Date __/__/____

Q1) $\Sigma xy = 182560$ $\Sigma y = 1737$

$\Sigma x = 810$ $\Sigma x^2 = 92400$

$S_{xy} = 34915$

$S_{xx} = 20250$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 1.72418$

$\bar{y} = 173.7$ $\hat{\alpha} = 27.14317$

$\hat{y} = 27.14317 + 1.72244\, x$

Q2) x = Share Price Dell y = Share price of IBM

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

$= 875.16$

$S_{xx} = 301.66$

$\beta = \frac{875.16}{301.66} = 2.90115$

$\hat{\alpha} = 3.74804$

95% conf for $\beta = \hat{\beta} \pm t_{10,25} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

$S_{yy} = 6409.7125$

$\hat{\sigma}^2 = \frac{1}{10}\left(6409.7125 - \frac{S_{xy}^2}{S_{xx}}\right) = 387.07477$

$\beta = (2.90115 \pm 2.52379)$

$= (0.37736, 5.42494)$

Page No.

Date __ /__ /____

iii) mean IBN = 40

$$se(\hat{y}_0) = \sqrt{\left(\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)\hat{\sigma}^2}$$

$= 10.59844$

95% conf Inter = $y_0 \pm t_{10, 2.5}\, se(\hat{y}_0)$

$= 115.34065 \pm 23.61443$

$= (86.17966, 143.4812)$

Q3) Assumption
Plot must be normal
Population have common valuey
The observations are independent

iii) H_0: Mean of Difference is same
H_1: The mean of it is not same

$SS_{TOT} = S_{yy}$

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

$1495 - \frac{(103)^2}{28}$

$SS_{TOT} = 1116.11$

$SS_{Rth} = \frac{1}{7}\left(64^2 + 44^2 + 8^2 (-73)^2\right) - \frac{103^2}{28}$

$= 576.11$

$SS_{Res} = SS_{TOT} - SS_{Res} = 60$

Date ___/___/_____

Anova

Under H_0 $\quad f = \frac{MSS\,Reg}{MSS\,Res} = \frac{172.04}{25} = 6.88$

Test statistic $= 6.88$

$f_{3,24,\,5\%} = 3.009$

Hence we have sufficient evidence to reject H_0 at 5% LOS. Thus tablets are unlikely difference b/w.

ii) $\bar{y}_1 = 9.14$

$\bar{y}_2 = 6.289$ $\qquad \bar{y}_{1*} > \bar{y}_{2*} > \bar{y}_{3*} > \bar{y}_{4*}$

$\bar{y}_3 = 1.14$

$\bar{y}_4 = -1.86$ $\qquad \hat{\sigma}^2 = \frac{SSRes}{n-k} = 25$

Least significance differ b/w any of mean is

$t_{24,\,0.025} \times \hat{\sigma} \sqrt{\frac{1}{7} + \frac{1}{7}} = 2.064 \times \sqrt{25} \times \sqrt{\frac{1}{7} + \frac{1}{7}}$

$= 5.52$

$\bar{y}_{1*} - \bar{y}_{2*} = 2.85$

$\bar{y}_{2*} - \bar{y}_{3*} = 5.15$

$\bar{y}_{3*} - \bar{y}_3 = 3$

$\bar{y}_{1*} > \bar{y}_{2*} > \bar{y}_{3*} > \bar{y}_{4*}$

Date __/__/____

Q4) $y_{ij} = \mu + \tau_i + e_{ij}$ $\quad i = 1, 2, \longrightarrow k$

$j = 1, 2, \longrightarrow n_i$

K = no. of treatments

n_i = no. of responses

y_i = Response from i^{th} treatment

Assumption e_{ij} is iid $N(0, \sigma^2)$

b) $n_1 = 7$

$n_2 = 8$ $\quad H_0$: Mean claim amounts of five companies are equal

$n_3 = 6$

$n_4 = 7$ $\quad H_1$: Mean claim amount of the companies are not equal

$n_5 = 5$

$n = 33$

$$\sum_{i=1}^{5} \sum_{j=1} y_{ij} = 1856$$

$$\sum_{i=1} \sum_{j=1}^{n_i} y_{ij}^2 = 174310$$

$$CF = \left(\sum \sum y_{ij}\right)^2 / n = 104385.94$$

SSTOT = 69930.06

SSREG = 227825.69

SSRes = 64604.37

Test Stat = $\dfrac{MSSReg}{MSSRes}$ = 3.283

$F_{4, 28, 0.05}$ = 2.714

We have sufficient evidence to reject H_0

Page No.

