-)
.11	.11	-.0475	.11	.11
.18	.18	.01175	.1175	.1175
.22	.22	-.6325	.095	.095
.31	.31	-.18096	.1225	.1225
.43	.43	-.055	.1175	.1175
.47	.47	-.0325	.095	.095
.56	.56	-.06	.1225	.1225
.58	.58	-.0175	.08	.08
.59	.59	-.035	.0275	.035
.72	.72	.03	.0325	.0325
.73	.73	.0825	-.02	.0825
.78	.78	-.095	-.0325	-.095
.90	.90	-.0375	.025	.0375
.96	.96	.04	.0225	.04

$D = \max \{D^+, D^-\}$ = .1225
Table value at 5% level of significance = .328

INFERENCE :- Since calculated value is less than table value we accept H_o at 5 % level of significance.
Hence the given samples follow the given distribution function.

2.
AIM:- To test whether there there is any difference in the life – times of two brands of batteries.
PROCEDURE:- Hence the Hypotheses to be tested is

H_o : F = G Vs H_1 : F ≠ G.
To test the hypotheses, the following steps are adepted.
Find the order statistics x_i i = 1,…….n.

Then, calculate. $\hat{F}(x_i) = ^{\neq} \frac{x_i \leq x}{n}$ I = 1,2……n

$\hat{G}(x_i) = \pm y_i \leq y/n.$ I = 1,2,..n

Calculate (x, X) = max $\{\hat{F}(x), \hat{G}(x_i)\}$

Then, D = max $\{\max(\hat{F}(x_i)), \hat{G}(x_i)\}$

Compare this D value with the table value. If D ≤table values accept H_o at 5 % level of significance otherwise reject null hypothesis.

CALCULATIONS:-

x_i	$\hat{F}(x_i)$	$\hat{G}(x_i)$	$\mathrm{Max}(\hat{F}, \hat{G})$
30	2/6	0	2/6
40	4/6	1/6	4/6
45	5/6	2/6	5/6
50	5/6	4/6	5/6
55	6/6	5/6	1
60	6/6	6/6	1

D = max $\{\max(\hat{F}(x_i)), \hat{G}(x_i)\}$

= 1

Table value at 5% level of significance = 5

CONCLUSION:-

Since calculated value is greater then. Table value reject H_o at 5 % level of significance.

Hence, are conclude that the two different brands of batteries are different w.r.t their average life-times.

Practical No.: 23

The following table is prepared on the basic of two independent samples.

	Sample 1	Sample 2
No. of observations above the combined median	7	17
No. of observations below the combined median	15	10

Apply the median test for testing the hypothesis that both samples come from the same population by using

a) Chi –Square approximation and

b) Normal approximation.

Compare the results of (a) and (b).

AIM:-

To apply median test for testing the hypothesis that both the samples come from same population by using (a) χ^2 – approximation (b) normal approximation

FORMULA:- Hence, the null hypothesis can be test as H_o : the two samples come from population. (a) By χ^2 – approximation to test H_o, the test statistic to, be used is,

$$\chi^2 = \frac{\sum o_i^2}{e_i} \text{ - N . ~ } \chi_1^2 .$$

where o_i is the observed frequency and e_i is the estimated frequency, N is grand total. If χ^2 calculated value is less than or equal to χ^2 – table value are accept H_o

(b) Suppose S is the no. of observations greater than combined median. Then,

$$E(S) = \frac{m(N-1)}{2N} \qquad V(S) = \frac{mn(N+1)}{4N^2}$$

Where m is the I sample sixe . by normal approximation method, under H_o

$$Z = \frac{S - E(S)}{\sqrt{V(S)}} \sim N(0,1)$$

If, calculated Z – value is less than or equal to 1.96 we accept H_o at 5 % level of significance.

CALCULATIONS:-

	I sample	II sample	Total
No. of observations above S	7	17	24
No. of observation below S	15	10	25
Totals	22	27	42

$$\chi^2 = \frac{49(70-225)^2}{25 \times 24 \times 22 \times 27} = 4.7025.$$

Table χ_1^2 (.05) = 3.84.

(b) $E(s) = \frac{m(N-1)}{2N} = \frac{22 \times 48}{2 \times 49} = 10.7755.$

$$V(s) = \frac{mn(N+1)}{4N^2}$$

$$= \frac{22 \times 27 \times 50}{4 \times (49)^2} = 3.0925$$

S = 24

$$Z = \frac{24 - 10.\quad 7755}{1.7586} = 7.5197.$$

INFERENCE:-

(a) By χ^2 – approximation,

Calculated value > table value. Hence we reject H_o at 5 % level of significance.

