ASSIGNMENT 2

Gargi Tendulkar
Roll No. 77.

Page :
Date : / /

Q1 (a) $\Sigma xy = 182560.$
$\Sigma x = 850$
$\Sigma y = 17.37.$
$\Sigma x^2 = 92500.$

$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 34.915.$$

$$S_{xx} = 20.250.$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 1.724198.$$

$$\bar{\alpha} = \bar{y} - \hat{\beta}\,\bar{x} = 27.14317.$$

$$\therefore \boxed{\hat{y} = 27.14317 + 1.72419x.}$$

(b) The slope of regression line refers to its slope parameter β.
Since β is significantly different from 0, there exist a positive linear relationship between x.

Page :
Date : / /

Q2 (i) Let x = share price of Dell
y = share price of IBM.

~~then~~, and

$S_{xy} = 875.16$

$S_{xx} = 301.66$

$S_{yy} = \cancel{640.9}\ 6409.7125$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 2.90115$$

$$\hat{\alpha} = \bar{y} + \hat{\beta}\bar{x} = 3.74809$$

$$\boxed{\hat{y} = 3.74809 + 2.90115x}$$

(ii) Assumptions:

$$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t_{n-2} \text{ with LOS } = 5\%$$

$$\therefore 95\% \text{ CI of } \beta = \hat{\beta} \pm t_{10,\,2.5\%}\sqrt{\frac{\hat{\sigma}}{S_{xx}}}$$

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

$$= \frac{1}{10}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = 387.07447$$

$$\therefore 95\% \text{ CI of } \beta = \hat{\beta} \pm t_{10,\,2.5\%}\sqrt{\frac{387.07447}{301.66}}$$

$\therefore$ 95% CI of $\beta = 2.90115 \pm 2.52379.$

$\therefore \boxed{95\% \text{ CI of } \beta = (0.37736, 5.424)}$

(iii) Mean of y: 113m $= \mu_0$.

$$\therefore \frac{\hat{\mu}_0 - \mu_0}{se(\mu_0)} \sim t_{n-2}$$

$\therefore \boxed{x_0 = 40}$

$$se(\mu_0) = \sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)\hat{\sigma}^2} = 10.54894.$$

$$\therefore \mu_0 = \hat{\alpha} + \hat{\beta} x_0 = 119.78409.$$

95% CI for μ_0

$$\hat{\mu}_0 \pm t_{10, 2.5\%} \, se(\hat{\mu}_0)$$

$$= 119.78409 \pm 23.61443.$$

$\therefore \boxed{95\% \text{ of } \mu_0 = (96.17866, 143.40852)}$

Q3 (i) City A hase the highest ~~mean~~ difference between the mean time taken for commute while city D has the lowest.

~~(ii)~~. All the 4 cities may have the underlying measures ~~of~~ of distribution to be different since the centres of ~~diff~~ distribution are different.

(ii) Assumptions underlying analysis of variance:-
Populations must follow Normal distribution
Populations must have a common variance.
Observations are independent identically distributed variables

(iii) H0 : mean of differences is same for all ~~cities~~ cities
H1 : mean of differences is not same for all cities.

$$SS_{Total} = S_{yy} = 1116.11$$

$$SS_{reg} = \frac{1}{7}(64^2 + 64^2 + 8^2 + 1.13)^2 - \frac{103^2}{8}$$

$$= 516.11$$

$$SS_{res} = SS_T - SS_{REG}$$
$$= 1116.11 - 516.11$$
$$= \underline{\underline{600}}$$

Anova table:

Q4. (a). Mathematical model for 1 way ANOVA

$$y_{ij} = \mu + \tau_i + e_{ij}$$

where, $i = 1, 2, 3 \ldots t$
$j = 1, 2, 3 \ldots n$

Where,
t = no. of treatments.
μ = mean response
τ_i = ith treatment.
e_{ij} = error term.
n = no. of responses from ith treatment

(b) H_0 = mean claim amounts of 5 companies are equal
H_1 = mean claim amts. of 5 companies are not equal.
$n_1 = 7$ $n_2 = 8$ $n_3 = 6$ $n_4 = 7$ $n_5 = 5$ $n = 33$

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

$$\sum_{i=1}^{5} \sum_{j=1}^{n_i} y^2_{ij} = 174316$$

$$CF = \frac{\left(\sum_{i=1}^{5} \sum_{j=1}^{n_i} y_{ij}\right)^2}{n} = 104385.94$$

$SS_{TOT} = 174316 - CF = 69,930.06$

$$SS_{reg} = \frac{354^2}{7} + \frac{386^2}{8} + \frac{872}{6}^2 + \frac{645^2}{7} + \frac{384^2}{5}$$

$SS_{reg} = \ldots - 104385.84$
$= 22325.69$

$\therefore SS_{res} = 47604.$

(c)

Salary.

paper cleared	3-5	5-8	8-10	10-12	Total
0-3	45	20	6	5	76
4-6	7	20	9	6	42
7-9	5	8	15	12	48
	57	48	30	23.	158

Ei

Papers clared	3-5	5-8	8-10	10-12
0-3	27.417	23.088	14.43	11.063
4-6	15.1518	12.75	7.47	6.11
7-9.	14.430	12.151	7.59	5.80

Contingency table:

Ho = Salary independent to papers

H1 = Salary not independent to papers.

