Date :

Page No.

Probability and Statistics - 2
Unit 3 & 4
:- Assignment Questions:

Ques 1:

(a) $\bar{x} = 85$

$\bar{y} = 173.7$

$$S_{xx} = \Sigma x_i^2 - n\bar{x}^2$$
$$= 92500 - 10(85)^2$$
$$\boxed{S_{xx} = 20250}$$

$$S_{yy} = \Sigma y_i^2 - n\bar{y}^2$$
$$= 372717 - 10(173.7)^2$$
$$\boxed{S_{yy} = 71000.1}$$

$$S_{xy} = \Sigma xy - n\bar{x}\bar{y}$$
$$= 182560 - \cancel{10} \times \frac{173.7}{\cancel{10}} \times 85$$
$$\boxed{S_{xy} = 34915}$$

$$\beta = \frac{S_{xy}}{S_{xx}} = \frac{34915}{20250}$$
$$\boxed{\beta = 1.7242}$$

$$\alpha = \bar{y} - \beta\bar{x}$$
$$= 173.7 - 1.7242 \times 85$$
$$\boxed{\alpha = 27.143}$$

$$y = \alpha + \beta x$$
$$\boxed{y = 27.143 + 1.7242x}$$

Date :

Page No.

(b) Gradient of regression line $= r\left(\frac{S_y}{S_x}\right)$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{34915}{\sqrt{20280 \times 71000.1}}$$

$$\boxed{r = 0.003456}$$

$$= 0.003456 \left(\frac{88.81948}{47.43416}\right)$$

$$\boxed{\text{Gradient of regression line} = 0.006413}$$

Ques 2:

$$= \frac{S_{xy}}{S_{xx}}$$

~~$\bar{x} = 32.1$~~

~~$\bar{y} = 96.875$~~

(1)

~~$S_{xy} = \Sigma xy - n\bar{x}\bar{y}$~~

$$\bar{x} = \frac{\Sigma x}{n} \qquad \bar{y} = \frac{\Sigma y}{n}$$

$$= \frac{385.2}{12} \qquad \bar{y} = \frac{1162.5}{12}$$

$$\boxed{\bar{x} = 32.1} \qquad \boxed{\bar{y} = 96.875}$$

$$S_{xy} = \Sigma xy - n\bar{x}\bar{y}$$

$$= 38{,}191.41 - 12 \times 32.1 \times 96.875$$

$$\boxed{S_{xy} = 875.16}$$

$S_{xx} = \Sigma x_i^2 - n\bar{x}^2 =$ ~~12666.58 - 12x(...)~~^{2}

$= 12666.58 - 12 \times (32.1)^2$

$\boxed{S_{xx} = 301.66}$

$\beta = \frac{S_{xy}}{S_{xx}} = \frac{875.16}{301.66}$

$\boxed{\beta = 2.90115}$

$\alpha = \bar{y} - \beta\bar{x}$

$\alpha = 96.875 - 2.90115 \times 32.1$

$\boxed{\alpha = 3.748}$

Equation : $y = \alpha + \beta x$

$\boxed{y = 3.748 + 2.90115x}$

(2). Slope coefficient of the model $= \beta = 2.90115$

$95\%\ C.I = (\bar{x} - \bar{y}) \pm t_{10} \sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$S_p^2 = (n_1-1)S_1^2 + (n_2-1)S_2^2 / n_1+n_2-2$

$= \frac{9 \times 27.424 + 9 \times}{18}$

$= \frac{27.424 + 583.6166}{2} = 305.5203$

Date :
Page No.

$$C.I = (96.875 - 32.1) \pm t_{10;\,2.5\%} \sqrt{305.5203\left(\frac{1}{12}+\frac{1}{12}\right)}$$

$$= 64.775 \pm 2.228 \sqrt{\frac{305.5203}{6}}$$

$$C.I = (48.8764,\ 80.674.)$$

(iii) Mean IBI = μ_0

$$Se(\hat{\mu}_0) = \sqrt{\left(\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)\hat{\sigma}^2}$$

$$= 10.59844.$$

$$95\%\ C.I = \mu_0 \pm t_{10;\,2.5\%}\, Se(\hat{\mu}_0)$$

$$= 115.5405 \pm 23.61443$$

$$= (36.17966,\ 143.4852)$$

Que 3 (i) Assumption
Plot must be normal.
Population have common values.
The observations are independent.

(ii) H_0: Mean is same
H_1: Mean is not same

$$SS_{TOT} = S_{yy} = \Sigma y^2 - n\bar{y}^2 = 1495 - \frac{(103)^2}{28}$$

Date:
Page No.

