Krishangi Trivedi, 82

i) (a) $\Sigma x^2 = 92500$

$\bar{x} = 850/10 = 85$

$\Sigma y^2 = 372777$

$\bar{y} = 173.7$

$\Sigma xy = 182560$.

$S_{xy} = 182560 - n\bar{x}\bar{y} =$

$= 34915$

$S_{xx} = 20250$

$S_{yy} = 71000.1$

$\beta = \frac{S_{xy}}{S_{xx}} = 1.724197531$

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

$= 27.1432$

$\therefore$ equation of regression line

$y_i = \alpha + \beta x_i$

$\therefore y_i = 27.1432 + 1.724197 x_i$

(b) It can be hence concluded that the money spent by Anand's girlfriend is affected by the amount of cash he withdrew at his previous visit to the ATM

Page No.
Date

2.

(i) $\cancel{\beta} = S_{xy} = 38191.41 - \frac{38 75.2 \times 1325}{12}$

$= S_{xy} = 875.16$

$S_{xx} = 301.66$

$S_{yy} = 6407.7125$

$\beta = \frac{S_{xy}}{S_{xx}} = 2.901146967.$

$\alpha = \bar{y} - 2.901146 \times \bar{x}$

$= 3.748181$

$y_i = 3.748181 - 2.901146 x_i$

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

$= 1.182761292.$

$\beta \pm t_{(n-2),\frac{5\%}{2}} \times 1.132761271$

$0.3773547,\ 5.424730162$

(iii) mean IBH share price

$S.e(\cancel{\ }) = \sqrt{\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)} \times \sigma$

$= \sqrt{0.290221303} \times \sigma$

$= 3.096042514$

Page No. | Date

$\mu_0 = 119.79406$

$CI = 112.9406373, 126.6474627$

Q.2 (i) The time difference in commutation of city A seems the highest meaning the time taken to commute during peak hours is drastically different than the time taken to commute during non-peak hours.

The time difference in commute for city B is high but it is lower than the time difference of city A.

The time difference in commute for city C is lower than city B. Variation of city C seems the lowest
The time difference in commute of city D is the least among the four cities. Variation seems higher.

(ii) i) Assumption of homoscedasticity i.e. the population variance is the same.
(ii) The observations are independent
(iii) They must follow normal distribution.

(iii) H_0: The mean difference is the same for all cities.

$$SS_{total} = 1495 - \frac{103^2}{28}$$

$$= 1116.11$$

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

$$= 516.11$$

$$SS_{res} = SS_{tot} - SS_{reg}$$

$$= 1116.11 - 516.11$$

$$= 600$$

ANOVA table :

~~SS_{reg}~~	df	SS	MSS
regression	3	516.11	172.04
residual	24	600	25
total	27	1116.11	

~~F~~ F test calculated $= \frac{172.4}{25} = 6.88$
which is much greater than
3.009 the 5% critical point of F distribution
∴ We have enough evidence to reject H_0
∴ we can conclude that there are underlying differences b/w the cities.

Page No.
Date

Q.4. Mathematical one way analysis.

$$y_{ij} = \alpha + \beta x_{ij} + e_{ij}$$

$i = 1, 2, 3, \ldots k$

$j = 1, 2, 3, \ldots n_i$

k = number of treatments

n_i = number of responses from ith ~~term~~ treatment.

y_{ij} = is the j^{th} response from the ith treatment

e_{ij} = is the error term.

$e_{ij} \approx N(0,1)$

c) H_0: Salaries are independent

H_1 = Salaries are dependent.

O_i

Paper Cler	3 - 5	5 - 8	8 - 10	10 - 12	total
0 - 3	45	20	6	5	76
4 - 6	7	20	9	6	42
7 - 9	5	8	15	12	40
Total	57	48	30	23	158

E_i

	3-5	5-8	8-10	10-12	Total
0-3	27.42	23.09	14.43	11.06	76
4-6	15.15	12.76	7.97	6.11	42
7-9	14.43	12.15	7.57	5.82	40
Total	57	48	30	23	158

$$\frac{\sum(O_i - E_i)^2}{E_i} = 49.919$$

$$df = (4-1)(3-1) = 6$$

$$\chi^2_{tab} = 12.59 .$$

$$\chi^2_{cal} > \chi^2_{tab}$$

∴ We have sufficient evidence to Reject H_0

∴ we conclude that Salaries is depend on the number of papers cleared.

Page No.
Date

~~Q.4.~~ Q.5 (i)

~~(a)~~ Assumptions are:

The population must be normal

The observations must be independant

The assumption of homoscadasticity.

The sample variances are different hence the assumption that population variance is same will not ~~(ii)~~ hold true.

(ii) new Sample mean of rate 1

$$\frac{1}{3}\left(\sqrt{29}+\sqrt{13}+\sqrt{21}\right)=4.52$$

New sample variance for rate 2

$$\frac{(\log_e 180-4.827)^2+(\log_e 90-4.827)^2+(\log_e 120-4.827)^2}{2}$$

$$=0.1213$$

(iii) For this assumption the sample variance is same hence the assumption of homoscedasticity holds true. hence this transformation is viable.

