P& S Assignment - 2.

Name: Ganesh Chaudhary.
Roll no: 10.

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

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

$S_{xx} = \Sigma x_i^2 - n\bar{x}^2$
$= \cancel{9765200}.\ 20250$.

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

$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$.
$= \cancel{0.23380096}.\ 1.724$.

$\alpha = \bar{y} - \hat{\beta}\bar{x}$
$= \cancel{147.3184}$.
$= 27.1432$.

Thus, equation
$y = 27.1432 + 1.724 x_i$

The gradient is ~~positive~~ uphill indicating the data is positively correlated.

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

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

$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$
$= 2.901147.$

$S_{yy} = \Sigma y^2 - n\bar{y}^2.$
$= 61409.712.$

$\alpha = 3.748182.$

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

ii]. 95% CI.

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

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

$= 6155.816.$

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

3.

i] The variance of all the cities appear to be mostly same.

The centres of all distributions differ.

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.

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

$$SS_{Reg} = \frac{1}{7}\left(54^2 + 44^2 + 8^2 + (-13)^2\right) - \frac{103^2}{28}$$

$$= 516.11$$

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

$$= 600.$$

ANOVA Table.

Source.	df	SS	MS.	F
Regression.	3	516.11	172.04	6.88.
Residual	24	600.00	25	
Total.	27.	1116.11		

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

$\therefore$ We reject H_0.

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

$= 69930.06$.

$SS_{Reg} = \sum \frac{y_{i.}^2}{n_i} - \frac{y^2_{..}}{n}$.

$= 22325.69$.

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

$= 47604.37$.

ANOVA Table.

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

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

∴ We reject H_0.

c] 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.

27.4	23.8	14.81	11.06
15.4	12.7	8.19	6.09.
14.13	12.6	7.8	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	df	SS	MSS.	F.
Regression.	3	21.153	7.051	45.638
Residual.	8	1.236.	0.155	
Total	11	22.38.		

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

$\therefore$ We reject H_0.

6.

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

$$S_{xy} = \sum xy - n\bar{x}\bar{y}$$
$$= 2853 - 10 \times 3.9 \times 56.2$$
$$= 661.2$$

$$S_{xx} = \sum x_i^2 - n\bar{x}^2$$
$$= 207 - 10(3.9^2)$$
$$= 54.9$$

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

i]. $r = 0.944662$

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

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

$$y = 9.229508 + 12.04372 x_i$$

$SS_{TOT} = S_{YY}$
$= 8923.6.$

$SS_{REG} = \frac{S^2_{XY}}{S_{XX}}$
$= 7963.305$

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

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

$= 0.892387.$

R^2 is the square of correlation coefficient r.

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]. $\beta = \frac{S_{xy}}{S_{xx}}$

$= 0.45451$

$\alpha = \bar{y} - \hat{\beta}\bar{x}$
$= 10.79056$.

$\therefore$ Equation.

$y = 10.79056 + 0.45451 x_i$

ii]. $\hat{\sigma}^2 = \frac{1}{11-2}\left(S_{yy} - \frac{S^2_{xy}}{S_{xx}}\right)$.

$= 22.201$.

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

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

classmate

Date
Page

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

$= 6.675675$

$t_{9, 0.025} = 2.262$.

$\therefore$ We reject H_0 ~~and~~

Hence, there is a linear relationship.

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

$= 0.912129$.

v) The residual plot shows that the 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 fewere 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)
PC1	0.456	65%
PC2	0.137	18.5%
PC3	0.08	11.4%
PC4	0.0165	2.4%
PC5	0.012	1.7%
	0.1015	100%

iii) As 1st 3 principle components explain over 95% of total variance, dimensionality can be reduced to 3 for this data set.

Date
Page

Q9 i) Appears to be negatively correlated.

ii) $S_{xx} = 75$

$S_{yy} = 2482.4$

$S_{xy} = -296.$

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

$= 0.686002.$

Hence (-)ve correlation

iii) $H_0 : \rho = 0.$

vs

$H_1 : \rho < 0.$

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

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

$Z_{cal} = -2.07893.$

$Z_{5\%} = 1.6449.$

$\therefore$ we reject H_0.

$\therefore \rho < 0.$

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

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

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

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) $x_0 = 1$
$y_0 = \hat{\alpha} - \hat{\beta}x_0$
$= 78.21333.$

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

$SS_{TOT} = 0.00126.$

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

$SS_{RES} = SS_{TOT} - SS_{REG}$
$= 0.0011.$

ANOVA Table.

Surce	df	SS	MS	F
Regression	1	0.00115	0.00115	10.454
Residual	1	0.00011	0.00011	
Total.	2.	0.00126		

$H_0 : \beta = 0,$

$H_1 : \beta \neq 0.$

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

$F_{cal} < F_{tab}$

$\therefore$ We will not reject H_0.