PLS

Assignment 2

Q1.

(a) $\Sigma xy = 182560$

$\Sigma x^2 = 92800$

$\Sigma x = 850$

$\Sigma y = 1737$

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

$= 34915$

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

$= 20250$

$\hat{B} = \frac{S_{xy}}{S_{xx}}$

$= 1.724198$

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

$\bar{y} = 173.7$

$\bar{x} = 85$

$\hat{\alpha} = 27.14317$

$\hat{y} = 27.14317 + 1.72419x$

classmate

Date
Page

Q2 X = Share price of Dell
Y = Share price of IBM

(i) $S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$
$= 875.16$

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

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 2.90115$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$
$= 3.74809$

$\hat{y} = 3.74809 + 2.90115x$

(ii) ASSUMPTIONS:
- distributions of X & Y are normal
- variables are independent

(iii) $\frac{\hat{\beta} - \beta}{\sqrt{\frac{s^2}{S_{xx}}}} \sim t_{n-2}$

95% CI for $\hat{\beta}$ = $\hat{\beta} \pm t_{10, 2.5\%} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

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

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

$$= 6409.7125$$

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

$$\hat{\sigma}^2 = 387.07447$$

95% CI for $\hat{\beta}$ = $\hat{\beta} \pm t_{10, 2.5\%} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

$$= 2.90115 \pm 2.52379$$

95% CI = [0.37736, 5.42484]

(iii) IBM Share price = μ_0

$$\frac{\hat{\mu}_0 - \mu_0}{sd(\hat{\mu}_0)} \sim t_{n-2}$$

$x_0 = 40$

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

$$\hat{\mu}_0 = 119.79409$$

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

$= \underline{10.59894}$

95% CI for $\mu_0 = \hat{\mu}_0 \pm t_{10, 2.5\%}\, Se(\hat{\mu}_0)$

$= 119.79408 \pm 23.61443$

95% CI for $\hat{\mu}_0 = [96.17966, 143.40812]$

Q3

(ii) Assumptions under ANOVA:

- populations must be normal
- populations have a common variance
- observations are independent.

(iii) H_0: mean of differences is same
H_1: mean of differences is not same

$SS_{Tot} = S_{yy}$

$= \sum y^2 - \frac{(\sum y)^2}{n}$

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

$= \underline{1116.11}$

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

$= 516.11$

$SS_{RES} = 600$

ANOVA

Source	DOF	SOS	MOS
Regressor	3	516.11	172.04
Residuals	24	600	25
TOTAL	27	1116.11	

Under H_0, $F = \frac{MSS_{REG}}{MSS_{RES}} = \frac{172.04}{25} = 6.88$

Test statistic = 6.88

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

(iv) Analysis of mean differences

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

4.

(a) $Y_{ij} = \mu = T_j + e_{ij}$, $i = 1, 2, \dots, k$
$j = 1, 2, \dots, n_j$

where,

- k = no. of treatments
- T_i is the jth response for ith treatment
- n_j = no. of responses
- μ is the mean response
- e_{ij} is the error term

ASSUMPTION : e_{ij} is iid

(b) H_0: mean claim amounts are equal
H_1: mean claim amounts are unequal

$n_1 = 7$, $n_2 = 8$, $n_3 = 6$, $n_4 = 7$
$n_5 = 5$, $n = 33$

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

$$\sum_{i=1}^{5} \sum_{j=1}^{n_i} y_{ij}^2 = 174{,}316$$

$$C.F = \left[\sum_{i=1}^{5} \sum_{j=1}^{n_i} y_{ij} \right]^2 / n = 104385.94$$

$$SS_{TOT} = 174316 - C.F$$
$$= 69930.06$$

$$SS_{REG} = \sum_{i=1}^{5} \left[\sum_{i=j}^{n_i} y_{ij}\right]^2 / n_i - C.F.$$

$$= \frac{354^2}{7} + \frac{386^2}{8} + \frac{87^2}{6} + \frac{645^2}{7} + \frac{389^2}{5} - 104380$$

$$= 22325.70$$

$$SS_{RES} = 47604.37$$

ANOVA

Sources	DOF	SS	MSS
Regression	4	22326	5581.42
Residuals	28	47604	1700.14
Total	32	69930	

$$\text{Test Stat} = \frac{MSS_{REG}}{MSS_{Res}}$$

$$= 3.283$$

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

Hence we reject H_0

(c) Observed

Papers	SALARY				TOTAL
	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

Papers	SALARY				TOTAL
	3-5	5-8	8-10	10-12	
0-3	27.41	23.08	14.43	11.06	
4-6	15.15	12.75	7.97	6.11	
7-9	14.43	12.15	7.59	5.81	

Contingency table

H_0 : salaries are independent
H_1 = Salaries are dependent

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

$= 48.91146$

DOF = $(r-1)(c-1)$
$= 6$

$\chi^2_{1, 5\%} = 12.59$

Hence we reject H_0

5.

(i) ASSUMPTIONS

- normal population
- populations have common variance
- observations are independent.