The scientist was correct in asserting that $\log_e$ transformation must be done.

Page No.
Date

Q.6.

(i) $S_{xx} = \sum x^2 - n\bar{x}^2$
$= 54.90$

$S_{yy} = \sum y^2 - n\bar{y}^2$
$= 8923.60$

$S_{xy} = \sum xy - n\bar{x}\bar{y}$
$= 661.20$

$r = \dfrac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = 0.945$

(ii) $\beta = \dfrac{S_{xy}}{S_{xx}} = \cancel{661}\ \dfrac{661.20}{54.9} = 12.04$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 9.23$

$\therefore \hat{y} = \hat{\alpha} + \hat{\beta}x = 9.23 + 12.04x$

(iii) $SS_{total} = S_{yy} = 8923.6$

~~SSreg~~ $SS_{res} = S_{yy} - \dfrac{S^2_{xy}}{S_{xx}}$

$= 8923.6 - \dfrac{(661.2)^2}{(54.9)} = 960.30$

$SS_{reg} = 8923.6 - 960.3 = 7963.3$

(iv) COD = $R^2 = r^2 = (0.945)^2 = 0.8927$

Q.7. (i) $S_{xx} = 4789.42$

$S_{yy} = 1189.21$

$S_{xy} = 2176.84$

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

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 10.79.$

$\therefore y = 10.79 + 0.4545x$

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

$$= \frac{1}{9} \times \left(1189.2 - \frac{(2176.84)^2}{4789.42}\right) = 22.2$$

$$S.e.(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{22.2}{4789.42}} = 0.0681$$

$H_0: \beta = 0$; $H_1: \hat{\beta} \neq 0$

$$\frac{\hat{\beta} - 0}{s.e(\hat{\beta})} = \frac{0.4545}{0.0681} = 6.672.$$

Page No.
Date

The errors of the regression are iid and follow $N(0, \sigma^2)$; β has t distribution with $n-2$ dof.

at dof 9 the critical value of t dist is 2.68

the calculated value is more hence there is sufficient evidence to reject H_0.
Hence there is no relationship between x & y

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

$= 0.91$

(iv) $x^0 = 25$.

$y = 10.79 + 0.4546 \times 25$
$= 22.15$

$\hat{\sigma}^2 = 22.2$

$$V(\mu_0) = \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \hat{\sigma}^2$$

$= 2.41$

* variance of individual response

$$\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \times \hat{\sigma}^2 = 24.61$$

95% CI for mean resp

$$22.20 \pm 2.262 \times \sqrt{2.41}$$

$$= (16.7, 25.7)$$

95% CI for individual response

$$22.2 \pm 2.262 \times \sqrt{24.61}$$

$$= (11, 33.4)$$

(v) The residual plot shows a pattern. The correlation coefficient is high but the model does not seem to be a good fit because of this.

Q.9.(i) The marks scored and the number of hrs spent on social media seem to be negatively correlated according to the scatter plot

(ii) $S_{xx} = \Sigma x^2 - n\bar{x}^2 = 75$

$S_{yy} = \Sigma y^2 - n\bar{y}^2 = 2482.4$

$S_{xy} = \Sigma xy - n\bar{x}\bar{y} = -296$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = -0.686$$

∴ It is confirmed that they data is negatively correlated

Page No.
Date

(iii) $H_0: \rho = 0$

$H_1: \rho < 0$

$$r = \frac{1}{2} \ln\left(\frac{1-0.686}{1+0.686}\right) = -0.8404$$

$$W \sim N\left(0, \frac{1}{7}\right)$$

$$t_s = \frac{-0.8404-0}{\sqrt{1/7}} = -2.22$$

which is higher. than the tabulated value hence we have ~~seen~~ enough evidence to reject H_0. $\therefore$ marks & number of hrs spent on social media is negatively correlated.

(iv) $\beta = \frac{Sxy}{Sxx} = -3.9467$

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

$y = 82.16 - 3.947x$.

(v) $R^2 = r^2 = 0.4706$

47.06% of the total variation of the model is explained.

(vi) For every additional hr spent of sm per day the total marks by reduce in by 3.95 basis of fitted equation

Q. 10.

H_0: No difference among
H_1: atleast one industry differs significantly from the mean.

$$SS_T = \frac{5^2 + 10^2 + 8^2}{17} = 3591$$

~~MSS~~ mean of resignation $= \frac{27+36+30}{3} = 31$

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

$= 840$

$$F_{2,57} = 840/2 \Big/ 3591/57 = 6.667$$

at ~~os~~ 1% los of F dist with dof 2, 60, is critical value is 4.977

∵ calculated value is higher than this we reject H_0

∴ resignation ~~is dif~~ rate is different accross different ~~com~~ countries