PAGE No.
DATE / /

prob and stats -2

(1) $\Sigma x^2 = 92500$, $\bar{x} = 85$
$\Sigma y^2 = 372717$; $\bar{y} = 173.7$
$\Sigma xy = 182560$

$S_{xx} = \Sigma x_i^2 - n(\bar{x})^2 \Rightarrow 20250$

$S_{yy} = \Sigma y^2 - n(\bar{y})^2 \rightarrow 71000.1$

$S_{xy} = \Sigma xy - n(\bar{x}\bar{y}) \Rightarrow 34915$

(a) $\beta = \dfrac{S_{xy}}{S_{xx}} = 1.7242$

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

So, $\hat{y} = 27.143 + 1.7242x$

(b) Gradient is no of hours he spend in upon amount withdrawn.

PAGE No.
DATE / /

(2) $\Sigma x = 385.2$ $\Sigma x^2 = 12666.58$

(i) $\Sigma y = 1162.5$ $\Sigma y^2 = 119026.9$

$\Sigma xy = 38191.41$

$S_{xx} = 301.66$ $S_{yy} = 6409.\cancel{7170}$ 7125

$S_{xy} = 875.16$

$\hat{\beta} = 2.901147$ $\hat{\alpha} = 3.74818$

$y = 3.74818 + 2.901147x$

(ii) $\hat{\sigma}^2 = \frac{1}{10}\left[S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right]$

$= 387.0745$

$-2.228 \le \frac{\hat{\beta} - \beta}{\sigma^2 \sqrt{S_{xx}}} \le 2.228$

$\hat{\beta} - \frac{2.228\,\sigma^2}{\sqrt{S_{xx}}} \le \beta \le \hat{\beta} + \frac{2.288\,\sigma^2}{\sqrt{S_{xx}}}$

$(0.377, 5.425)$

(III) $\hat{\mu}_0 = 119.794$

$se(\hat{\mu}_0) = 10.5989$

95% confidence interval of μ_0

$\Rightarrow 119.794 \pm 2.228(10.598)$

$\Rightarrow 119.794 \pm 23.6144$

$\Rightarrow (96.17956, 143.4084)$

(3) (I) comments on plot:

- The centres of distribution differ for all four cities. Thus theres a Prima facie case for suggesting that underlying means are different.
- The difference between mean time taken to communicate to office in peak hours & non peak hours are in order city A, city D [highest, lowest].

(II) Assumptions

(•) the population must be normal

(•) The population must have a common variance.

(•) The observations are independent

(III) H_0 : The mean difference is same for each city.

H_0 : The mean difference are not same

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

$$SS_B = \frac{1}{7}\left(64^2 + 44^2 + 8^2 + (1-13)^2\right) - \frac{103^2}{28} = 51.611$$

$$SS_R = SS_T - SS_B = 600.00$$

ANNOVA

Source	df	SS	MS	F
treatments	3	516.11	172.04	6.88
Residual	24	600.00	25.00	
Total	27	1116.11		

Under H_0 = $F = \frac{172.04}{25} = 6.88$ Using $F_{3,24}$ distributⁿ

The 5% critical pint is 3.009, so we have sufficient to reject Ho.

$\bar{y}_1 > \bar{y}_2 > \bar{y}_3 > \bar{y}_4$; $\hat{\sigma}^2 = \frac{SS_R}{n-k} = 2$

$C_{24,\,0.025} * \hat{\sigma} \sqrt{\left(\frac{1}{7}+\frac{1}{7}\right)} = 2.064 * \sqrt{2}$

$=$

$\bar{y}_1 - \bar{y}_2 = 2.85$, $\bar{y}_2 - \bar{y}_3 = 5.15$, $\bar{y}_3 -$

observing that all 3 differences less than 5.52, we can that they have no significan difference.

Examining $\bar{y}_1 - y_3 = 8$

hence there's a significant diff means 1 4 3.

PAGE No.
DATE / /

(iii)

(4)

(1) Already in Sol (3), (ii)

(ii) $H_0 : H_1 = H_2 \dots H_5$
H_a : atleast one of the mean is not same

$$\frac{T^2}{N} = \frac{(\Sigma y_1 + \Sigma y_2 + \Sigma y_3 + \Sigma y_4 + \Sigma y_5)^2}{N} = 104385.9394$$

$$SS_{Reg} = \frac{(354)^2}{7} + \frac{(386)^2}{8} + \frac{(87)^2}{6} + \frac{(645)^2}{7} + \frac{(384)^2}{5} - \frac{T^2}{N}$$
$$= 22325.68917$$

$$SS_{TOT} = 69930.0606$$

$$SS_{RES} = SS_{TOT} - SS_{REG}$$
$$= 47604.37143$$

ANOVA

Sources of variation	df	SS	MS	F
Reg	4	22325.69	5581.4225	
Res	28	47604.3143	1700.154	3.28
Total	33	69930.066		

PAGE No.

PAGE No.
DATE / /

So, $F_{4,28}$ tabulated is : 2·21
and our calculated value is 3·28
So, we have sufficient evidences to reject Ho.

(C)

Ho: They are independent of each other

Expected values

Ha: They are dependent

paper cleared	3-5	5-8	8-10	10-12
0-3	27·42	23·089	14·43	11·06
4-6	15·15	12·76	7·97	6·14
7-9	14·43	12·15	7·59	5·82

$$\chi^2_{cal} = \frac{(E_i - O_i)^2}{E_i}$$

$\Rightarrow$ 11·27 + 0·4133 + 4·925 + 3·3204 + 4·384 + 4·1079 + 0·133 + 0·00319 + 6·162 + 1·4175 + 7·234 + 6·56

= 39·929

Since $\chi^2_{cal} > \chi^2_{Tab}$; hence we have sufficient evidences to reject Ho.