One sided test : χ^2 test.

$$TS = \frac{(\Sigma Oi - Ei)^2}{Ei} = 49.9446.$$

$$\chi^2_{6, 5\%} = 12.59$$

Reject Ho.

Salary dependent on papers.

Q6. $S_{xy} = 661.2$
$S_{xx} = 54.9$
$S_{yy} = 8923.6$.

(i) $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = 0.94466$.

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 12.04371$

(ii) fitted regression eqn.

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 9.2295$.

$\therefore$ $\boxed{y = 9.2295 + 12.04371x}$

(iii) $SS_{Total} = S_{yy} = 8923.6$.
$SS_{reg} = \frac{S^2_{xy}}{S_{xx}} = 7963.304$

$SS_{res} = SS_{Tot} - SS_{reg} = 960.29$

$R^2 = \frac{S^2_{xy}}{S_{xx} S_{yy}} = \boxed{89.238703.}$

Q7. (i) $S_{xx} = 4788.4221$

$S_{xy} = 2176.84$

$S_{yy} = 1189.20767$

$$\hat{\beta} = \cancel{\frac{S_{xx}}{S_{xy}}} \quad \frac{S_{xy}}{S_{xx}} = 0.45451$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 10.78056$$

$$\boxed{\hat{y} = 10.78056 + 0.45451x}$$

(ii) $H_0 : \beta = 0$

$H_1 : \beta \neq 0$

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

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

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

$\therefore t_{9, 2.5\%} = 2.262.$

~~t~~ $TS > t$

Reject H_0.

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

(iv) Mean Response $= \mu_0.$

$$\therefore se(\hat{\mu}_0) = \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}}\ \hat{\sigma} = \underline{1.55181}$$

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

$$95\%\ CI\ \text{of}\ \mu_0 = \hat{\mu}_0 \pm t_{9,\ 2.5\%}\ se(\hat{\mu}_0)$$

$$= (22.1533 \pm 3.51018)$$

$$= \boxed{(18.64313,\ 25.66349)}$$

Individual response (y_0).

$$\hat{y}_0 = \hat{\alpha} + \hat{\beta} x_0 = 22.153.$$

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

$$= \cancel{4.90}\quad 4.9607$$

$$\frac{\hat{\mu}_0 - \mu_0}{se(\mu_0)} \sim t_{n-2}$$

$$95\%\ CI\ \text{for}\ \mu_0 = \hat{\mu}_0 \pm t_{9,\ 2.5\%}\ se(\hat{\mu}_0)$$

$$= 22.15331 \pm 11.22131$$

$$= \boxed{(10.932,\ 33.37462)}$$

Page :
Date : / /

Q9 (i) Based on the scatterplot, there appears to be a negative correlation between the marks and the hours spent.

(ii) $S_{xx} = 75$
$S_{yy} = 2482.4$
$S_{xy} = -294$

$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = -0.686002$

$\because$ -ve
$\therefore$ they have a weak correlation

(iii) $H_0 : \rho = 0$
$H_1 : \rho < 0$

$Z_{cal} = -2.07893$
$Z_{5\%} = 1.6449$

$\because |Z_{cal}| > Z_{tab}$
$\therefore$

Reject H_0
$\therefore \rho < 0$.

(iv) $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = -3.94667$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 82.16$

$\boxed{\hat{y} = 82.16 - 3.94667x}$

(v) $R^2 = \frac{S_{xy}}{S_{xx} S_{yy}} = \boxed{47.0598\%}$

(vi) $x_0 = 1$

$$\hat{\mu}_0 = \hat{\alpha} - \hat{\beta}_0 x_0$$

$$= \boxed{78.21333}$$

Q10 S_{xy}

$S_{xx} = 0.0042$

$S_{yy} = 0.00126$

$S_{xy} = 0.0022$

$SS_{Tot} = S_{yy} = 0.00126$

$SS_{reg} = \dfrac{S^2_{xy}}{S_{xx}} = 0.00115$

$SS_{res} = 0.0011$

ANOVA TABLE: