PROBABILITY & STATS

Q1) A. $\hat{y_i} = \hat{\alpha_i} + \hat{\beta} x_i$

b). gradient of the regression line $= \hat{\beta} = 1.7241$

Q2) i) $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$

$\hat{\alpha_i} = \bar{y} + \hat{\beta}\bar{x_i}$

$S_{xx} = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n} = 12666.5 - \frac{(385.2)^2}{12}$

$= 301.66$

$S_{yy} = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n} = 119026.9 - \frac{(1162.5)^2}{12} = 6409.7125$

$S_{xy} = \Sigma x_i y_i - \frac{(\Sigma x_i)(\Sigma y_i)}{n} = 38191.41 - \frac{(385.2)(1162.5)}{12}$

$= 875.16$

$\therefore \hat{\alpha_1} = 3.7481 \qquad \hat{\beta} = 2.9011$

$y_i = 3.7481 + 2.9011\, x_i$

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

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

$\hat{\sigma}^2 = \frac{1}{10}\left[6409.7125 - \frac{(875.16)^2}{301.66}\right] = 387.075$

$$\Rightarrow \quad P\left[-2.228 < \frac{2.901-\beta}{1.132} < 2.228\right] = 0.95$$

$$P\left[-2.228(1.132) - 2.901 < -\beta < 2.28(1.132) - 2.901\right]$$

95% CI for $\beta = \{0.378904, 5.423096\}$

(iii) $$PQ = \frac{\hat{\mu}_i - \mu_i}{Se(\hat{\mu}_i)} \sim t_{m-2}.$$

$$\hat{\mu}_i = 3.74818 + 2.901146(40)$$
$$= 119.79$$

$$Se(\mu_i) = \sqrt{\left[\frac{1}{12} + \frac{(7.9)^2}{301.66}\right] 381.07} = 10.5988$$

$$P\left[-2.228 < \frac{119.79 - \mu_i}{10.5988} < 2.228\right] = 0.95$$

95% CI for $\mu_i = \{119.79 - 2.228(10.5988), 119.79 + 2.28(10.5$

$$= \{96.17, 143.40\}$$

classmate

Date
Page

Q3) i) Comment on the plot:-

→ the underlying means are different.

→ the variation in the data C is large as compared to city D.

ii] Underlying Assumptions:-

→ Population follows normal distribution

→ It has a common variance.

→ observation are independent

iii) H_0: mean of difference is same.

H_1: Mean of difference is not same

$$SS_{TOT} = S_{yy} = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}$$

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

$$= 1116.11$$

$$SS_{REG} = \frac{1}{7}\left[(64)^2 + (44)^2 + (8)^2 + 169\right] - \frac{103^2}{28}$$

$$= 516.11$$

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

$$= 600$$

ANOVA table

Source of variation	df	Sum of Squares	Mean
Regression	3	516.11	172.04
Residual	24	600	25
total	27	1116.11	24

under Ho :-

$$f = \frac{MSS_{Reg}}{MSS_{Res}} = \frac{172.04}{25} = 6.88$$

test statistic = 6.88

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

we reject Ho at 5% los

IV] Analysis :-

$\bar{y}_{1*} = 9.14$
$\bar{y}_{2*} = 6.29$
$\bar{y}_{3*} = 1.14$
$\bar{y}_{4*} = -1.86$

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

$$\therefore \hat{\sigma}^2 = \frac{SSRES}{n-2} = 25$$

$\bar{y}_{1*} - \bar{y}_{2*} = 2.85$
$\bar{y}_{2*} - \bar{y}_{3*} = 5.15$
$\bar{y}_{3*} - \bar{y}_{4*} = 3$

$\bar{y}_{1*} > \bar{y}_{2*} > \bar{y}_3 > \bar{y}_{4*}$

Q4) a) $y_{ij} = \mu + \tau_i + e_{ij}$, $i = 1,2,\dots k$
$j = 1,2 \dots n_i$

we assume that e_{ij} are iid variables $\sim N(0,\sigma^2)$

b) H_0: Mean claim amounts of 5 companies are equal
H_1: Mean claim amounts are not equal.