⑤ 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 (i) correlation coefficient

$S_{yy} = 8923.6$
$S_{xx} = 54.9$
$S_{xy} = 661.2$

$$r = \frac{S_{xy}}{\sqrt{S_{yy} * S_{xx}}} = 0.94466$$

(ii) $\hat{\beta} = 12.0437$
$\hat{\alpha} = 9.22957$
$y = 9.22957 + 12.0437\,x$

PAGE No.
DATE / /

(iii) $SS_{RES} = \left[S_{yy} - \frac{S_{xy}^2}{S_{xx}} \right] = 960.2951$

$S_{TOT} = 8923.6$

$S_{REG} = 7963.305$

(iv) $R^2 = \frac{SS_{REG}}{SS_{TOT}} = \frac{7963.305}{8923.6} = 0.89148$

Hence: 89% of variation is shown by our model

7. $S_{xx} = \Sigma x^2 - n(\bar{x})^2$
$= 4789.4221$

$S_{yy} = \Sigma y^2 - n(\bar{y})^2 = 1189.2077$

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

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

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 10.9056$

so required eqn = $y = 10.79056 + 0.45451x$

PAGE No. DATE / /

(II) $H_0 : \beta = 0$ $H_a : \beta \neq 0$ $\hat{\sigma}^2 = 22.20136$

$$Z_c = \frac{\hat{\beta} - \beta}{\hat{\sigma}/\sqrt{S_{xx}}} = \frac{0.45451}{\sqrt{22.20136/4789.4221}}$$

$$Z_c = 6.67567$$

$$Z_{tab} = 2.262$$

since $= Z_{cal} > Z_{tab}$

Hence, we have insufficient evidence to reject H_0

(III) $r = \cancel{seq}\ \frac{S_{xy}}{S_{xx} \cdot S_{yy}} = \frac{2176.84}{\sqrt{4789.4221 \times 1189.2077}}$

$$= 0.91213$$

(IV) given : $x^3 = 25 \Rightarrow 2.924$

wow $= 1$

For mean response

$$\hat{H}_0 = 10.79 + 0.4545(2.924)$$
$$\hat{H}_0 = 12.11773$$

$$V(\hat{H}_0) = \hat{\sigma}^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right]$$

PAGE No.

DATE / /

$$= \underline{22.20136} \left[\frac{1}{11} + \frac{(2.724 - 15.83)^2}{4789.4221} \right]$$

$$\boxed{Val(\hat{H}_0) = 2.790417}$$

95% CI :

$$\hat{H}_0 \pm 2.262 \sqrt{2.790417}$$

$$\Rightarrow 12.11773 \pm 2.262 \sqrt{2.790417} \Rightarrow (8.3392, 15.8965$$

for indivisual response

$$\hat{y}_0 = 12.11773$$

$$v(\hat{y}_0) = 24.991777$$

95% CI :

$$\hat{y}_0 \pm 2.262 \sqrt{24.991777}$$

$$\Rightarrow 12.11773 \pm 2.262 \sqrt{24.991777}$$

$$= (0.8096, 23.42587)$$ soln

(IV) since there is some pattern hence, we can't say it exibits normal distribution.

PAGE No.
DATE / /

~~(iii)~~

8: ~~9~~ (i) The main purpose of factor analysis is to reduce many indivisual item into a fewer number of dimensions principle component analysis is a method to decrease dimensions without data loss.

(ii)

PC	Diagnol Entry	Pci/sum (PCi)
PC1	0.456	65%
PC2	0.137	18.5%
PC3	0.08	11.4%
PC4	0.0165	2.4%
PC5	0.72	1.7%
	0.1015	100%

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

PAGE No.	
DATE	/ /

9. (i) → Appears to be negatively correlated.

(ii) $S_{xx} = 75$, $S_{xy} = -296$

$S_{yy} = 2482.4$

$$r = \frac{-296}{\sqrt{2482.4(75)}} = -0.686$$

This shows ~~ne~~ (-ve) ~~correlatn~~ coefficient.

(iii) $H_0 : \rho = 0$ $H_a : \rho < 0$.

$$\omega = \tanh^{-1}(r) = -0.84036$$

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

$$Z_c = \frac{-0.84036 - 0}{\sqrt{\frac{1}{7}}} \Rightarrow -2.2234$$

$Z_{tab} = -1.96$

Since = $Z_c < Z_{tab}$

hence we reject null hypothesis

10 (iv) $\hat{\beta} = \dfrac{S_{xy}}{S_{xx}} = -3.94667$

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

$y = 8216 - 3.94667x$.

(v) ~~(iv)~~ $R = r^2 = 0.470596$

so, 47% of variation is shown by our model,

so, when $x = 1$

$y = 82.16 - 3.94667$
$= 78.21333$

when $x = 2$

$y = 74.26666$

for every additional 1 hour spent, expeted change is

$\boxed{-3.94667}$

PAGE No.
DATE / /

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

$SS_{TOT} = 0.00126$

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

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

ANNOVA TABLE

Source	df	SS	MSS	F
Regression	1	0.00115	0.00115	10.454
Residual	1	0.00011	0.00011	
Total	2	0.00126		

$H_0 : \hat{\beta} = 0$
$H_1 = \beta \neq 0$
$F_{1,1}\ 5\% = 161.42$

$F_{cal} < F_{tab}$

∴ Hence will not reject H_0.