Nimitt Chauhan
PAGE No.
DATE Roll/no.- 11

Probability & Statistics-II
Assignment-2

Q1]
(a) $\hat{y}_i = \hat{\alpha}_i + \hat{\beta} x_i$

(b) Gradient of line $= \hat{\beta}$

$\therefore \hat{\beta} = 1.72419$

Q2] $S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 301.66$

$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 6409.7125$

$S_{xy} = \sum x_i y_i - \cancel{(\sum x_i y_i)^2} \frac{\sum x_i \sum y_i}{n} = 875.16$

(i) $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{875.16}{301.66}$

$\boxed{\hat{\beta} = 2.9011}$

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

$\hat{\alpha} = 3.7482$

$\therefore \hat{y} = 3.7482 + 2.9011 x$

(ii) $PQ = \frac{\hat{\beta} - \beta}{S_c(\hat{\beta})} \sim t_{n-2}$

$\therefore S_c(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \cdot \sqrt{\frac{\frac{1}{10}\left(6409.7125\right) - \frac{875.16}{301.66}}{301.66}}$

$S_c(\hat{\beta}) = 1.132.$

95% CI for $\hat{\beta}$ = $\left(-2.228(1.132)+2.901,\ 2.228(1.132)+2.901\right)$

$$\boxed{CI = (0.3789,\ 5.4231)}$$

(iii) $PQ = \dfrac{\hat{\mu} - \mu_i}{Se(\hat{\mu})} \sim t_{n-2}$

$\therefore \hat{\mu} = 3.7482 + 2.9011(40)$

$\hat{\mu} = 119.79$

$Se(\mu) = \sqrt{\left(\frac{1}{12} + \frac{79^2}{301.66}\right) 387.07}$

$Se(\mu) = 10.5988$

95% CI -

$CI = (119.79 - 2.228(10.5988),\ 119.79 + 2.228(10.5988))$

$$\boxed{CI = (96.17,\ 143.40)}$$

Q4.

(a) one-way ANOVA model is given by -

$$y_{ij} = \mu + \tau_i + e_{ij}$$

where

$i = 1, 2, 3, \ldots, k$ ← no. of treatments

$j = 1, 2, 3, \ldots m$ ← no. of responses

~~$y_{ij} =$~~

y_{ij} = jth response from ith treatment

τ_i = ith treatment

μ = mean response

e_{ij} = error term.

(b) $H_0 \Rightarrow$ mean claim is equal

$H_1 \Rightarrow$ mean claim is not equal

$n_1 = 7$, $n_2 = 8$, $n_3 = 6$,

$n_4 = 7$, $n_5 = 5$, $\therefore n = 33$

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

$$\therefore \sum_{i=1}^{5} \sum_{j=1}^{n} y_{ij}^2 = 174316$$

$$CF = \left(\sum_{i=1}^{5} \sum_{j=1}^{n} y_{ij} \right)^2 / n$$

$$CF = 104385.94$$

$$\therefore SS_{TOT} = 174316 - 104385.94$$

$$SS_{TOT} = 69930.06$$

$$SS_{REG} = \frac{354^2}{7} + \frac{386^2}{8} + \frac{87^2}{6} + \frac{643^2}{7} + \frac{384^2}{3} - 104385.94$$

$SS_{REG} = 22325.69.$

$SS_{RES} = SS_{TOT} - SS_{REG}$
$= 47604.37$

Source of Variation	df	sum of sqr	mean sum of sqr
Regression	4	22326	5581
Residual	28	47604	1700
Total	32	69930	

Test statistic $= \frac{MSS_{REG}}{MSS_{RES}} = \frac{5581}{1700}$

Test statistic $= 3.283$

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

$\therefore$ we will reject H_0. [$\because$ TS > F value]

(c) O_i –

	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

E_i –

	3-5	5-8	8-10	10-12
0-3	27.41772	23.08861	14.43038	11.06329
4-6	15.15189	12.75949	7.97468	6.11392
7-9	14.43038	12.15189	7.59494	5.82278
				158

$H_0 \Rightarrow$ Sales are independant
$H_1 \Rightarrow$ Sales are dependant.

$$T.S. = \frac{\Sigma(O_i - E_i)^2}{E_i} = 49.91146$$

degree of freedom $= 2 \times 3$
$= 6.$

$\therefore \chi^2_{6, 5\%} = 12.59$

$\therefore$ We reject H_0. $[\because TS > \chi^2 \text{ value}]$

Q5.]