(b) By normal approximation,

Calculated Z – value >. Table value at 5 % level. Hence we reject H_o at 5% level. In both the cases. We are rejecting H_o . Hence, we conclude that the two samples not come from the same population.

Practical – 24
Stratified Random Sampling

1. The following data shows this stratification of all the farms in a country by farm size and average acres farm per farm in each stratum, for a sample of 100 fa. Compute the sample size in each stratum under

i. Proportional allocation

ii. Neymann allocation

Compare the precision of these methods with the SRS when finite population correction is ignored.

Farm size (acres)	No. of farms	Average corn acres	S.D (Sh)
0 – 40	394	5.4	8.3
41 – 80	461	16.3	13.3
81 – 120	391	24.3	15.1
121 – 160	334	34.5	19.8
161 – 200	169	42.1	24.5
201 – 240	113	50.1	26.0
241 – above	148	63.8	35.2
Total/Mean	2010	26.3	

Aim:-

To compute the sample sizes of each stratum from proportional allocation and Neymann allocation and also compare the precision of these methods with that of simple Random sampling when fpc is ignored.

Procedure:-

$$V_{prop} = \frac{1}{nN} \sum_h N_h S_h^2$$

$$V_{ney} = \frac{1}{nN} \left(\sum_h N_h S_h^2 \right)^2$$

$$V_{Ran} = \frac{1}{n} \left[\frac{\sum (N_h - 1) S_h^2 + \sum N_h (\overline{Y}_h - \overline{Y})^2}{N - 1} \right]$$

Relative precision for proportional allocation is

$$\frac{\frac{1}{V_{prop}}}{\frac{1}{V_{Ran}}} \times 100$$

Relative precision for Neymann allocation is

$$\frac{\frac{1}{V_{ney}}}{\frac{1}{V_{ran}}} \times 100$$

Proportional allocation is $n_h = (\frac{n}{H})\ N_h$

Neymann allocation is $n_h = \frac{nN_h\ S_h}{\sum_h N_h\ S_h}$

Calculation:-

Stratum No.	N_h	S_h	$N_h\ S_h$	S_h^2	$N_h S_h^2$
1	394	8.3	3270.2	68.89	27142.66
2	461	13.3	6131.3	176.89	81546.29
3	391	15.1	5904.1	228.01	89151.91
4	334	19.8	6613.2	392.04	130941.36
5	169	24.5	4140.5	600.25	101442.25
6	113	26.0	2938.0	676	76388
7	148	35.2	5209.6	1239.04	183377.92
Total	= 2010	= 142.2	= 34206.9	= 3381.12	= 689990.39

Stratum No.	N_h	$\overline{Y}_h$	$N_h \overline{Y}_h$	$(\overline{Y}_h - \overline{Y})^2$	$N_h(\overline{Y}_h - \overline{Y})^2$	$S_h^2 (N_h - 1)$
1	394	5.4	2127.6	437.2616	172281.0704	27073.77
2	461	16.3	7514.3	100.2161	46199.6221	81369.40
3	391	24.3	9501.3	4.0433	1580.9303	88923.90
4	334	34.5	11523	67.0630	22399.0420	130549.30
5	169	42.1	7114.9	249.2988	42131.4972	100842.00
6	113	50.1	5661.3	565.9260	63949.6380	75712.00
7	148	63.8	9442.4	1405.4401	208005.1348	182138.88
Total	= 2010	= 236.5	= 52884.8		=556546.9348	=686609.20

$$\overline{Y} = \frac{\sum N_h\ \overline{Y}_h}{N}$$

$$= \frac{52884.8}{2010} = 26.3108$$

$$V_{prop} = \frac{1}{nN} \sum_h N_h\ S_h^2$$

$$= \frac{1}{100(2010)} (689990.39)$$

$= 3.4328$

$$V_{ney} = \frac{1}{nN^2} \left(\sum (N_h S_h)\right)^2$$

$$= \frac{1}{100(2010)^2} (34206.9)^2$$

$$= 2.8962$$

$$V_{Ran} = \frac{1}{n} \frac{\sum (N_h - 1)S_h^2 + \sum N_h (\overline{Y}_h - \overline{Y})^2}{N - 1}$$

