1) $\Sigma x^2 = 92500$ $\qquad \Sigma x = 850$

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

$\Sigma xy = 184720$

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

$S_{xx} = 20250$

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

$S_{xy} = 34780$

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

$S_{yy} = 73832.4$

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

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

$= 30.40995$

∴ The regression line of y on x is

$y = \alpha + \beta x = 30.40995 + 1.7175x$.

b) The gradient of regression line B is 1.717 which is positive. ∴ x and y have positive linear relationship with each other

Date:

2) $\Sigma x = 385.2$

$\Sigma x^2 = 12666.58$

$\Sigma y = 1162.5$

$\Sigma y^2 = 119026.7$

$\Sigma xy = 38191.41$

i) $S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 301.66$

$S_{xy} = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n} = 875.16$

$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 6409.7125$

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

$\alpha = \bar{y} - B\bar{x} = \frac{1162.5}{12} - 2.90115\left(\frac{382.5}{12}\right).$

$= 96.875 - 2.90115\ (32.1)$

$\alpha = 3.748085$

$\therefore$ Least squares fit regression line.

$y = \alpha + Bx$

$y = 3.7481 + (2.90115)x$

ii) 95% C.I for $B = \hat{\beta} \pm t_{n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

$\therefore$ 95% CI for $B = 2.90115 \pm t_{10}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

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

$= \frac{1}{10}\left(6409.7125 - \frac{(875.16)^2}{301.66}\right)$

$= 387.0745$

95% C.I for $B \in \left(2.90115 \pm 2.228\sqrt{\frac{387.0745}{301.66}}\right)$

95% CI for $B \in (2.90115 \pm 2.5238)$

95% CI for $B \in (0.37735,\ 5.42494)$

Date:

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

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

$= 112.3375$

CI for $\hat{y}_i$ $\quad y_i \in \left(\hat{y}_i \pm t_{10} \sqrt{\text{var}(\hat{y})} \right)$

$\hat{y}_i = (3.748085 \pm 2.228 \sqrt{112.3375})$

$= (3.748085 \pm 23.6144)$

$= (-19.866, \ 27.3626)$

3.

i) On basis of the plot, first two are +vely skewed, third is symetric; fourth is -vely skewed

ii) Assumption of ANOVA :-

1) Data is normally distributed
2) Homogenity of variance
3) Observations are independent of each other

iii) $SSreg = \frac{S_{xy}^2}{Sxx} = \frac{875.16^2}{301.66} = 2538.967$

$SStotal = Syy = 6409.7125$

$SSres = SS\,total - SSreg = 3870.744.$

Under Ho,

$\frac{SSreg/1}{SSres/n-2} \sim F_{1,n-2}$

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

$F_{1,10} = 4.965$

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

7. $\Sigma x = 1734.13$ $\qquad \bar{x} = 15.83$

$\Sigma y = 197.84$ $\qquad \bar{y} = 17.985$

$\Sigma x^2 = 7545.9$

$\Sigma y^2 = 4747.45$

$\Sigma(x-\bar{x})(y-\bar{y}) = 2176.84$

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

$= \Sigma x^2 - n\left(\frac{\Sigma x}{n}\right)^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$

$= 4789.4221$

$S_{xy} = \Sigma(x-\bar{x})(y-\bar{y}) = 2176.84$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{2176.84}{4789.42} = 0.4545$

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

$= 17.985 - (0.4545 \times 15.83)$

$= 10.79$

∴ The fitted regression equation is

$\hat{y} = 10.79 + (0.4545)x$

(ii) $S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 1189.21$

$S_{xx} = 4789.4221$

$S_{xy} = 2176.84$

$\hat{\sigma}^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = \frac{1}{9}\left(1189.21 - \frac{(2176.84)^2}{4789.42}\right)$

$\hat{\sigma}^2 = 22.20$

$\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{22.20}{4789.42}} = 0.0681$

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

$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \frac{0.4545 - 0}{0.0681} = 6.674$

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

$\therefore t_{9,\,0.025} = 2.68$

$\therefore$ ~~the~~ critical value at 95% l o s is less than test static

$\therefore$ we can reject null hypothesis.

$\therefore$ No significant relationship b/w x & y

iii) Pearson's correlation coefficient

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

$$r = \frac{2176.84}{\sqrt{(4789.42)(1189.21)}} = 0.91$$

iv) $x^3 = 25,$ $\quad y = 10.79 + (0.4545)(25)$

$= 22.15$

$\hat{\sigma}^2 = 22.20$

$\therefore$ Variance of mean response estimator

$$\left[\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right]\hat{\sigma}^2 = \left[\frac{1}{11} + \frac{84.0889}{4789.42}\right](22.20)$$

$$= 2.41$$

Variance of individual response estimator

$$\left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right]\hat{\sigma}^2 = \left[1 + \frac{1}{11} + \frac{84.0889}{4789.42}\right](22.2)$$

$$= 24.61$$

using t_9 distribution, the 95% C.I. for mean and individual responses are

$22.20 \pm 2.262\sqrt{2.41} = (18.7, 25.7)$

$22.20 \pm 2.262\sqrt{24.61} = (11.0, 33.4)$

v) The residual plot shows a definite pattern

Model is inappropriate

Model leads to underestimation of premium rates.

8) Factor analysis is a powerful data reduction technique that enables researchers to investigate concepts that can't easily be measured.

	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
cumulative	0.65	0.845	0.959	0.9825	0.9995

till component 4 PC4 should be retained.

9) i) The scatterplot suggests an inverse relation between marks obtained and hours spent on social media per day.

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

$S_{yy} = 43956 - \frac{(644)^2}{10} = 2482.4$

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

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

-69% correlation coefficient also implies a moderate negative linear relation between the two variables as visible from the scatterplot.

iii) Null hypothesis $H_0 : \rho = 0$ against $H_1 : \rho < 0$

Need to assume that data come from a bivariate normal distribution.

$r = 0.5 \ln\left(\frac{1-0.686}{1.686}\right) = -0.8404$

$N(0, 1/7)$

~~He~~ Fisher's standardized statistic $= \frac{(-0.8404 - 0)}{\sqrt{1/7}}$

$= -2.22$

~~The~~ This gives the p-value $= P(z < -2.22)$

$= 0.013$

which is quite small and hence shows that ~~we~~ we can reject null hypothesis at 95% CI. ∴ Marks and hours spent on social media are -vely correlated.

10) H_0: There is no difference among industries

H_1: Atleast one industry differs significantly from overall mean

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

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

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

$= 840$

$F_{2,57} = \frac{(840/2)}{(3591/57)} = 6.667$

~~$F_{2,57}$~~ $F_{2,60}$ at 1% $= 4.977$

$F_{cal} > F_{tab}$

$\therefore$ we can reject H_0

$\therefore$ Resignation rate is impacted by industry.