(I) Assumptions –

→ Poplulations are normal
→ Observations are independant.
→ Population have common variance.

(II) mean at rate 1 $= \dfrac{\sqrt{29} + \sqrt{13} + \sqrt{21}}{3} = 4.5244$

var. at rate 2 $= \dfrac{\Sigma(x_i - \bar{x})^2}{n} = 0.1213.$

(III) The test was correct in telling us that the transformation must be done before One-way ~~test~~ ANOVA test.

(IV) (a) $y_{ij} = \mu + \tau_{ij} + e_{ij}$

→ y is the $\log_e$ transformed value for j^{th} obs. of no. of germination per sq. foot. observed when i^{th} rate was applied.

→ μ is population mean

→ τ is deviation ; $\Sigma \tau = 0$

→ e is independant error term

(b) For ANOVA -

$H_0 \Rightarrow \tau = 0$, $H_1 \Rightarrow \tau \neq 0$

ln -

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	0.148	389.063
	20.110	0.618	973.373

$$SS_{RATE} = \sum_{i=1}^{4} \frac{Y_i^2}{3} - \frac{Y_{.}^2}{12} = 21.153$$

$$SS_{RES} = \sum_{i=1}^{4} \left(\sum_{j=1}^{3} (Y_{ij} - \bar{Y}_i)^2 \right)$$
$$= 1.236.$$

Source of Variation	df	SS	MSS
Rate	3	21.153	7.051
Residual	8	1.236	0.155
	11	22.389	

$$TS = \frac{SS_{Rate}}{SS_{RES}} = \frac{7.051}{0.155}$$

$$T.S. = 45.688$$

$$F_{3,8,1\%} = 7.951$$

$\therefore$ we reject H_0.

Q6.

$\Sigma x = 39$, $\Sigma x^2 = 307$, $\Sigma y = 562$, $\Sigma y^2 = 40508$
$\Sigma xy = 2853$, $n = 10$

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 207 - \frac{39^2}{10} = 54.9$$

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 40508 - \frac{562^2}{10} = 8923.6$$

$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 2853 - \frac{39 \times 562}{10} = 661.2$$

(i)

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{661.2}{\sqrt{54.9 \times 8923.6}}$$

$$\boxed{r = 0.9447}$$

(II) $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{661.2}{54.9}$

$\boxed{\hat{\beta} = 12.0437}$

$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x} = \frac{\Sigma y}{n} - 12.0437 \cdot \left(\frac{\Sigma x}{n}\right)$

$\boxed{\hat{\alpha} = 92295}$

(III) $SS_{TOT} = S_{yy} = 8923.6$

$SS_{REG} = \frac{S^2_{xy}}{S_{xx}} = \frac{661.2^2}{54.9} = 7963.3049$

$SS_{RES} = SS_{TOT} - SS_{REG} = 8923.6 - 7963.3049$

$SS_{RES} = 906.2951$

(IV) $r^2 = \frac{S^2_{xy}}{S_{xx} S_{yy}} = \frac{661.2^2}{54.9 \times 8923.6}$

$r^2 = 89.24\ \%$

∴ it is a good fit since it explains almost 90% of output variance.

Q7] $\Sigma x = 174.13$, $\Sigma y = 197.84$, $\Sigma x^2 = 7545.9$,
$\Sigma y^2 = 4747.45$, $\Sigma(x-\bar{x})(y-\bar{y}) = 2176.84$, $n = 11$

$$S_{xx} = 7545.9 - \frac{(174.13)^2}{11} = 4789.4221$$

$$S_{yy} = 4747.45 - \frac{(97.84)^2}{11} = 1189.2077$$

$$S_{xy} = 2176.84$$

(i) $$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{2176.84}{4789.4221} = 0.4545$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 10.7806$$

$$\therefore \boxed{\hat{Y} = 10.7806 + 0.4545x}$$

(ii) $H_0 \Rightarrow \beta = 0$, $H_1 \Rightarrow \beta \neq 0$

$$T.S. = \frac{\hat{\beta} - \beta}{se(\hat{\beta})} \sim t_{n-2}$$

$$\hat{\sigma}^2 = \frac{1}{11-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = 22.2014$$

$$Se(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = 0.06808$$

$$T.S. = \frac{0.4545}{0.06808} \sim t_{n-2} = t_{9,5\%}$$

$$\therefore 6.6759 \sim 2.262$$

$\therefore$ we reject H_0.