Mean at Rate 1 = $\frac{\sqrt{29} + \sqrt{13} + \sqrt{21}}{3}$

$= \underline{4.5244}3$

Mean at Rate 2 = $\frac{\Sigma(x - \bar{x})^2}{n_1} = 0.1212$

(b) For ANOVA

$H_0 : T_i = 0$, $i = 1, 2, 3, 4$

$H_1 : T_i \neq 0$ for atleast one

(lw)

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	0.48	389.06
TOTAL	20.110	0.618	973.37

Here, $Y_i^2 = \left(\sum_{j=1}^{3} y_{ij}\right)^2 = [3 \times \text{Mean}]^2$

$$SS_{RATE} = \sum_{i=1}^{4} \frac{Y_i^2}{3} - \frac{Y^2}{12} = \frac{973.375}{3} - \frac{(3 \times 20.110)^2}{12}$$

$$SS_{RATE} = 21.153$$

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

$$= 1.236$$

Sources	DOF	SS	MSS
Rate	3	21.153	7.051
residuals	8	1.236	0.155
TOTAL	11	22.38	

$$\text{Test stat} = \frac{MSS\ Rate}{MSS\ Res} = 45.638$$

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

Q6

$n = 10$

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

$S_{xy} = \Sigma xy - n\bar{x}\bar{y}$
$= 661.2$
$S_{xx} = \Sigma x^2 - n\bar{x}^2$
$= 54.9$
$S_{yy} = \Sigma y^2 - n\bar{y}^2$
$= 8723.6$

$r = 0.94466$

(ii) Fitted model

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$
$\hat{\alpha} = 9.2295$

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

~~$\hat{y} = 9.2295 +$~~

$\hat{y} = 9.2295 + 12.04371x$

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

$SS_{REG} = \frac{S^2_{xy}}{S_{xx}} = 7963.30492$

$SS_{RES} = SS_{TOT} - SS_{REG} = 960.27508$

$R^2 = \frac{S^2_{xy}}{S_{xx}\,S_{yy}}$

$= 89.238703\%$

Q7.

(i) linear regression :-

$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n},$

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

$S_{xx} = 4789.4221$

$S_{yy} = 1189.20767$

$S_{xy} = 2176.84$

$$\hat{B} = \frac{S_{xy}}{S_{xx}}$$

$$= 0.45451$$

$$\hat{a} = \bar{y} - \hat{B}\bar{x}$$

$$= 10.79056$$

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

(ii) $H_0 : B = 0$

$H_1 : B \neq 0$

$$\frac{\hat{B} - B}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \sim t_{n-2}$$

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

$$= \frac{1}{4}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$$

$$= 22.20136$$

Test stat $= \frac{\hat{B} - 0}{\sqrt{\hat{\sigma}^2 / S_{xy}}}$

$$= 1.67568$$

$t_{9, 2.5\%} = 2.262$

Hence we will reject H_0
Linear relation exists.

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

$= 0.91213$

(iv) Now μ_0

$x_0 = 25$

$\hat{\mu}_0 = \hat{\gamma} + \hat{\beta} x_0$

$= 22.15331$

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

$= 1.55181$

95% CI for $\mu_0 = \hat{\mu}_0 \pm t_{9, 2.5\%}, \; Se(\hat{\mu}_0)$

$= 22.15331 \pm 3.51018$

95% CI $= [18.64313, 25.66349]$

- Individual Response

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

$= 22.15331$

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

$= 4.96079$

95% CI $y_0 = \hat{y}_0 \pm t_{9, 2.5\%} \times Se(\hat{y}_0)$

$= 22.15331 \pm 11.32131$

$= [10.932, 33.37462]$

Q8.

(ii)

PC	Diagonal	PC / (Sum PC)
1	0.486	65%
2	0.137	18.5%
3	0.08	11.4%
4	0.0165	2.4%
5	0.012	1.7%
	0.1015	100%

Q9. (i) there is a negative correlation

(ii) $Sxx = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$

$= 75$

$Syy = \Sigma y^2 - \frac{(\Sigma y)^3}{n}$

$= 2482.4$

$Sxy = \Sigma xy - (\Sigma x)(\Sigma y)/n$

$= -286$

$r = \frac{Sxy}{\sqrt{Sxx \times Syy}}$

$= 0.0686003$

(iv) $\hat{B} = \frac{S_{xy}}{S_{xx}}$

$= -3.94667$

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

$= 82.16$

$\hat{y} = 82.16 - 3.94467x$

(v) $R^2 = \frac{S^2xy}{S_{xx}\,S_{yy}}$

$= 47.05983\%$

- Model explains 47% of the variability

(vi) $x_0 = 1$

$\hat{\mu}_0 = \hat{\alpha} - \hat{B}x_0$

$= 78.4133$

10.

$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 0.0042$

$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 0.00126$

$S_{xy} = \sum xy - n\bar{x}\bar{y} = 0.0022$

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

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

$SS_{RES} = 0.00014$

ANOVA TABLE

SOURCES	DOF	SS	MSS
Regression	1	0.00115	0.00115
Residuals	1	0.00011	0.00011
	2	0.00126	

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$ (One which)

Test Stat $\frac{MSS_{REG}}{MSS} = 10.9545$

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

$F_{cal} < F_{tab}$, do not reject H_0