(15)

Date : ....   Page No. ....

(1) ~~$\Sigma x =$~~

a. $\Sigma x^2 = 92500$ $\Sigma y^2 = 372717$ $\Sigma xy = 182560$

$\bar{x} = 85$ $\bar{y} = 173.7$

$S_{xx} = 20250$

$S_{yy} = 71000.1$

$S_{xy} = 34915.$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 1.7242$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 27.143$

$\Rightarrow y = 27.143 + 1.7242x$

b. gradient is number of hours he spent in atm upon amount withdrawn.

GOENKAN

Date : ....   Page No. ....

(2) $\Sigma x = 385.2$ $\Sigma y : 1162.5$ $\Sigma xy = 38191.41$

$\Sigma x^2 = 12666.58$ $\Sigma y^2 = 119026.9$

i. $S_{xx} = 12666.58 - 12\left(\frac{385.2}{12}\right)^2 = \underline{301.66}$

$S_{yy} = 119026.9 - 12\left(\frac{1162.5}{12}\right)^2 = \underline{6409.7125}$

$S_{xy} = 38191.41 - 12\left(\frac{385.2}{12}\right)\left(\frac{1162.5}{12}\right) = 875.16$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{875.16}{301.66} = 2.901147$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{1162.5}{12} - 2.901147\left(\frac{385.2}{12}\right)$

$\hat{\alpha} = 3.74818$

$\Rightarrow \quad y = 3.74818 + 2.901147\, x$

GOENKAN

ii. $\hat{\sigma}^2 = \frac{1}{10}\left[6409.7125 - \frac{875.16^2}{301.66}\right] = 387.0245$

95% CI of β

$$-2.228 < \frac{\hat{\beta} - \beta}{\hat{\sigma}/\sqrt{S_{xx}}} < 2.228$$

$$\hat{\beta} - 2.228\frac{\hat{\sigma}}{\sqrt{S_{xx}}} < \beta < \hat{\beta} + 2.228\frac{\hat{\sigma}}{\sqrt{S_{xx}}}$$

$$(0.377 < \beta < 5.425)$$

iii. $x_0 = 40$

$$\hat{\mu}_0 = 3.74818 + 2.901147(40) = 119.794$$

$$s.e(\hat{\mu}_0) = \sqrt{\left[\frac{1}{12} + \frac{(40-32.1)^2}{30166}\right] 387.0745}$$

$$= 10.5989$$

95% Confidence Interval of μ_0

$$\Rightarrow \quad 119.794 \pm \cancel{2.88}\ 2.2\cancel{28}(10.5989)$$

$$\Rightarrow \quad 119.794 \pm 23.6144$$

$$\Rightarrow \quad (96.17956, 143.4084)$$

(3) i. - The centers of the distributions differ for all the four cities.

- The difference between mean time are in the order City A (highest), City (D) lowest

- The variation in data city C is lowest compared to City D (highest).

ii. Assumptions required to carry out ANOVA.
- Populations must be normal
- Populations have a common variance.
- Observations are independent.

iii. Ho: The mean of differences is same for each city.

Ha: The mean of differences are not the same for all of the cities.

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

$$SS_{REG} = \frac{1}{7}\left(64^2+44^2+8^2+(-13)^2\right) - \frac{103^2}{28} = 516.11$$

$$SS_{RES} = 600$$

ANOVA

	df	SS	MS	F
REGRESSION	3	516.11	172.04	6.88
RESIDUAL	24	600	25	
TOTAL	27	1116.11		

Under H_0, $F = \frac{172.04}{25} = 6.88$.

$F_{0.05} = 3.009$, As $F_{cal} > F_{tab}$, we reject null hypothesis.

$\bar{y}_1 = 9.14$ $\bar{y}_2 = 6.29$ $\bar{y}_3 = 1.14$ $\bar{y}_4 = -1.86$

$t_{24, 0.05} * \hat{\sigma} \sqrt{\frac{1}{7} + \frac{1}{7}} = 2.064 * \sqrt{25} * \sqrt{\frac{2}{7}} = 5.52$

$\bar{y}_1 - \bar{y}_2 = 2.85$ $\bar{y}_1 - \bar{y}_3 = 8$

$\bar{y}_2 - \bar{y}_3 = 5.15$ $\bar{y}_2 - \bar{y}_4 = 8.15$

$\bar{y}_3 - \bar{y}_4 = 3$

No significant difference between $\bar{y}_1 - \bar{y}_2$, $\bar{y}_2 - \bar{y}_3$, $\bar{y}_3 - \bar{y}_4$

Significant difference between $\bar{y}_1 - \bar{y}_3$, $\bar{y}_2 - \bar{y}_4$.

4 a. Already done in qs. 3 part (ii).

b. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_a: At least one of the mean isn't same

$$\frac{T^2}{N} = \frac{(354+386+87+645+384)^2}{33} = 104385.9394$$

$$SS_{RES} = \frac{(354)^2}{7} + \frac{(386)^2}{8} + \frac{(87)^2}{6} + \frac{(645)^2}{7} + \frac{(384)^2}{5} - 104385.9394$$

$SS_{REG} =$ 22325.68917

$SS_{TOT} =$ 69930.0606

$SS_{RES} =$ 69930.0606 − 22325.68917

$=$ 47604.37143

ANOVA:

		Df	MS	F
REG	22325.68917	4	5581.4225	3.2829
RES	47604.37143	28	1700.154	
TOT	69930.0606	32		

$F_{4,28} =$ 2.21 As $F_{cal} > F_{tab}$, we reject H_0 at 5% level of significance

C. H_0: They are independent of each other
H_a: They are not independent.

Salary per annum (in Rs.)