Page No.

c) Observed Frequency (O_i)

	3-5	5-8	8-10	10-12	
0-3	45	20	6	6	76
4-6	7	20	3	5	42
4-9	5	8	15	12	40
	57	48	30	23	188

Expected frequency

	3-5	5-8	8-10	10-12
0-3	27.41	23.08	14.43	11.07
4-6	15.15	12.75	7.57	6.11
7-9	14.43	12.15	7.54	5.18

Contigency table

H_0: Sales are independent no. of acting p
H_1: Sales are dependent on the actual p

Test stat $\sum \frac{(O_i - E_i)^2}{E_i}$

Total stat $= 49.91148$

DOF $= (r-1)(c-1) = 2\times3 = 6$

$\chi^2_{0.05} = 12.59$

Here, we have sufficient evident to reject H_0

Salaries are depended on no. of actual perf

Saathi

Date __/__/__

Q5) The 4 samp variances are very different from each other.

ii) mean at rate $= \frac{\sqrt{24} + \sqrt{13} + \sqrt{4}}{3} = 4.52443$

Var at Rat 2 $= \frac{\Sigma(x_i - \bar{x})^2}{4-1} = 0.12127$

for lary [m]

iii) $y_{ij} = \mu_i + t_j + e_{ij} \quad i = 1,2,3,4 \quad j = 1,2,3$

y_{ij} lan transformed value of jth observation with overall popl mean

τ_i is the deviation of the ith mean such

e_{ij} are the independent error terms

$e_{ij} \sim N(0, \sigma^2)$

for Anova

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

$H_1 = \tau_i \neq 0 \qquad$ for atleast one i

ln(n)

66

Rate	mean	Variance	y_i^2
1	2.992	0.163	80.582
2	4.827	0.121	209.678
3	5.716	0.186	294.053
4	6.575	0.—	389.060
Total	20.110	0.618	973.373

Page No.

Page No.

Date ___ / ___ / ______

$$\text{Here } y_i^2 = \left(\sum_1^3 y_{ij}\right)^2 = (3 \times \text{mean}_{T})^2$$

$$SS_{Res} = \sum \frac{y_i^2}{3} - \frac{y^2}{12} = \frac{973.373}{3} - \frac{(3 \times 2 \ldots 1.4)^2}{12}$$

$$SS_{Rate} = 21.153$$

$$SS_{Res} = \sum_{i=1}^{4} \left\{ \sum_{1}^{3} \left(y_{ij} - \bar{y}_i \right)^2 \right\}$$

$$\sum_1^1 2 \times \text{Variance} = 2 \times 0.618 = 1.236$$

Source of Variation	d.f	SS	MSS
Rate	3	21.153	—
Residual	8	1.236	0.15
Total	11	22.21	

Test stat = 45.638

$f_{3,8,17\ldots} = 7.957$

Q6)

x	5	9	2	3	1	1	6	5	4
y	73	120	34	46	24	26	93	45	66

$n = 10$

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

$S_{xy} = 661.2$

$S_{xx} = 54.9$

$S_{yy} = 8923.6$

$r = 0.94466$

Saathi

Date __ / __ / __

Fitted regression eq^n

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

$$\alpha = 9.2295$$

$$\hat{\beta} = \frac{Sxy}{Sxx} = 1.204371$$

$$\hat{y} = 9.2295 + 1.204371 x$$

$$SS_{TOT} = Syy = 8923.6$$

$$SS_{reg} = \frac{S^2xy}{Sxx} = 7963.30492$$

$$SS_{res} = SS_{TOT} - SS_{reg} = 960.29508$$

$$R^2 = \frac{S^2xy}{Sxx\, Syy} = 89.238703$$

Comment

Model is good fit as the model explains 89% of the variability in the output. Better models can be considered carrying more ...