$SS_{TOT} = 1116.11$

$$SS_{REG} = \frac{1}{7}\left(64^2 + 44^2 + 8^2 + 4^2\right) - \frac{100^2}{28}$$

$= 516.11$

$SS_{RES} = SS_{TOT} - SS_{REG}$

$\boxed{SS_{RES} = 600}$

(iii) Anova

Under H_0 $F = \frac{MSS_{REG}}{MSS_{RES}} = \frac{172.04}{25} = 6.88$

$T.S = 6.88$

$F_{3,24,5\%} = 3.009.$

Hence, we have sufficient evidence to reject H_0 at 5% LOS

(iv) $\bar{y}_1 = 9.14$

$\bar{y}_2 = 6.298$

$\bar{y}_3 = 1.14$

$\bar{y}_4 = -1.86$

$\bar{y}_1 > \bar{y}_2 > \bar{y}_3 > \bar{y}_4.$

$\hat{\sigma}^2 = \frac{SS_{RES}}{n-4} = 25$

Least significance diff. b/w. of means is

$t_{24, 0.025}\, \hat{\sigma}\sqrt{\frac{1}{7}+\frac{1}{7}} = 5.52.$

Date :
Page No.

$\overline{Y}_{.4.} - \overline{Y}_{.1} = 2.85$

$\overline{Y}_{.} - \overline{Y}_{3} = 5.15$

$Y_{31} - \overline{Y}_{3} = 3$

$\overline{Y}_{1} > \overline{Y}_{2} > \overline{Y}_{3.} > \overline{Y}_{7}$

same trend.

Ques 4.

(i) $y_{ij} = \mu + \tau_i + e_{ij}$ $\quad i = 1, 2, 3, \ldots k$
$\quad j = 1, 2, \ldots k_i$

K = no. of treatments
k_i = no. of responses
τ_i = Responses from I treatment.

Assumption e_{ij} is iid $N(0, \sigma^2)$.

(ii) $n_1 = 7$
$n_2 = 8$
$n_3 = 6$
$n_4 = 7$
$n_5 = 5$
$n = 33$

H_0: Mean claim amount is equal.
H_1: Mean claim amount is not equal.

$$\sum_{i=1}^{5} \sum_j y_{ij} = 1856$$

$$\sum_{i=1} \sum_{j=1} y_{ij}^2 = 174310$$

$$CF = \left(\sum \sum y_{ij}\right)^2 / n = 104385.94.$$

Date :
Page No.

$SS_{TOT} = 69930$, $SS_{RES} = 22\,7325.0$ $SS_{REG} = 64604.37$

$T.S = \frac{MSS_{REG}}{MSS_{M}}$ 3.283

$F_{1,4,0.05} = 2.714$

we have sufficient evidence to reject H_0.

O_i

(iii)

	3-5	5-8	8-10	10-12	
0-3	47	20	6	6	76
4-6	7	20	3	5	42
7-9	5	8	15	12	40
	54	48	30	23.	188

E_i	3-5	5-8	8-10	10-12
0-3	27.41	23.08	14.43	11.07
4-6	15.15	12.75	7.57	6.11
7-9	14.43	12.15	7.54	5.88

Contingency table.

H_0: Dalus are independent w.r.t. of oriraind.

H_0 " " dependent " "

$$T_S = \frac{\Sigma (O_i - E_i)^2}{E_i}$$

$T.S = 49.91148$.

$DOF = (r-1)(c-1) = 2 \times 3 = 6$

$\chi^2_{0.1,} = 12.59$

Date:
Page No.

Hence we have sufficient evidence to reject H_0
i.e. factors are dependent on sex of actual papers.

Q.5 (i) The sample variances are very different from each other.

(ii) Mean at $n = \frac{\sqrt{24} + \sqrt{13} + \sqrt{92}}{3} = 4.50443$

Var at n = 2 $= \frac{\sum(x_i - \bar{x})^2}{n} = 0.12124$

(iii) $Y_{ij} = \mu + \tau_i + e_{ij}$

For Anova

$H_0 \; \tau_i = 0$
$H_1 \; \tau_i \neq 0$

Rate	Mean	Variance	y_i^2
1	2.992	0.163	80.582
2	4.827	0.121	209.678
3	5.716	0.186	294.053
4	6.577	0.2	389.01
Total	20.110	0.618	973.373

$y_i^2 = \left(\sum_{i=1}^{3} y_{ij}\right)^2 = \sum m_i y^2)$

$SS_{\text{rest}} = \sum \frac{y_{i\cdot}^2}{3} - y_{\cdot\cdot}^2 = \frac{973.373}{3} - \left(\frac{2x\sim}{1\sim}\right)$

$\boxed{SS_{xc} = 21.153}$

Source of Variance	Dif	SS	MSS
Rate	3	21.153	50
Residual	8	1.226	0.15
Total	11	22.33	

