Name - Maitreyi Pawar
FY BSc ASQF Div B
Subject : Probability and statistics
Assignment 2

1. From the given data,

(a) S_{x} $\Sigma x^2 = 92500$ $\Sigma x = 850$

$\Sigma y^2 = 385002$ $\Sigma y = 1764$

$\Sigma xy = 184720$

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 92500 - \frac{(850)^2}{10}$$
$$= 20250$$

$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 184720 - \frac{(850 \times 1764)}{10}$$
$$= 34780$$

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 385002 - \frac{(1764)^2}{10}$$
$$= 73832.4$$

$$\beta = \frac{S_{xy}}{S_{xx}} = \frac{34780}{20250} = 1.7175$$

$$\alpha = \bar{y} - \beta\bar{x} = 176.4 - 1.7175(85)$$
$$= 30.40995$$

$\therefore$ The regression line of y on x is
$$y = \alpha + \beta x = 30.40995 + 1.7175x$$

(b) The gradient of regression line β is 1.717 which is a positive no. This denotes that x and y shows a positive linear relationship with each other.

2. given :

$\Sigma x = 385.2$

$\Sigma x^2 = 12666.58$

$\Sigma y = 1162.5$

$\Sigma y^2 = 119026.9$

$\Sigma xy = 38191.41$

i) $S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 12666.58 - \frac{(385.2)^2}{12}$

$= 301.66$

$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = (119026.9) - \frac{(1162.5)^2}{12}$

$= 6409.7125$

$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 38191.41 - \frac{(385.2 \times 1162.5)}{12}$

$= 875.16$

$\beta = \frac{S_{xy}}{S_{xx}} = \frac{875.16}{301.66} = 2.90115$

$$\alpha = \bar{y} - \beta\bar{x} = \frac{\cancel{385.2}\ 1162.5}{12} - 2.90115\left(\frac{385.2}{12}\right)$$
$$= 96.875 - 2.90115(32.1)$$
$$= 3.748085$$

∴ least squares fit regression line:

$$y = \alpha + \beta x$$
$$y = 3.748085 + 2.90115x$$

ii) 95% c.i for $\beta \in \left(\hat{\beta} \pm t_{n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right)$

∴ 95% c.i for $\hat{\beta} = \left(2.90115 \pm t_{10}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right)$

$$\hat{\sigma}^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$$
$$= \frac{1}{n-2}\left(6409.7125 - \frac{(875.16)^2}{301.66}\right)$$
$$= \frac{1}{10} \times 3870.745$$
$$= 387.0745$$

∴ 95% c.i. for $\beta \in \left(2.90115 \pm 2.228\sqrt{\frac{387.0745}{301.66}}\right)$

95% c.i for $\beta \in (2.90115 \pm 2.5238)$

95% c.i for $\beta \in (0.37735, 5.42494)$

(iii) $Var(\hat{y}) = \left\{ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx} \right\} \sigma^2$

$= \left\{ \frac{1}{12} + \frac{(40 - 32.1)^2}{301.66} \right\} 387.0745$

$= 112.3375$

C.I for $y\hat{i} \in \left(\hat{y}_i \pm t_{10} \sqrt{Var(\hat{y})} \right)$

$\hat{y}_1 = (3.748085 \pm 2.228\sqrt{112.3375})$
$= (3.748085 \pm 23.6144)$
$= (-19.866, 27.3626)$

3.
i) On basis of the plot, first ~~one~~ two ~~is~~ are positively skewed, third is symmetric and third is negatively skewed.

ii) Assumptions of ANOVA.
→ data is normally distributed.
→ It assumes homogeneity of variance; which means that the variance among the groups should be approximately equal.
→ ANOVA also assumes that the observations are independent of each other.

(iii) $SS_{REG} = \frac{S^2_{xy}}{S_{xx}} = \frac{875.16^2}{301.66} = 2538.967$

$SS_{TOT} = S_{yy} = 6409.7125$

$SS_{RES} = SS_{TOT} - SS_{REG} = 6409.7125 - 2538.967$
$= 3870.744$

Under Ho, Test Statistic

$$\frac{SS_{REG}/1}{SS_{RES}/n-2} \sim F_{1,n-2}$$

$$TSV = \frac{2538.967}{3870.744/10}$$
$$= 6.55938$$

$F_{1,10} = 4.965$

$F_{cal} > T_{tab}$ $\therefore$ Ho is rejected.

