Date : \ \ \

P & S Assignment

Roll - 80

Q1) $\Sigma x_i^2 = 92500$ $\Sigma x = 850$

$\bar{x} = 85$

$\Sigma xy = 182500$ $\Sigma y = 1737$

$\overline{x}\,\overline{y} = 14764.5$

$S_{xx} = \Sigma x_i^2 - n\bar{x}^2$

$= 20250$

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

$= 34915$

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

$= 1.724$

Date : \ \ \

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

$= 27.1432$

$\therefore$ Equation is,

$y = 27.1432 + 1.724x_i$

Q2) i) $S_{xx} = \sum x_i^2 - n\bar{x}^2$

$= 301.66$

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

$= 875.16$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ $\qquad S_{yy} = \sum y^2 - ny^2$

$= 1409.712$

$= \cancel{2.901174}$ 2.901147

$\alpha = 3.748182$

$\therefore$ equation,

$y = 3.748182 + 2.901147x_i$

ii) 95% CI

$\hat{\beta} = t_{0.0025,10}\sqrt{\frac{\sigma^2}{S_{xx}}}$

Date : \ \ \

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

$$= 6155.816$$

95% CI is $(-7.16357, 12.965)$

4) b) $SS_{TOT} = \sum y_{ij}^2 - \frac{y^2}{n}$

$$= 69930.06$$

$$SS_{Reg} = \sum \frac{y_i^2}{n_i} - \frac{y^2}{n_i}$$

$$= 22325.69$$

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

$$= 47604.37$$

Date : \ \ \

ANOVA TABLE

Source	df	SS	MSS	F_1
Regression	4	22325.64	5581.42	3.28
Residual	28	47604.37	1700.156	
Total	32	69930.06		

F_2 3.28 $F_{4,28} = 2.714$ at 5%

$\therefore$ We Reject H_0

c) Contingency Table

Salary

Papers	45	20	6	5
C.board	7	20	9	6
	5	8	15	20

$$\text{Expected} = \frac{\text{row total} \times \text{column total}}{\text{table total}}$$

Date : \ \ \

$\therefore$ Expected

27.4	23.8	14.81	11.06
15.4	12.7	8.19	6.09
14.3	12.6	7.8	5.08

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= 49.919$$

$$\chi^2_{0.025,7} = 17.53$$

Thus, we reject H_0

$\therefore$ No. of papers cleared are dependent on salary

Date : \ \ \

3) i) The variance of cities are mostly same.

ii) Assumptions
Population must be normal
Must have common variance
Observations are independent

iii) H_0 : Mean is same vs H_1 : Mean is not same

$$SS_T = 1495 - \frac{103^2}{28} = 1116.11$$

$$SSREG = \frac{1}{7}(64^2 + 44^2 + 8^2 + (-13)^2) - \frac{103^2}{28}$$

$$= 516.11$$

$$SS_{RES} = SSTOT - SSREG$$

$$= 600$$

Date : \ \ \

ANOVA Table

Source	df	SS	MS	F
Regression	3	516.11	172.04	6.88
Residual	24	600	25	
Total	27	1116.11		

$F = 6.88$ ~~$F_{3,24} =$~~

$F_{3,24} = 3.009$, at 5%

$\therefore$ Reject H_0

Date : \ \ \

Q5)

Source	df	SS	MSS	F
Regression	3	21.153	7.051	45.638
Residual	8	1.236	0.155	
Total	11	22.38		