(iii) $r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = 0.9121$

(iv) $x_0 = 25$

$\mu_0 = PQ = \frac{\hat{\mu} - \mu}{se(\hat{\mu})} \sim t_{n-2}$

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

$= 1.55$

$95\%\ CI = \hat{\mu} \pm t_{9, 2.5\%} \times se(\hat{\mu})$

$CI = (18.64313,\ 25.66349)$

for individual ~~data~~ response -

$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} x_0$

$= 22.15331$

$Se(\hat{Y}_0) = 4.9608$

$95\%\ CI = \hat{Y} \pm t_{9, 2.5\%} \times SE(\hat{Y}_0)$

$CI = 22.15331 \pm 11.22131$

$CI = (10.932,\ 33.37462)$

(v) $\therefore$ We can see relation between residual, and explanatory variables. $\therefore$ Model is not a good fit for the data.

Q8] (i) Main purpose is to reduce the many individual items into fewer dimensions. It is used to simplify data.

Principle component analysis is a method to reduce dimensions of data with increasing interpretation.

(ii)

Principal Component	Diagonal entries	$PC_i / \Sigma PC_i$
PC1	0.456	65%
PC2	0.137	19.5%
PC3	0.080	11.4%
PC4	0.0165	2.4%
PC~~8~~5	0.012	1.7%
	0.7015	

(iii) AS the first ~~8~~ 3 PC_i covers 95% variance, demonstrating can be reduced to the data set.

Q9.] $\Sigma x = 45$, $\Sigma x^2 = 277.5$, $\Sigma y = 644$, $\Sigma y^2 = 43956$
$\Sigma xy = 2602$, $n = 10$

(i) There seems to be negative correlation between marks and hours spent of social media.

(ii) ~~Sa~~ $S_{xx} = 75$
$S_{yy} = 2482.4$
$S_{xy} = -296.$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{xy}}} = -0.686002$$

$\therefore$ weak negative correlation exists.

(III) $H_0 \Rightarrow \rho = 0$, $H_1 \Rightarrow \rho \leq 0$

$\tanh^{-1}(r) - \tanh^{-1}(\rho) \sim N(0,1)$

$$\therefore T.S. = \frac{\tanh^{-1}(r) - \tanh^{-1}(\rho)}{\sqrt{1/n-3}} \sim N(0,1)$$

$$= -2.0789$$

$z_{cal} = -2.0789,$

$z_{tab} = 1.6448$

$\because |z_{cal}| > z_{tab}.$

we reject Ho

$\therefore \rho < 0$

(IV) $$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = -3.9467$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 82.16$$

$$\therefore \boxed{\hat{Y} = 82.16 - 3.9467x}$$

(V) $$R^2 = \frac{S^2_{xy}}{S_{xx} \cdot S_{yy}} = 47.0598\%$$

$\therefore$ Model is not a good fit as it explains only 47% data.

(vi) $X_0 = 1$

expected change $= \hat{\beta} x_0 = -3.9467$

Q10. $\Sigma X = 0.93$, $\Sigma X^2 = 0.2925$, $n = 3$

$\Sigma Y = 0.23$, $\Sigma Y^2 = 0.0189$, $\Sigma XY = 0.0735$

$S_{xx} = 0.0042$

$S_{yy} = 0.00126$

$S_{xy} = 0.0022$

$SS_{TOT} = S_{yy} = 0.00126$

$SS_{REG} = \dfrac{S_{xy}^2}{S_{xx}} = 0.00115$

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

$SS_{RES} = 0.00011$

ANOVA Table

Source of Variance	d.f.	Sum of Squares	mean SS
Regression	1	0.00115	0.00115
Residual	1	0.00011	0.00011
total	2	0.00126	

$H_0 \Rightarrow \beta = 0$, $H_1 \Rightarrow \beta \neq 0$

$T.S. = \dfrac{0.00115}{0.00011} \sim F_{1,1}$

$= 10.4545 \sim F_{1,1}$

DATE / /

$F_{cal} = 10.4545$

$F_{tab} = 161.4$

$\therefore F_{cal} < F_{tab}$

$\therefore$ we doo not reject H_0.