PAGE No.
DATE / /

P&S Assignment -2

Name :- Rohit Dubey
Roll no : 19

Q1) $\Sigma x_i^2 = 92500$ $\Sigma x = 850$
$\bar{x} = 85$

$\Sigma xy = 182560$ $\Sigma y = 1737$
$\bar{x}\bar{y} = 14764.5$

$S_{xx} = \Sigma x_i^2 - n\bar{x}^2$
$= 20250$

$S_{xy} = \Sigma xy - n\bar{x}\bar{y}$
$= 34915$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$
$= 1.724$

$\alpha = \bar{y} - \hat{\beta}\bar{x}$
$= 27.1432$

Thus, equation
$y = 27.1432 + 1.724\, x_i$

The gradient is uphill indicating the data is positively correlated.

Q2) i) $S_{xx} = \Sigma x_i^2 - n\bar{x}^2$
$= 301.66$

$S_{xy} = \Sigma xy - n\bar{x}\bar{y}$
$= 875.16$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ $\qquad S_{yy} = \Sigma y^2 - n\bar{y}^2$
$= 1409.712$

$= 2.901147$

$\alpha = 3.748182$

$\therefore$ Equation
$y = 3.748182 + 2.901147 x_i$

ii) 95% CI

$$\hat{\beta} \pm t_{0.025, 10} \sqrt{\frac{\sigma^2}{S_{xx}}}$$

$$\sigma^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$$
$= 6155.816$

95% CI is $(-7.16351, 12.96581)$

3)
i) The variance of all the cities appear to be mostly same.
The centre's of all distributions differ.

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

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

$$SS_{TOT} = \frac{1}{7}\left(64^2 + 44^2 + 8^2 + (-13)^2\right) - \frac{103^2}{25}$$

$$= 516.11$$

$$SS_{Res} = SS_{TOT} - SS_{Reg}$$

$$= 600$$

ANOVA Table.

Source	dof	SS	MSS	F
Regression	3	516.11	172.04	6.88
Residual	24	600	25	
Total	27	1116.11		

$F = 6.88$ $F_{3,24} = 3.009$ at 5%

∴ ~~will be~~

∴ We will reject H_0.

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

$= 69930.06$

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

$= 22325.69$

$SS_{Res} = SS_{TOT} - SS_{Reg}$

$= 47604.37$

ANOVA Table.

Source	DoF	SS	MSS	F
Regression	4	22325.69	5581.42	3.28
Residual	28	47604.37	1700.156	
Total	32	69930.06		

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

∴ we will reject H_0

1] Contingency Table

Salary

Papers Cleared	45	20	6	5
	7	20	9	6
	5	8	15	20

$$\text{Expected} = \frac{\text{Row total} \times \text{Column total}}{\text{Table total}}$$

∴ Expected Table

27.4	23.8	14.4	11.06
15.4	12.7	8.19	6.09
14.43	12.6	9.8	5.8

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

$= 49.919$

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

Thus, we reject H_0.

∴ No of papers cleared are dependent on salary.

Q5)

Source	d.o.f	SS	MSS	F
Regression	3	21.153	7.051	45.638
Residual	8	1.236	0.155	
Total	11	22.38		

$F_{3,8,5\%} = 7.951$

$\therefore$ We reject H_0

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

$S_{xy} = \Sigma xy - n\bar{x}\bar{y}$
$= 661.2$

$S_{xx} = \Sigma x_i^2 - n\bar{x}^2$
$= 54.9$

$S_{yy} = \Sigma y_i^2 - n\bar{y}^2$
$= 8923.6$

i) $r = 0.944662$

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

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

$$y_i = 9.229508 + 12.04372\ x_i$$

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

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

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

$$R^2 = \frac{SS_{REG}}{SS_{TOT}}$$
$$= 0.892387$$

R^2 is the square of correlation Coefficient
Both of them show positive relationship

7)

$S_{xy} = 2176.84$ given

$S_{xx} = \Sigma x_i^2 - n\bar{x}^2$
$= 4789.422$

$S_{yy} = \Sigma y_i^2 - n\bar{y}^2$
$= 1189.208$

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

PAGE No.

DATE / /

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

$= 10.79056$

∴ Equation.

$y_i = 10.79056 + 0.45451 x_i$

11) $\sigma^2 = \frac{1}{11-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$

$= 22.201$

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

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

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

$= 6.675675$

$t_{9, 0.025} = 2.262$

We reject H_0.

Here, there is a linear relationship.

PAGE No.
DATE / /

iii) $\gamma = \dfrac{S^2xy}{\sqrt{Sxx\ Syy}}$

$= 0.912129$

v) The residual plot shows that the errors are forming a pattern thus showing they might not form a non linear relationship.

Q8)

i) The main purpose of factor analysis is to reduce many individual item into a fewer number of dimension.
Principle component analysis is a method to decrease dimensions without any data loss.

ii)

PC	Diagonal Entry (PC_i)	PC_i / sum(PC_i)
PC_1	0.456	65%
PC_2	0.137	18.5%
PC_3	0.06	11.4%
PC_4	0.0165	2.4%
PC_5	0.012	1.7%
	0.1015	100%

iii) As 1^{st} 3 principle components explained over 95% of total variance, dimensionally can be reduced to 3 for this data set.

PAGE No.

DATE / /

Q9) i) Apperas to be negatively correlated.

ii) $S_{xx} = 75$

$S_{yy} = 2482.4$

$S_{xy} = 296$

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

$= 0.686002$

Hence (-)ve correlation

iii) $H_0 : \rho = 0$

vs

$H_1 : \rho < 0$

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

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

$Z_{cal} = -2.07893$

$Z_{5\%} = 1.6449$

∴ We reject H_0

∴ $\rho < 0$.

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

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

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

v) $R^2 = \frac{S_{xy}}{S_{xx} S_{yy}} = 47.05983\%$

As model only explains 47% of the data, it is not a good fit.

vi) $x_0 = 1$
$y_0 = \hat{a} - \hat{\beta} x_0$
$= 78.2133.$

10) $S_{xx} = 0.0042$
$S_{yy} = 0.00126$
$S_{xy} = 0.0022$

$SS_{TOT} = 0.00126$
$SS_{REG} = \frac{S^2_{xy}}{S_{xx}} = 0.00115$

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

$= 0.0011$

PAGE No.
DATE / /

ANOVA Table

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

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

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

$F_{cal} < F$

$\therefore$ We will not reject H_0.