Probabilty & Statistics - 2
Assignment - 2.

1. $\Sigma x = 850$, $\Sigma y = 1737$; $\Sigma xy = 182560$, $\Sigma x_i^2 = 92500$, $\Sigma y_i^2 = 372717$. , $n = 10$.

a. Regression of line: y_i

$$y_i = \hat{\alpha} + \hat{\beta} x_i$$

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

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

$$= 182560 - \frac{850 \times 1737}{10}$$

$$= 34915$$

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

$$= 92500 - \frac{(850)^2}{10}$$

$$= 20250$$

$$\hat{\beta} = \frac{34915}{20250}$$

$$= 1.724197531$$

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

$$= 1737 - 1.724197531(850)$$

$$= 271.4321$$

$\therefore$ Regression of line is : $\hat{y} = 271.4321 + 1.724197531$
$= \boxed{273.1562975}$

b) Gradient of regression line = $\hat{\beta} = 1.724197531$.

Q.2. i) $\hat{\beta} = \dfrac{S_{xy}}{S_{xx}}$

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

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

$$= 12666.58 - \frac{(385.2)^2}{12}.$$

$$S_{xx} = 301.66$$

$$S_{xy} = \Sigma x_i y_i - \frac{(\Sigma x_i)(\Sigma y_i)}{n}.$$

$$= 38191.41 - \frac{(385.2)(1162.5)}{12}.$$

$$= 874.839$$

$\hat{\beta} = 2.900082875$

$\hat{y}_f$ $\hat{\alpha}_i = 3.748181$ $\hat{\beta} = 2.900082875$

$\hat{y}_i = 3.748181 + 2.900082875 x_i$ //

≈ 6.648263875 //

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

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

$\sigma^2 = \dfrac{1}{10}\left[6409.7125 - \dfrac{(875.160)^2}{301.66}\right]$

$= 387.0744703$

$Se(\hat{\beta}) = 1.13277.$

$P\left[-2.228 < \dfrac{2.901 - \beta}{1.13277} < 2.228\right]$

$= 0.95$

$P[-2.228(1.113277) - 2.901 < -\beta < 2.228(1.132) - 2.901]$

$= 0.52.$

95% C.I for $\beta = \{2.901 - 2.228(1.13277), 2.901 + 2.228(1.13277)\}$

$= \{\cancel{0.31904}\ 0.37718549, 5.42481156\}$ //

iii) $PQ = \dfrac{\hat{\lambda}_i - \lambda_i}{Se(\hat{\lambda}_i)} \sim t_{m-2}$

$Se(\lambda_i) = \sqrt{\left(\dfrac{1}{12} + \dfrac{(7.9)^2}{301.66}\right) 38707}$

$= 10.5988.$

$P\left[-2.228 < \dfrac{119.79 - \lambda_i}{10.5988} < 2.228\right]$

$= 0.95.$

95% CI for $\lambda_i = \{119.79 - 2.228(10.5988), 119.79 + 2.228(10.5988)\}$

$= \{96.17, \cancel{143} \; 143.40\}$ //

Q.3. ii) Assumptions underlying analysis of variance:

- The poplulation should have a common variance.
- The poplulation must be normal.
- The observations are independent.

iii) H_0: The mean of differences is same for every city.
H_1: The mean of difference is not same for every city.

$SS_{TOT} = S_{yy}$

$$= \Sigma \hat{y}_i^{2} - \frac{(\Sigma y_i)^2}{n}.$$

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

$$= 1116.107143$$

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

$$= 516.1071429$$

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

$= 600$

ANOVA Table.

Source of Variation	d.f.	Sum of squares	Mean sum of squares.
Regression	3	516.1071429	172.04
Residual	24	600	25
Total	27	1116.107143	

Under H_0:

$$F = \frac{MSS_{REG}}{MSS_{RES}} = \frac{172.04}{25} = 6.88.$$

TEST Statistic = 6.88

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

∴ We have sufficient evidence to reject H_0 at L.S of 5%.

∴ There is difference between the cities.

iv) Analysis of mean difference.

Since $\sqrt{y_1} = 9.14$

$\sqrt{y_2} = 6.29$

$\sqrt{y_3} = 1.14$

$\sqrt{y_4} = -1.86$.

$\sqrt{y_1} > \sqrt{y_2} > \sqrt{y_3} > \sqrt{y_4}$.