$m_1 = 7$ $\quad m_4 = 7$
$m_2 = 8$ $\quad m_5 = 5$
$m_3 = 6$

$n = 33$

$\sum_{i=1}^{5}\sum_{j=1}^{n_i} y_{ij} = 1856$ $\qquad \sum_{i=1}^{5}\sum_{j=1}^{n_i} = 174316$

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

$= 104385.94$

$SS\,REG = \frac{352^2}{4} + \frac{386^2}{8} + \frac{872}{6} + \frac{643^2}{7} + \frac{384^2}{3} - 10438594$

$= 22325.69$

$SS\,REG = SSTOT - SSREG = 47604.37$

Sources of variation	df	sum of squares	Mean.
Regression	4	22326	5581
Residual	28	47604	1700
total	32	69930	

Test statistic $= \frac{MSS\,REG}{MSS\,RES} = 3.283$

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

we reject Ho

Q7.

Oi	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

Ei	3-5	5-8	8-10	10-12
0-3	27.41772	23.088	14.430	11.06
4-6	15.15189	12.759	7.97	6.11
7-9	14.43038	12.15189	7.59	5.822

contingency table :-

Ho: sales are independent
H1: sales are dependent

$$TS = \frac{\Sigma (O_i - E_i)^2}{E_i} = 49.91146$$

df = 6

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

Thus, we reject Ho

Q5) i) Assumptions for ANOVA
→ population is normal
→ " have common variance.
→ observations are ~~indep~~ iid variables

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

$= 4.52443$

$$\text{Var (Rate 2)} = \frac{\Sigma (x_i - \bar{x})^2}{n-1} = 0.12127.$$

iii) Test was correct.

6) $\Sigma x = 39$ $\Sigma x^2 = 207$ $\Sigma y = 562$ $\Sigma y^2 = 40508$

$\Sigma xy = 2853$, $n = 10$.

i) $S_{xx} = 207 - \frac{39^2}{10} = \boxed{54.9}$.

ii) $S_{yy} = 40508 - \frac{562^2}{10} = \boxed{8923.6}$.

$S_{xy} = 2853 - \frac{562 \times 39}{10} = \cancel{6612}\ \boxed{661.2}$.

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

ii) $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 12.04$.

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

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

~~$SS_{REG} = \frac{S^2_{xy}}{S_{xx}}$~~ $S_{TOT} = S_{yy} = 8923.6$

$SS_{RES} = SS_{TOT} - SS_{REG}$
$= 960.29$

$R^2 = \frac{S^2_{xy}}{\quad} = 89.24\%$

classmate

Date ____

Page ____

Q7) i) $\Sigma x = 174.13$ $\Sigma x^2 = 7545.9$

$\Sigma y = 197.84$ $\Sigma y^2 = 4747.45$

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

$n = 11.$

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

$$= 4789.4221$$

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

$$S_{yy} = 4747.45 - \frac{197.84^2}{11}$$

$$= 1189.20767.$$

$S_{xy} = 2176.84$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 10.780$$

$$\hat{y} = 10.780 + 0.45451\bar{x}$$

ii) $H_0 : \beta = 0$, $H_1 : \beta \neq 0$.

$$TS = \frac{\hat{\beta} - \beta}{Se(\hat{\beta})} \sim t_{n-2}$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S^2_{xy}}{S_{xx}}\right) = 22.2013$$

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

t_9 ~~at~~ $2.5\% = 2.262$

Assumptions :-

→ Population has a common variance

→ " follows $\sim N(0,1)$

→ observations are iid.

iii) $r = \frac{Sxy}{\sqrt{Sxx \cdot Syy}} = 0.91213$

Q8) i) Main purpose of factor analysis is to reduce many individual items into fewer dimensions

→ Principal component Analysis is technique for reducing dimensionality and minimising information loss.

Principal components chosen should be

i) Uncorrelated

ii) linear

iii) max variance

Q9) $\Sigma x = 45$ $\Sigma x^2 = 277.5$ $\Sigma y = 644$

$\Sigma y^2 = 43956$ $\Sigma xy = 2602$

$n = 10$

i) negative correlation.

ii) $Sxx = 277.5 - \frac{45^2}{10} = 75$

$Syy = 2482.4$

$Sxy = -296$

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

As stated above, -ve correlation

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

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

$$\hat{y} = 82.16 - 3.9467X$$

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

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

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

$$S_{xx} = 0.2925 - \frac{0.93^2}{3} = 0.0042$$

$$S_{yy} = 0.0189 - \frac{0.23^2}{3} = 0.00126$$

$$S_{xy} = 0.0022$$

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

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

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

$= 0.00011$

Anova

Var	DOf	SS	MSS
Reg	1	0.00115	0.00115
Res	1	0.00011	0.00011
Tot	2	0.00126	

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

$$TS = \frac{0.00115}{0.00011} \sim f_{1,1}$$

$10.4545 \sim f_{1,1}$

$f_{cal} < f_{tab}$, $\therefore$ we do not reject H_0