FOR EDUCATIONAL USE

Sundaram

4.

(a) Assumption of ANOVA

→ data is normally distributed

→ It assumes homogeneity of variances, which means that the variance among the groups should be approximately equal.

→ ANOVA also assumes that the observations are independent of each other.

(b) <u>ANOVA</u>

Test Statistic :

$$\frac{SS_{REG} / k-1}{SS_{TOT} / n-k} = \frac{SS_{REG}/4}{SS_{TOT}/n-5}$$

6. $\Sigma x = 39$ $\qquad \Sigma y = 562$ $\qquad \Sigma xy = 2853$

$\Sigma x^2 = 207$ $\qquad \Sigma y^2 = 40508$

i) $$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 207 - \frac{(39)^2}{10} = 54.9$$

$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 2853 - \frac{(39 \times 562)}{10} = 661.2$$

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 40508 - \frac{(562)^2}{10} = 8923.6$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{661.2}{\sqrt{54.9 \times 8923.6}} = 0.94466$$

ii) $$\beta = \frac{S_{xy}}{S_{xx}} = \frac{661.2}{54.9} = 12.0437$$

$$\alpha = \bar{y} - \beta \bar{x} = \frac{\Sigma y}{n} - 12.0437 \left(\frac{\Sigma x}{n}\right)$$

$$= \frac{562}{10} - 12.0437 \times \left(\frac{39}{10}\right)$$

$$= 9.229$$

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

$$y = 9.229 + 12.0437x$$

iii) $SSREG = \frac{S^2xy}{Sxx} = \frac{661.2^2}{54.9} = 7963.304$

$SSTOT = Syy = 8923.6$

$SSRES = SSREG - SSTOT = 960.2951$

iv) $R^2 = \frac{SSREG}{SSTOT} = \frac{7963.304}{8923.6} = 0.89239$

$R^2 = r^2$ is the relationship

7. $\Sigma x = 174.13$ $\Sigma x^2 = \cancel{74} \, 7545.90$
$\Sigma y = 197.84$ $\Sigma y^2 = 4747.45$
$\Sigma xy = 2176.84$

i) $Sxx = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$

$= 7545.90 - \frac{174.13^2}{10}$

$= 4513.774$

$Sxy = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 2176.84 - \left(\frac{174.13 \times 197.84}{10}\right)$

$= -1268.148$

$$S_{yy} = \Sigma_i y^2 - \frac{(\Sigma_i y)^2}{n} = 833.383$$

$$\beta = \frac{S_{xy}}{S_{xx}} = \frac{-1268.148}{4513.774} = -0.28095$$

$$\alpha = \bar{y} - \beta\bar{x} = 19.784 - (-0.28095)17.413$$
$$= 24.6762$$

$$y = \alpha + \beta x$$
$$= 24.6762 - 0.28095x$$

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

$$TSV : \frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \qquad C.V = t_{n-2}$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$$
$$= \frac{1}{8}\left(833.383 - \frac{(-1268.148)^2}{4513.774}\right)$$
$$= 59.63699$$

$$TSV = \frac{-0.28095 - 0}{\sqrt{\frac{59.63699}{4513.774}}} = -2.444$$

$$C.V = t_{8,0.025} = -2.306$$

$\therefore$ Ho is rejected at 5% significance.

iii) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \dfrac{-1268.148}{\sqrt{4513.774 \times 833.383}}$

$= -0.65385$

iv) $x = 25$

then $y = 24.6762 - \cancel{\beta}\, 0.28095x$
$= 24.6762 - 0.28095(25)$
$= 17.65245$

Mean

$var(\hat{y}) = \left\{ \dfrac{1}{10} + \dfrac{(25 - 17.413)^2}{4513.774} \right\} \times 59.63699$