$$= \frac{1}{100} \left[\frac{686609.27 + 556546.9348}{2010 - 1}\right]$$

$$= 6.1879$$

Relative precision for proportion allocation

$$\frac{\frac{1}{V_{prop}}}{\frac{1}{V_{ran}}} \times 100 = \frac{\frac{1}{3.4328}}{\frac{1}{6.1879}} \times 100$$

$$= 180.26\%$$

Relative precision for Neymann allocation

$$\frac{\frac{1}{V_{ney}}}{\frac{1}{V_{ran}}} \times 100 = \frac{\frac{1}{2.8962}}{\frac{1}{6.1879}} \times 100$$

$$= 213.6558\%$$

Proportional allocation $n_h = \left(\frac{n}{N}\right) N_h$

$$= \left(\frac{100}{2010}\right)(394)$$

$$= 19.6020$$

Neymann allocation $n_h = \frac{nN_h S_h}{\sum_h N_h S_h}$

$$= \frac{100(3270.2)}{34206.9}$$

$$= 9.5600$$

Stratum No.	N_h	$N_h S_h$	Proportion allocation (n_h)	Neymann allocation (n_h)
1	394	3270.2	19.6020 $\simeq$ 20	9.5600 $\simeq$ 10
2	461	6131.3	22.9353 $\simeq$ 23	17.9242 $\simeq$ 18
3	391	5904.1	19.4527 $\simeq$ 19	17.2599 $\simeq$ 17
4	334	6613.2	16.6169 $\simeq$ 17	19.3329 $\simeq$ 19
5	169	4140.5	8.4080 $\simeq$ 8	12.1042 $\simeq$ 12
6	113	2938.0	5.6219 $\simeq$ 6	8.5889 $\simeq$ 9
7	148	5209.6	7.3632 $\simeq$ 7	15.2296 $\simeq$ 15
			100	100

Conclusion:-

V_{prop} = 3.4328

V_{ney} = 2.8962

V_{ran} = 6.1879

Relative precision for Proportional allocation is 180.26

Relative precision for Neymann allocation is 213.6558%

Practical No: 25

-: Gain in precision due to stratification:-

The following data is derived for the stratified sample of tires dealers were assigned to strata according to the no. of new tires held at a previous senses. The sample mean are the mean no. new tires per dealer.

a. Estimate the gain in precision due to stratification.

b. Compare the result with gain that would have been attained from proportion allocation.

Stratum Boundaries	N_h	$\overline{Y}_h$	s_h^2	n_h
1 – 9	19850	4.1	34.8	3000
10 – 19	3250	13.0	92.2	600
20 – 29	1007	25.0	174.2	340
30 – 39	606	38.2	320.4	230
Total	24713			= 4170

Aim:-

To estimate the gain in precision due to stratification and also compare this result with the gain that would have been attained from proportional allocation.

Procedures:

$$\vartheta(\overline{Y}_{st}) = \sum_h \frac{w_h^2 s_h^2}{n_h} - \sum_h \frac{w_h^2 s_h^2}{N_h}$$

$$\vartheta_{prop}(\overline{Y}_{st}) = \frac{N-n}{Nn} \sum w_h s_h^2$$

$$\vartheta(\overline{Y}) = \frac{N-n}{n(N-1)} \left[\sum_h w_h s_h^2 + \sum_h w_h \overline{y}_h^2 - (\sum_h w_h \overline{y}_h)^2 \right]$$

Gain in precision due to stratification

$$\frac{1}{\vartheta(\overline{y}_{st})} - \frac{1}{\vartheta(\overline{y})}$$

Gain in precision due to proportional allocation

$$\frac{1}{\vartheta_{prop}(\overline{y}_{st})} - \frac{1}{\vartheta(\overline{y})}$$

S.No.	N_h	w_h	s_h^2	$w_h s_h^2$	$w_h^2 s_h^2$	n_h	$\frac{w_h^2 s_h^2}{n_h}$	$\frac{w_h^2 s_h^2}{N_h}$
1	19850	0.8032 w_h^2 =0.6451	34.5	27.9513	22.4494	3000	0.0074	0.0013
2	3250	0.1315 w_h^2 =0.0172	92.2	12.1243	1.5858	600	0.0026	0.0004
3	1007	0.0407 w_h^2 =0.0016	174.2	7.0899	0.2787	340	0.0008	0.0002
4	606	0.0245 w_h^2 =0.0006	320.4	7.8498	0.1922	230	0.0008	0.0003
Total	24713			55.0153	24.5061	4170	0.0116	0.0022