Q7) Sxx 4780

Sxy 2177.4221

$Syy = 1189.207$

$$\hat{\beta} = 0.45484$$

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

$$= 10.78088$$

$$\hat{y} = 10.78087 + 0.454812 x$$

ii) $H_0: \beta = 0$

$H_1: \beta \neq 0$

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{Sxx}}} \sim t_{n-2}$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\left(Syy - \frac{S^2xy}{Sxx}\right) = 22.20136$$

Page No.

Saathi

Date ___/___/___

Test stat $\frac{\hat\beta - 0}{\sqrt{\hat\sigma^2/S_{xx}}} = 6.67568 =$

$t_{9, 2.5} = 2.262$

Hence we have sufficient evidence to reject H_0 $\beta \neq 0$
Hence a linear relation exist

Assumption:
Population have common variance
Population follows the normal distribution
Deviation are independent

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

iv) For Mean response

$x_0 = 25$

$\hat{y_0} = \hat\alpha + \hat\beta x_0 = 22.1524$

$se(\hat{y_0}) \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \hat\sigma^2} = 1.55781$

$\frac{\hat{y_0} - y_0}{se(y_0)} \sim t_9$

95% CI for $\hat{y_0}$ $= y_0 \pm t_{9, 2.5}\, se(\hat{y_0})$
$= (18.84313, 25.66349)$

Individual response

$(10.932, 33.37462)$

Page No.

Date __/__/__

vii) There is clearly some relationship b/w the residuals and the explanatory variable values. Hence our assumption of normality might be wrong

Q8)

Principle component	Diagonal/ Eigen value	PC/Cum
PC_1	0.436	65%
PC_2	0.131	19.5%
PC_3	0.08	11.4%
PC_4	0.0165	2.4%
PC_5	0.012	1.7%
		100%

iii) As 1st 3 principle component explain over 95% of total variance dimension can be reduced to 3 per this data. The 3rd principle components can be used for building further classification regression modelling purp

Q9) i) There appear to be a negative correlation b/w the marks in and hrs spent per day on social media

$S_{xx} = 75$

$S_{yy} = 2482.4$

$S_{xy} = -296$ $\qquad r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = -0.686$

Hence they have -ve weak corr

Page No.

H_0 $\rho = 0$
H_1 $\rho < 1$

$\tanh^{-1} R - \tanh^{-1}(\rho) \sim N\left(0, \frac{1}{n-3}\right)$

$z_{cal} = -2.07893$
$z_{ty.} = 1.6490$
$|z|_{cal} > z_{tab}$
We have sufficient evidence to reject H_0
Hence $\rho < 0$

iv) $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = -3.94667$ $\hat{\alpha} = \hat{y} - \hat{\beta} x$
$= 82.16$
$\hat{y} = 82.16 - 3.94 x$

v) $R^2 = 47.059832\cancel{90}$

Hence model explains 47% of variation

Model not a good fit

vi) $x_0 = 1$
$\hat{y} = \hat{\alpha} - \hat{\beta} x_0$ $\hat{y}_0 = 78.2133$

d(a)

x	0.27	0.36	0.3
y	0.05	0.1	0.02

$S_{xx} = 0.0042$
$S_{yy} = 0.00120$
$S_{xy} = 0.0022$
$SS_{tot} = 0.00126$ $SS_{reg} = 0.00115$ $SS_{res} = 0.00011$

Date __ /__ /____

Anova Table

Reg	1	0.00115
Res	1	0.0011
Tr	2	0.00126

$H_0 = \beta = 0$

$H_1\ \beta \neq 0$

Test stat $\frac{MSR}{MSE} = 10.4$

$F_{1,\ 5\%} = 161.4$

Fcal < Ftab

Do not reject H_0