$\hat{\sigma}^2 = \frac{SS_{RES}}{m-2} = 25$.

6. $\Sigma x = 39$, $\Sigma x^2 = 207$, $\Sigma y = 562$, $\Sigma y^2 = 40508$,
$\Sigma xy = 2853$.

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

$$= 54.9$$

$$S_{yy} = 40508 - \frac{562^2}{10}$$

$$= 8923.6$$

$$S_{xy} = \cancel{562}\ 2853 - \frac{(562)(39)}{10}$$

$$= 661.2.$$

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

$$= \frac{661.2}{\sqrt{54.9 \times 8923.6}}$$

$$\boxed{r = 0.944662390}$$ //

ii. $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{661.2}{54.9} = 12.04371585$

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

$= 56.2 - 12.04371585 \times (3.9)$

$\boxed{\hat{\alpha} = 9.229508185}$ //

iii. $SS_{TOT} = S_{yy}$

$= 8923.6$ //

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

$= \frac{661.2^2}{54.9}$

$= 7963.30492$ //

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

$= 960.29508$ //

iv. $r^2 = \frac{S^2_{xy}}{S_{xx} . S_{yy}}$

$= 89.24\%$

v. Comment:

Model explains 89% of variability. Hence it it a good fit.

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

$S_{xx} = 4789.4221$

$S_{yy} = 1189.207673$

$S_{xy} = 2176.84$

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

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$
$$= 10.78056.$$

$$\boxed{\hat{y}_i = 10.78056 + 0.45451\, x_i}$$

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

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

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

$$= 22.20136.$$

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

$$T.S = \frac{0.45451 - 0.}{\sqrt{\frac{22.20136}{4789.4221}}} \sim t_{(n)}$$

$$= 6.6755 \sim t_m$$

$$t_m > 2.5\% = 2.262.$$

∴ We have sufficient evidance to Reject H_0 at 5% L.S.

∴ $B \neq 0$ has a linear relationship.

__Assumptions.__

1) Population have a common variance.
2) Population follow a normal distribution.
3) The observations are independent.

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

$$\boxed{r = 0.9121286 58}$$

9. $\Sigma x = 45$, $\Sigma x^2 = 277.5$, $\Sigma y = 644$, $\Sigma y^2 = 43956$, $\Sigma xy = 2602$

i. Between the marks scored and hours spent per day on social media seems to have a negative correlation from the above scatterplot.

ii. $S_{xx} = 75$.

$S_{xy} = -296$.

$S_{yy} = 2482.4$

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

$= -0.217996208$.

$\therefore$ There is a weak negative correlation.

iii. $H_0 : \rho = 0$, $H_1 : \rho < 0$.

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

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

$T.S. = -2.07893$.

$Z_{5\%} = 1.6448$

$|Z_{cal}| > Z_{tab}$

Hence, we have sufficient evidence to Reject H_0

$f < 0$.

iv. $\hat{\beta} = \dfrac{S_{xy}}{S_{xx}}$

$= -3.94667$.

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

$= 82.16$.

$\boxed{\hat{y} = 82.16 - 3.94667x}$ //

v. $r^2 = \dfrac{S^2_{xy}}{S_{xx} \cdot S_{yy}}$

$= 47.058\ \%$.

$\therefore$ The model explains only 47% of the variability.

$\therefore$ The model is not a good fit for the data.

vi) $x_0 = 1$

Expected change $= \hat{\beta}x_0$.

$= -3.9467 \times 1$

$= -3.9467$.

10. $\Sigma x = 0.93$, $\Sigma x^2 = 0.2925$, $\Sigma y = 0.23$, $\Sigma y^2 = 0.0189$,
$\Sigma xy = 0.0735$, $n = 3$.

$S_{xx} = 0.0042$.
$S_{yy} = 0.0126$.
$S_{xy} = 0.0022$

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

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

$SS_{RES} = 0.0011$

ANOVA Table.

Source of Variation	d.F.	Sum of Squares	Mean of Squares.
Regression	1	0.00115	0.00115
Residual	1	0.00011	0.00011
Total	2	0.00126	

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

$$T.S = \frac{0.00115}{0.00011} \sim F_{11}$$

$= 10.4545 \sim F_{11}$

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

$F_{cal} < F_{tab}$

$\therefore$ We do not reject H_0.

$\therefore$ The ressisation rate has no effect on the industry