SAKSHI TRIPATHI Roll No. 81

Probability and Statistics

Assignment-2

Q.1

x	40	10	100	110	120	180	20	90	80	130
y	56	62	195	240	170	270	48	196	214	286

$\Sigma xy = 1,82,560$

$\Sigma x^2 = 92,00.$

$\Sigma x = 850$

$\Sigma y = 1737$

$S_{xy} = \Sigma xy - \dfrac{\Sigma x \Sigma y}{n} = 34915$

$S_{xx} = \Sigma x^2 - \left(\dfrac{\Sigma x}{n}\right)^2$

$20,250.$

$\hat{\beta} = \dfrac{S_{xy}}{S_{xy}} = 1.724198$

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

$\bar{y} = 173.7$

$\bar{x} = 85$

$\hat{\alpha} = 27.14317$

$\hat{y} = 27.14817 + 1.72419x$

b) The gradient of the regression line refers to its slope parameter β since β is significantly different from 0, there exists a positive linear relationship b/w x and y.

DOMS Page No.
Date / /

Hence, there is a +ve linear relationship b/w the amount withdrawn and the number of hours until the next visit

Q.2 x = share price of Dell
y = share price of IBM.

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

$$= 875.16$$

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

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

$$\hat{\alpha} = 3.74809 + 2.80115\, x$$

ii) Assumptions :-
The distribution for x & y are normal
The explanatory variable values all indep of each other

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

Level of significance - 5%
95% confidence for $\beta = \hat{\beta} \pm t_{0.025, n-2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

Q.3

Comment of the plot:

- The centre of the distribution differs for all the 4 cities thus there is prima facie case for reg suggesting that the underlying means are different
- The variation in the data for city C is lowest compared to city D which appears to be highest. However with only 7 observations for each city, we can't be certain that there is a real underlying difference in variance

ii) Assumption underlying analysis of var

→ The population must be normal
→ The population have a common var
→ The observation are independent

iii) Ho : The mean of differences is same for each city

H_1 : The mean of differences is ~~same~~ ~~for each city~~ not the same

$SS_{TOT} = S_{y-y}$

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

Q4 a) $y_{ij} = \mu + T_i + e_{ij}$, $i = 1, 2, \dots k$
Assumption $j = 1, 2, \dots n_i$
e_{ij}, i.i.d $N(0, \sigma^2)$.

b) H_0 : mean claim amount of 5 companies are equal
H_1 : Mean claim amounts of the companies are not equal

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

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

$$CF = \left(\sum_{i=1}^{5} \sum_{j=1}^{n_i} y_{ij}\right)^2 / n = 104.385-94$$

$$SS_{TOT} = 174316 - CF$$
$$= 69.930-06.$$

$$SS_{Reg} = \sum_{j=1}^{5} \left(\sum_{i,j}^{n_i} y_{ij}\right)^2 / n_i - CF$$

$$= \frac{351^2}{7} + \frac{386^2}{8} + \frac{87^2}{6} + \frac{645^2}{7} \frac{4884^2}{5}$$

$$- 104385-84 = 22325 \cdot 69$$

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

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 6409.7125$$

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

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

$= 2.90115 \pm 2.52379$

95% of confidence interval for β = $(0.37736, 5.42484)$

iii) Mean IBM share price = μ_0

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

$x_0 = \mu_0$

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

$\hat{\mu}_0 = 119.79409$

$$se(\hat{\mu}_0) = \sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)\hat{\sigma}^2} = 10.59894$$

95% confidence int for μ_0

$= \hat{\mu}_0 \pm t_{10, 2.5\%}\, se(\hat{\mu}_0)$

$= 119.79409 \pm 23.61443$

95% confidence interval for μ_0 = $(96.17866, 143.40852)$

DOMS Page No.
Date / /

Here $Y_i^2 = \left(\sum_{j=1}^{3} Y_{ij}\right)^2 = (3 \times \text{Mean}_i)^2$

$SS_{Rate} = \sum_{i=1}^{4} \frac{Y_i^2}{3} - \frac{Y^2}{12} = \frac{973.373}{3} - \frac{(3\times 0.110)^2}{12}$

$SS_{Rate} = 21.153$

$SS_{Res} = 2 \times 0.618 = 1.236$

Source of Variation	d.f	SS	MSS
Rate	3	21.153	7.051
Residual	8	1.2.536	0.155
Total	11	22.389	

Test statistics $= \frac{MSS_{rate}}{MSS_{Res}} = 45.638$

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

Q.6

x	5	9	8	8	3	1	1	6	5	4
y	73	120	54	46	35	24	26	93	45	66

$n = 10.$

$r = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} \Rightarrow S_{XY} = \underline{\sum xy - n\bar{x}\bar{y}} = 661.2.$

$S_{XX} = \underline{\sum x^2 - n\bar{x}^2} = 54.9$

Q.5) Assumptn ANOVA :
Pop are normal
pop have common variance
observation have are indpendent

4 sample variances are very different from each other.
Hence, the common variance

ii) Mean at Rate 1 . $\frac{\sqrt{28} + \sqrt{13} + \sqrt{21}}{3}$

. 4.82443

Rate 2 . 0.12127

iii) ANOVA as for the transformation it can be claimed that 14 assumption of common var

iv) a $e_{ij} \sim N, 0\ \sigma^2$
b) for ANOVA
$H_0 : T_i = 0$
$H_1 = T_i \neq 0$ for atleast one i
ln(x)

Rate	Mean	Variance	$Y_{i\cdot}^2$
1	2.882	0.163	80.582
2	4.827	0.14	209.678
3	5.716	0.186	294.083
4	6.675	0.148	388.04
Total	20.110	0.616	973.373

DOMS Page No.
Date / /

$$SS_{TOT} = 1116.11$$

$$= \frac{1}{7}\left(64^2 - 44^2 + 8^2 + (-15)^2\right) - \frac{105^2}{28}$$

$$= 516.11$$

$$SS_{RES} = SS_{TOT} - SS_{REG} = 600$$

ANOVA TABLE :

Source	Degree of Freedom	Sum of Squares	Mean sum of Squares
Regression	3	516.11	172.04
Residual	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

$F_{3, 24, 5\%} = 3.009$ {Right tailed test)

Hence we have ~~did~~ sufficient evidence to reject H_0

iv) Analysis of mean diff

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

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

Source of Variation	df	SS	MSS
Regression	4	22226	558142
Residual	28	47604	170016
Total	32	69,830	

Test stat = 3.283.

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

Suff evidence to reject H_0.

c) Observed freq (O_i)

Papers cleared	Salary per annum in lakhs				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
Total	57	48	30	23	158

Expected Frequency : (E_i)

	3-5	5-8	8-10	10-12
0-3	27.41772	23.08861	14.43038	11.0632
4-6	15.15189	12.75849	7.97461	6.1158
7-9	14.43038	12.15189	7.59484	5.[illegible]

$S_{yy} = \Sigma y^2 - n\bar{y}^2 = 8928.6$

$r = 0.94466.$

ii) fitted regression eq^n

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

$$\hat{\alpha} = 9.2285$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 12.04371$$

$$\hat{y} = 9.2285 + 12.04371\, x.$$