$= 1.0024$

C. I for $y_i \in \left(\hat{y}_i \pm t_{8, 0.025} \sqrt{Var(\hat{y})} \right)$

95% CI for $y_i \in \left(17.65245 \pm 2.306 \sqrt{1.0024}\right)$

95% C.I for $y_i \in (17.65245 \pm 2.30877)$
95% C.I for $y_i \in (15.34367, 19.96122)$

Individual

$Var(\hat{y}) = \left\{ 1 + \dfrac{1}{10} + \dfrac{(25 - 17.413)^2}{4513.774} \right\} \times 59.63699$

$= 66.361218$

95% CI for $y_i \in \left(17.65245 \pm 2.306 \sqrt{66.3612}\right)$
95% CI for $y_i \in (17.65245 \pm 18.7852)$
95% CI for $y_i \in (-1.13277, 36.4376)$

v) The residual plot shows that the residuals are not randomly distributed about mean.

Drawbacks here is that model leads to underestimation of premium rates.

8.

i) Factor analysis is a powerful data reduction technique that enables researchers to investigate concepts that cannot easily be measured directly.

ii)

	PC1	PC2	PC3	PC4	PC5
	0.456	0.137	0.08	0.0165	0.012
%	$\frac{0.456}{0.7015}$	$\frac{0.137}{0.7015}$	$\frac{0.08}{0.7015}$	$\frac{0.0165}{0.7015}$	$\frac{0.012}{0.7015}$
	= 0.65	0.195	0.114	0.0235	0.017

iii)

Cumulative	0.65	0.845	0.959	0.9825	0.9995

Till component 4 PC4 it should be retained.

9.

i) The above scatterplot tells that marks scored and hours spent per day on social media is negatively linear relation.

ii) $S_{xx} = 277.5 - \frac{(45)^2}{10} = 75$

$S_{yy} = 4395.6 - \frac{(644)^2}{10} = 2482.4$

$S_{xy} = 2602 - \frac{644 \times 45}{10} = -296$

$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{-296}{\sqrt{75 \times 2482.4}} = -0.686$

The correlation coefficient is negative which suggests that the data is negatively linear relation.

iii) $H_0 : \rho = 0$
$H_1 : \rho < 0$
Fischer Z transformation

$Z_r = \frac{1}{2} \log\left(\frac{1+r}{1-r}\right) = \frac{1}{2} \log\left(\frac{1-0.686}{1+0.686}\right) = -0.365$

$Z_\rho = \frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right) = \frac{1}{2} \log 1 = 0$

Test Statistic : $\frac{-0.365-0}{\sqrt{1/7}} = -0.9657$

P. value $= P(Z < -0.9657) = 0.013$

This is small than 0.05 therefore we ~~cannot~~ can reject H_0, $\rho < 0$

iv) $\beta = \frac{S_{xy}}{S_{xx}} = \frac{-296}{75} = -3.94667$

$\alpha = \overline{y} - \beta\overline{x} = \frac{\Sigma y}{10} - \beta \frac{\Sigma x}{10}$

$= 64.4 + 3.94667 \times 4.5$

$= 64.4 + 17.76$

$= 82.16$

$\therefore$ Regression line is

$y = \alpha + \beta x$

$= 82.16 + (-3.94667)x$

$= 82.16 - 3.94667x$

v) $R^2 = r^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}\right)^2 = (-0.686)^2 = 0.470596$

10. H_0 : There is no difference among industries

H_1 : Atleast one industry differs significantly from overall mean

$$SS_R = 19[5^2 + 10^2 + 8^2] = 3591$$

$$\bar{X}_{Resignation} = \frac{(27+36+30)}{3} = 31$$

$$SS_B = 20[(27-31)^2 + (36-31)^2 + (30-31)^2]$$
$$= 840$$

$$F_{2,57} = \frac{SS_B/2}{SS_R/59-2} = 6.667$$

$F_{2,20}$ at 1% $= 4.977$

$F_{cal} > F_{tab}$

∴ We can reject H_0

∴ There is no difference among industries.

∴ Resignation rate is impacted by industry.