S. No	w_h	$\bar{y}_h$	$w_h \bar{y}_h$	$\bar{y}_h^2$	$w_h \bar{y}_h^2$
1	0.8032	4.1	3.29312	16.81	13.5017
2	0.1315	13.0	1.7095	169	22.2235
3	0.0407	25.0	1.0175	625	25.4375
4	0.0245	38.2	0.9359	1459.24	35.751
Total			6.956		96.9127

$$V(\bar{Y}_{st}) = \sum_h \frac{w_h^2 s_h^2}{n_h} - \sum_h \frac{w_h^2 s_h^2}{N_h}$$

$$= 0.0116 - 0.0022 \qquad = 0.0094$$

$$V_{prop}(\bar{Y}_{st}) = \frac{N - n}{Nn} \sum w_h s_h^2$$

$$= \frac{24713 - 4170}{4170(24713 - 1)} (55.0153)$$

$$= 0.01096$$

$$V(\bar{Y}) = \frac{N - n}{n(N - 1)} \left[\sum_h w_h s_h^2 + \sum_h w_h \bar{y}_h^2 - (\sum_h w_h \bar{y}_h)^2 \right]$$

$$= \frac{24713 - 4170}{4170(24713 - 1)} [55.0153 + 96.912 - (6.956)]$$

$$= 0.0206$$

Gain in precision due to stratification

$$\frac{1}{V(\bar{y}_{st})} - \frac{1}{V(\bar{y})} = \frac{1}{0.0094} - \frac{1}{0.0261}$$

$$= 106.3829 - 48.5436 \quad = 57.8393$$

Gain in precision due to proportional allocation

$$\frac{1}{V_{prop}(\bar{y}_{st})} - \frac{1}{V_{prop}}$$

$$= \frac{1}{0.0109} - \frac{1}{0.0206}$$

= 42.7913

Conclusion:-

$\vartheta(\bar{Y}_{st})$ = 0.0094

$\vartheta(\bar{Y}_{st})$ = 0.01096

$\vartheta(\bar{Y})$ = 0.02064

Gain in precision due to proportional allocation is 42.7

Gain in precision due to stratification is 57.8393

Gain in precision due to stratification is greater than the gain in precision due to proportional allocation.

Practical – 26

-: PPS Sampling:-

A sample survey was conducted to study the yield of wheat in Haryana. A sample of 20 farms from a total of 100 was taken, with probability proportional to the area under wheat crop with replacement method. The total area under wheat crop (x) was 484.5 hectares. The area under crop x and yield (y) were noted in hector and quintals per hector respectively. The sample selected by the cumulative total method was

Area under crop (x_i):	5.2	5.9	3.9	4.2	4.7	4.8	4.9	6.8
Yield of crop (y_i):	28	20	30	22	24	25	28	37
Area under crop(x_i):	4.7	5.7	5.2	5.2	4.9	4.0	1.3	7.4
Yield of crop (y_i):	26	32	25	38	31	16	06	61
Area under crop (x_i):	7.4	4.8	6.2	6.2				
Yield of crop (y_i):	61	29	47	47				

(i) Estimate the average yield per form using pps with replacement

(ii) Estimate the gain in precision due to pps sampling over simple random sampling with replacement.

$$\hat{\bar{y}}_{pps} = \frac{1}{nN}\sum_{i=1}^{n}(y_i/p_i)$$

$$\hat{y}_{pps} = N\ \hat{\bar{y}}_{pps}$$

$$V(\hat{\bar{y}}_{pps}) = \frac{1}{n(n-1)N^2}\left[\sum_{i=1}^{n}(y_i/p_i)^2 - N\,\hat{y}^2_{pps}\right] \qquad \text{-----} \qquad (1)$$

$$V(\hat{y}_{pps}) = N^2 V(\hat{\bar{y}}_{pps})$$

$$V(\hat{y}_{pps}) = \frac{1}{n^2}\left[\sum_{i=1}^{n}\frac{y_i}{p_i} - n\,\hat{y}^2_{pps}\right] + \frac{1}{n}V(\hat{y}_{pps})$$

$$\vartheta_{pps}(\hat{\bar{y}}_{sr}) = \frac{1}{N^2} V_{pps}(\hat{y}_{sr}) \qquad \text{------} \quad (2)$$

$$\text{Gain in precision} = \left(\frac{1}{V(\hat{\bar{y}}_{pps})} - \frac{1}{V_{pps}(\hat{\bar{y}}_{sr})} \right) \times 100$$