$T.S = 45.638$

$F_{3,5,1\%} = 7.951$

Q6

x	5	9	2	3	4	1	6	5	4
y	13	120	34	46	24	26	93	25	66

$n = 10$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{661.2}{\sqrt{54.9 \times 8923.6}}$$

$$\boxed{r = 0.94466}$$

Date :
Page No.

fitted regression eqⁿ

$\hat{\alpha} = \hat{y} - \hat{\beta}\bar{x}$

$\boxed{\alpha = 9.22415}$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 12.04337$

$\boxed{\hat{y} = 9.2245 + 12.04334x}$

$SS_{TOT} = S_{yy} = 8923.6$

$SS_{REG} = \frac{S^2_{xy}}{S_{xx}} = 7463.30492$

$SS_{RES} = 960.24806$

$\boxed{R^2 = 89.23807}$ $\left(\frac{S^2_{xy}}{S_{xx}S_{yy}}\right)$

Model is a good fit as it contains almost 89% variability in the output.

Q4

$S_{xy} = 4785$

$S_{yy} = 2176.424$

$S_{xx} = 1189909$

$\boxed{\hat{\beta} = 0.45454}$

$\hat{\alpha} = \hat{y} - \hat{\beta}\bar{x}$

$\boxed{\hat{\alpha} = 10.73086}$

Date :
Page No.

$$\boxed{\hat{y} = 10.78086 + 0.45457x}$$

(ii) $H_0 : \beta = 0$
$H_1 : \beta \neq 0$

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \sim t_{n-2}$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S^2_{xy}}{S_{xx}}\right) = 22.20136.$$

$$T.S = \frac{\hat{\beta} - 0}{\sqrt{\hat{\sigma}^2/S_{xx}}} = 6.67568$$

$\therefore$ we have insufficient evidence to reject H_0

Assumption :-
Population have common variance.
Population follows normal distribution.

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = 0.91213.$$

(iv) For Mean response.
$x_0 = 25$
$\hat{y}_0 = \hat{\alpha} + \hat{\beta} x_0 = 22.1534$

$$S_y(\hat{y}_0) = \sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)\hat{\sigma}^2} = 1.85781$$

Date :

Page No.

$\frac{\hat{\mu}_0 - \mu_0}{se(\hat{\mu}_0)}$ is t_9.

95% C.I for μ_0

$= \mu_0 \pm t_{9,2.5\%} \cdot S.E.$

$= (18.84313, 25.66340)$

Individual response

$= (10.932, 33.37461)$

(v) There is clearly some relationship bw. the residuals and the explanatory variable values. Hence our assumption of normality might be true.

Qu 8: (i)

Principal component	Diagonal Entries	PC %
PC_1	0.436	65%
PC_2	0.137	19.5%
PC_3	0.08	11.4%
PC_4	0.0165	2.4%
PC_5	0.012	1.7%
		100%

(ii) As the 3rd principle component explain 95% of total variance which can be further used for building classified regression models.

Date :
Page No.

A: (i) $S_{xx} = 75$
$S_{yy} = 2482.93$
$S_{xy} = -296$

$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$

$\boxed{r = -0.686}$

Hence my have -ve correlation coefficient

(ii) H_0 $\tau = 0$
H_1 $\tau < 1$
$\tan^{-1}(r) = \tanh^{-1}(\rho) \sim N\left(0, \frac{1}{n-3}\right)$

$Z_{cal} = -2.07893$
$Z_{tab} = 1.6449$
$Z_{cal} > Z_{tab}$
Hence we have sufficient evidence to reject H_0

(iii) $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = -3.94667$ $\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$
$= 82.10$

$\boxed{\hat{y} = 82.10 \mp 3.94467 \hat{x}}$

(iv) $R^2 = 47.05987$

Hence model explains 47% variability
∴ Model is not a good fit.

(v) $x = 1$
$\hat{y} = \hat{\alpha} - \hat{\beta} \hat{x}_0$ $\boxed{\hat{y}_0 = 78.21873}$

Date:
Page No.

Ques 10:

x	0.29	0.36	0.3
y	0.05	0.1	0.08

$S_{xy} = 0.0042$

$S_{yy} = 0.00120$

$S_{xx} = 0.0022$

$SS_{TOT} = 0.00126$

$SS_{reg} = 0.00115$

$SS_{res} = 0.00011$

Anova Table

Reg	1	0.00115
Res.	1	0.00011
Tot	2	0.00126

$H_0 \Rightarrow B = 0$

$H_1 \Rightarrow B \neq 0$

$T.S = \frac{MSS_{reg}}{SS_r} = 10.45$

$F_{tab} = 161.4$

Hence we have insufficient evidence to reject H_0.