Papers Cleared	3-5	5-8	8-10	10-12	
0-3	45	20	6	5	76
4-6	7	20	9	6	42
7-9	5	8	15	12	40
	57	48	30	23	158

Expected Values:

	3-5	5-8	8-10	10-12
0-3	27.42	23.088	14.43	11.06
4-6	15.15	12.76	7.97	6.14
7-9	14.43	12.15	7.59	5.82

$\chi^2 = 39.926$

$\chi^2_6 = 12.592$

As $\chi^2_{cal} > \chi^2_{tab}$ we reject null hypothesis at 5% level of significance.

(5) i. Already mentioned in qs. 3 part (ii).

ii. $x \to \sqrt{x}$, value of sample mean for rate 1

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

$x \to \log_e x$, the value of sample mean for rate 2

$\frac{1}{2}[(\log 180 - 4.827)^2 + (\log 90 - 4.827)^2 + (\log 120 - 4.827)^2]$

$= 0.1213$

iii. $\log_e$ transformation must be done before carrying out ANOVA as assumption of common variance for the underlying population holds.

iv. ANOVA on $\log x$ data.

We would assume the following model

$y_{ij} = \mu + \tau_i + e_{ij}$, $i = 1,2,3,4$, $j = 1,2,3$

$H_0: \tau_i = 0$, $i = 1,2,3,4$

$H_a: \tau_i \neq 0$ for at least one i.

Date : ....

$SS_{\cancel{REG}} = 21.153$

$SS_{REG} = 1.236$

ANOVA :

	Df	SS	MS	F
REG	3	21.153	7.051	45.638
RES	8	1.236	0.155	
TOT	11	22.389		

$F_{3,8} = 7.951$ at 1% l.o.s

As $F_{cal} > F_{tab}$, we reject Ho at 1% level of significance

⑥ $S_{xx} = 54.9$

$S_{yy} = 8923.6$

$S_{xy} = 661.2$

i. $r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = 0.94466$

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

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 9.22957$

$\Rightarrow \underline{y = 9.22957 + 12.0437x}$

iii. $SS_{TOT} = S_{yy} = 8932.6$

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

$SS_{RES} = 960.2951$

iv. $R^2 = \frac{SS_{REG}}{SS_{TOT}} = \frac{7963.305}{8932.6} = 0.891488$

89.148% variation is shown by model.

7. $S_{xx} = 4789.4221$

i. $S_{yy} = 1189.2077$

$S_{xy} = 2176.84$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 0.45451$

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

$y = \cancel{0.45451}\ 10.79056 + 0.45451x$

$\hat{\sigma}^2 = \frac{1}{9}\left[1189.2077 - \frac{2176.84^2}{4789.4221}\right] = 22.20136$

ii. H_0: There is not a linear relationship $(\beta = 0)$
H_a: There is linear relationship $(\beta \neq 0)$

$$Z_c = \frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \frac{\hat{\beta}}{\sqrt{22.20136 / 4789.4221}}$$

$$= \frac{0.45451}{0.068085}$$

$$= 6.67567$$

$$t_{9, 0.025} = 2.262$$

As $t_{cal} > t_{tab}$, we have sufficient evidence to reject H_0.

iii. $$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = 0.91213$$

iv. $$\hat{\mu}_0 = 12.11773$$

For mean response

$\Rightarrow \hat{\mu}_0 \pm 2.262 \; s.e(\hat{\mu}_0)$ ... $(s.e(\hat{\mu}_0) = \sqrt{2.790417})$
$\Rightarrow 12.11773 \pm 2.262 (2.790417)$

$$(8.3392, \; 15.8963)$$

for individual response

$\hat{y_o} = 12.11773$

$s.e(\hat{y_o}) = \sqrt{24.991777}$

$\Rightarrow \quad \hat{y_o} \pm 2.262\, s.e(\hat{y_o})$

$= (0.8096, 23.42587)$

v. Since there is some pattern seen
Hence, we can't say it exhibits normal distribution

⑨ i. Given plot shows -ve correlation between marks scored and hours spent per day on social media.

ii. $S_{xx} = 75$

$S_{yy} = 2482.4$

$S_{xy} = -296$

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

This shows -ve correlation.

iii. $H_0: \rho = 0$

$H_a: \rho < 0$

$$w = \tanh^{-1} r = -0.84036$$

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

$$w \sim N\left(0, 1/7\right)$$

$$\Rightarrow Z_c = \frac{w - \tanh^{-1} 0}{\sqrt{1/7}} = \frac{-0.84036}{\sqrt{1/7}} = -2.2234$$

$$Z_{tab} = -1.96$$

As $Z_c < Z_{tab}$, we reject null hypothesis at 5% level of significance

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

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

$$\Rightarrow y = 82.16 - 3.94667x$$

v. $$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = 0.470596$$ $\therefore$ 47.0596% variation is shown by modelled.

$x=1$

vi. $y_1 = 78.21333$

$x=2$

$y_2 = 74.26666$

For every additional 1 hour spent,

expected change = $\underline{-3.94667}$.

(10)