Calculation:-

$$\hat{\bar{y}}_{pps} = \frac{1}{nN} \sum_{i=1}^{n} (y_i/p_i)$$

$$= \frac{1}{20(100)} (58433.4871)$$

$$= 29.2167$$

$$\hat{y}_{pps} = N \hat{\bar{y}}_{pps}$$

$$= 100 \times 29.2167 = 2921.67$$

$$V(\hat{\bar{y}}_{pps}) = \frac{1}{n(n-1)N^2} \left[\sum_{i=1}^{n} (y_i/p_i)^2 - N \hat{y}^2_{pps} \right]$$

$$= \frac{1}{3800000} [7423459.5]$$

$$= 1.9535$$

y_i	x_i	$p_i = \frac{x_i}{484.5}$	y_i / p_i	$(y_i / p_i)^2$	y_i^2 / p_i
28	5.2	0.0107	2616.8224	6847759.473	73271
29	5.9	0.0122	2377.0492	5650362.899	68934.4
30	3.9	0.0080	3750	14062500	112500
22	4.2	0.0087	2528.7356	6394503.735	55632.
24	4.7	0.0097	2474.2268	6121798.258	59381.
25	4.8	0.0099	2525.2525	6376900.189	63131.
28	4.9	0.0101	2772.2772	7685520.874	77623
37	6.8	0.140	2642.8571	6984693.651	97785
26	4.7	0.0097	2680.4124	7184610.634	69690.
32	5.7	0.0118	2711.8644	7354208.524	86779
25	5.2	0.0107	2336.4486	5458992.06	58411.
38	5.2	0.0107	3551.4019	12612455.46	134953
31	4.9	0.0101	3069.3069	9420644.846	95148
16	4.0	0.0082	1927.7108	3716068.928	30843.
06	1.3	0.0027	2222.2222	4938271.506	13333
61	7.4	0.0153	3986.9281	15895595.67	243202
61	7.4	0.0153	3986.9281	15895595.67	243202
29	4.8	0.0099	2929.2929	8580756.894	84949
47	6.2	0.0128	3671.8754	13482666.02	172578
47	6.2	0.0128	3671.8754	13482666.02	17257
Total			=58433.4871	=178146571.3.	= 2013

$$V(\hat{y}_{pps}) = N^2 V(\hat{\bar{y}}_{pps})$$

$$= (100)^2 (1.9535)$$

$$= 19535$$

$$V(\hat{y}_{SR}) = \frac{1}{n^2}\left[\sum_{i=1}^{n}\frac{y_i}{p_i} - n\,\hat{y}_{pps}^2\right] + \frac{1}{n}V(\hat{y}_{pps})$$

$$= \frac{1}{(20)^2}[100(2013930.935) - 20(2921.67)^2] + \frac{1}{(20)}(19535)$$

$= 76674.95425 + 976.75$

$= 77651.70425$

$$V(\hat{\bar{y}}_{pps}) = \frac{1}{n^2} V_{pps}(\hat{y}_{RS})$$

$$= \frac{1}{(100)^2}[77651.70425]$$

$= 7.7652$

Gain in precision $= \dfrac{1}{V(\hat{\bar{y}}_{pps})} - \dfrac{1}{V_{pps}(\hat{\bar{y}}_{SR})} \times 100$

$$= \frac{1}{1.9535} - \frac{1}{7.7652} \times 100$$

$= 38.3122\%$

Conclusion:-

$\hat{\bar{y}}_{pps} = 29.2167$

$\hat{y}_{pps} = 2921.67$

$V(\hat{\bar{y}}_{pps}) = 1.9535$

$V(\hat{y}_{pps}) = 19535$

$V(\hat{y}_{RS}) = 77651.70425$

$\vartheta(\hat{\bar{y}}_{RS}) = 7.7652$

Gain in precision is 38.3122%

Practical – 27

-: Ratio of method of estimation :-

A sample of 34 villages was selected from a population of 170 villages in a region. The following table gives the data of cultivated area under wheat in 1963(y) and 1961(y) for these sample villages.