$F_{3,8,\ 5\%} = 7.95\%$

$\therefore$ We reject H_0

Q6) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

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

$= 2853 - 10 \times 3.9 \times 56.2$

$= 661.2$

$S_{xx} = \Sigma x_i^2 - n\bar{x}^2$

$= 207 - 10(3.9)^2$

$= 54.9$

$S_{yy} = \Sigma y_i^2 - n\bar{y}^2$

$= 8923.6$

Date : \ \ \

i) $r = 0.944662$

ii) $\beta = \frac{S_{xy}}{S_{xx}} = 12.04372$

$\alpha = \bar{y} - \beta\bar{x}$
$= 9.229508$

$y = 9.229508 + 12.043724 x_i$

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

$SS_{REG} = \frac{S^2_{xy}}{S_{xx}}$
$= 7963.305$

$SS_{RES} = SS_{TOT} - SS_{REG}$
$= 960.295$

$R^2 = \frac{SS_{REG}}{SS_{TOT}}$

$= 0.892387$

1) $S_{xy} = 2176.84$

$S_{xx} = \Sigma x_i^2 - n\bar{x}^2$

$= 4789.422$

$S_{yy} = \Sigma y_i^2 - n\bar{y}^2$

$= 1189.208$

i) $\beta = \frac{S_{xy}}{S_{xx}} = 0.45451$

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

$= 10.79056$

$\therefore$ Equation,

$y = 10.79056 + 0.45451x_i$

ii) $\sigma^2 = \frac{11}{11-2}\left(S_{yy} - \frac{S^2_{xy}}{S_{xx}}\right)$

$= 22.201$

$H_0 : \beta \neq 0$ vs $H_1 : \beta \neq 0$

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

Date : \ \ \

$$\cancel{\frac{0.4545}{\sqrt{\frac{22.201}{4\ldots}}}}$$

$$\frac{0.4545}{\sqrt{\frac{22.201}{4789.422}}}$$

$= 6.675675$

$t_{9, 0.025} = 2.262$

$\therefore$ We reject H_0

iii) $r = \cancel{\frac{S^2_{xy}}{\sqrt{}}}$ $\frac{S^2_{xy}}{\sqrt{S_{xx} S_{yy}}}$

$= 0.912129$

v) The residual plot shows that they might form a non-linear relationship.

Date : \ \

Q8)

i) The main purpose of factor analysis is to reduce many individual item into fewer no. of dimension.

Principal component Analysis is a method to decrease dimension without any data loss

ii)

PC	Diagonal Entry (PC)	PCi / Sum
PC1	0.456	65%
PC2	0.137	18.5%
PC3	0.08	11.4%
PC4	0.0165	2.4%
PC5	0.043	1.7%
	0.1015	100%

iii) As 1st principle component explain that over 95% of variance can be reduced to 3 for this dataset

Date : \ \

Q9) i) Negatively Correlated

ii) $S_{xx} = 75$

$S_{yy} = 2482.4$

$S_{xy} = -296$

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

$= 0.686002$

$\therefore$ ~~(ve) corre~~ $-$ve correlation

iii) $H_0 : \rho = 0$

vs

$H_1 : \rho < 0$

$\tanh^{-1}(r) - \tanh^{-1}(\rho) \sim N\left(0, \frac{1}{n-3}\right)$

$$\frac{\tanh^{-1}(r) - \tanh^{-1}(\rho)}{\sqrt{\frac{1}{n-3}}}$$

$Z\ cal = -2.07893$

25% $= 1.6449$

$\therefore$ we reject H_0

$\rho < 0$

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

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

$= 82.16$

$\hat{y} = 82.16 - 3.94667x_i$

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

Not a good fit

vii) $x_0 = 1$

$y_0 = \hat{\alpha} - \hat{\beta}x_0$

$= 78.2133$

Date : \ \ \

10) $S_{xx} = 0.0042$

$S_{yy} = 0.00126$

$S_{xy} = 0.0022$

$SS_{TOT} = 0.00126$

$SS_{REG} = \dfrac{S^2_{xy}}{S_{xx}} = 0.0015$

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

$= 0.0011$

ANOVA TABLE

Source	df	SS	MSS	F
Regression	1	0.00115	0.00115	10.454
Residual	1	0.00011	0.00011	
Total	2	0.00126		

Date : \ \ \

$H_0 : \beta = 0$
$H_1 : \beta \neq 0$

$F_{1,1}$ 5% = 161.4

$F_{cal} < F_{tab}$

$\therefore$ Do not reject H_0