S. No.	x	y
1	70	50
2	163	149
3	320	284
4	440	381
5	250	278
6	125	111
7	558	634
8	254	278
9	101	112
10	359	355
11	109	99
12	481	498

S. No.	x	y
13	125	111
14	5	6
15	427	339
16	78	80
17	75	105
18	45	27
19	564	515
20	238	241
21	92	85
22	247	221
23	134	133
24	131	144

S. No.	x	y
25	129	103
26	190	175
27	363	335
28	235	219
29	73	62
30	62	79
31	71	60
32	137	100
33	196	141
34	255	263

i. Estimate the area under wheat in 1964 by method of ratio estimation using information on wheat area and x = 21288 acres for 1963.
ii. Determine the efficiency the ratio estimation as compared to the usual SRS estimate.

Aim:-

To estimate the area under wheat in 1964, by ratio estimate method by given data on wheat area x = 21288 acres for 1963 and to determine the efficiency the ratio estimation as compared to the usual estimate.

Procedure:-

$$\hat{y}_R = \frac{\bar{y}}{\bar{x}} x$$

where $\bar{y}$, $\bar{x}$ are sample means

$$v(\hat{y}_R) = \frac{N(N-n)}{n(n-1)} [\sum_{i=1}^{n} y_i^2 + \hat{R}^2 \sum_{i=1}^{n} x_i^2 - 2\hat{R} \sum_{i=1}^{n} x_i y_i] \quad \text{------} \quad (1)$$

where $\hat{R} = \frac{\bar{y}}{\bar{x}}$

$$v(\hat{y}) = \frac{N^2(N-n)}{n(n-1)} \frac{s^2 y}{n} \quad \text{------} \quad (2)$$

where s^2y =

Relative efficiency = $\frac{v(\hat{y})}{v(\hat{y}_R)}$ x 100

Calculation:-

S. No.	x_i	y_i	x_i^2	y_i^2	x_i y_i
1	70	50	4900	2500	3500
2	163	149	26569	22201	24287
3	320	284	102400	80656	90880
4	440	381	193600	145161	167640
5	250	278	62500	77284	69500
6	125	111	15625	12321	13875
7	558	634	311364	401956	353772
8	254	278	64516	77284	70612
9	101	112	10201	12544	11312
10	359	355	128881	126025	127445
11	109	99	11881	9801	10791
12	481	498	231361	248004	239538
13	125	111	15625	12321	13895
14	5	6	25	36	30
15	427	339	182239	114921	144753
16	78	80	6084	6400	6240
17	75	105	5625	11025	7875
18	45	27	2025	729	1215
19	564	515	318096	265225	290460
20	238	241	54756	58081	56394
21	92	85	8464	7225	7820
22	247	221	61009	48841	54587
23	134	133	17956	17689	17822

24	131	144	17161	20736	18864
25	129	103	16641	10609	13287
26	190	175	36100	30625	33250
27	363	335	131769	112225	121605
28	235	219	55225	47961	51465
29	73	62	5329	3844	4526
30	62	79	3844	6241	4898
31	71	60	5041	3600	4260
32	137	100	18769	10000	13700
33	196	141	38416	19881	27636
34	255	263	65025	69169	67065

n = 34

$$\bar{x} = \frac{\sum x_i}{n} = \frac{7102}{34} = 208.8824$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{6773}{34} = 199.2059$$

$$\hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{199.2059}{208.8824} = 0.9537$$

$$\hat{y}_R = \frac{\bar{y}}{\bar{x}} x = = 20301.8311$$

$$s^2y = \frac{\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2}{n-1}$$

$$= \frac{2093121 - 34(199.2059)^2}{33}$$

= 22542.4036

$$V(\hat{y}_R) = \frac{N(N-n)}{n(n-1)} [\sum y_i^2 + \hat{R}^2 \sum x_i^2 - 2\hat{R} \sum x_i\, y_i\,]$$

$$= \frac{170(170-34)}{34(33)} [2093121 + (0.9537)^2 (2231000) - 2(0.9537)(2145743)]$$

= 608349.2333

$$V(\hat{y}) = \frac{N(N-n)}{n} s^2y$$

$$= \frac{170(170-34)}{34} . 22542.4036$$

= 15328834.44

$$\text{Relative efficiency} = \frac{v(\hat{y})}{v(\hat{y}_R)} \times 100$$

$$= \frac{1532884.44}{608349.23} \times 100$$

$= 2520\%$

Conclusion:-

$\therefore \hat{y}_R = 20301.8311$

$\vartheta(\hat{y}) = 15328834.44$

$\vartheta(\hat{y}_R) = 608349.23333$

Relative efficiency = 2520%

Practical – 28

-: Regression method of estimation:-

An experienced former makes an eye estimate of the weight of apples on each tree in an orchard of N = 20 trees. He finds a total weight of x = 11600 lbs. The apples of picked and weighted on a SRS of 10 trees with the following results.

Tree number	Actual wt (y_i)	Estimated wt (x_i)
1	61	59
2	42	47
3	50	52
4	58	60
5	67	67
6	45	48
7	39	44
8	57	58
9	71	76
10	53	58

Compute the regression estimate for total actual weight of y and its standard error.

Aim:-

To compute the regression estimate for total actual weight y and to find its standard error.

Procedure:-

Regression estimate of population mean is

$$\hat{\bar{y}}_{/r} = \bar{y}_{/r} = \bar{y} + b(\bar{x} - \bar{x})$$

$$\hat{\bar{y}} = N\hat{\bar{y}}_{/r} = N\bar{y}_{/r} \quad \text{-----} \quad (1)$$

$$V(\bar{y}_{/r}) = \frac{1-f}{n(n-2)} \left[\sum_{i=1}^{n} (y_i - y) - b(x_i - \bar{x})\right]^2$$

where, $b = \dfrac{\sum_{i=1}^{n} (y_i - y)(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

$$V(\bar{y}_{/r}) = \frac{N-n}{Nn(n-2)} \sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{\left\{\sum_{i=1}^{n}(y_i - y)(x_i - \bar{x})\right\}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$V(\hat{y}_{/r}) = N^2 V(\bar{y}_{/r})$$

Standard error of $\hat{y}_{/r} = \sqrt{V(\hat{y}_{/r})}$ ------ (2)

Calculation:-

Tree number	Actual wt (y_i)	Estimated wt (x_i)	y_i^2	x_i^2	x_i y_i
1	61	59	3721	3481	3599
2	42	47	1764	2209	1974
3	50	52	2500	2704	2600
4	58	60	3364	3600	3480
5	67	67	4489	4489	4489
6	45	48	2025	2304	2160
7	39	44	1521	1936	1716
8	57	58	3249	3364	3306
9	71	76	5041	5776	5396
10	53	58	2809	3364	3074
	= 543	= 569	= 30483	= 33227	= 31794

$$\bar{y} = \frac{\sum y_i}{n} = \frac{543}{10} = 54.3$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{543}{10} = 56.9$$

$$\bar{X} = \frac{11600}{200} = 58$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

$= 30483 - 10(54.3)^2$

$= 998.1$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i - n\bar{x}^2$$

$= 33227 - 10(56.9)^2$

$= 850.9$

$$\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$= 31794 - 10(54.3)(56.9)$

$= 897.3$

$$b = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} =$$

$$= \frac{897.3}{850.9}$$

$= 1.0545$

$\hat{\bar{y}}_{/r} = \bar{y}_{/r} = \bar{y} + b(\bar{X} - \bar{x})$

$= 54.3 + 1.0545(58 – 56.9)$

$= 55.46$

$\hat{y}_{/r} = N\,\bar{y}_{/r}$

$= N\,\bar{Y}_{/r}$

$= 200 \times 55.46 \quad = 11092$

$$V(\bar{Y}_{/r}) = \frac{(N - n)}{Nn(n - 2)}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{\left\{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})\right\}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

$$= \frac{(200 - 10)}{200\,(10)\,(8)}\left[(998.1) - \frac{(897.3)^2}{(850.9)}\right]$$

$= 0.6160$

$\vartheta(\hat{y}_{/r}) = (200)2\ (6160)$

$= 24638$

Standard error of $\hat{y}_{/r} = \sqrt{V(\hat{y}_{/r})}$

$= \sqrt{23340} \quad = 156.9655$

Conclusion:-

$\hat{y}_{/r} = 11092$; $\hat{\bar{y}}_{/r} = 55.46$

$\vartheta(\hat{y}_{/r}) = 24638$, $\vartheta(\bar{y}_{/r}) = 0.6160$

Standard error of $\hat{y}_{/r} = 156.9655$