Probability & Stats

Assignment

Sol ① $\Sigma x^2 = 92500$; $\bar{x} = 85$

$\Sigma y^2 = 372717$; $\bar{y} = 173.7$.

$\Sigma xy = 182560$

$S_{xx} = \underline{\Sigma x_i^2 - n(\bar{x})^2} \Rightarrow 20250$.

$S_{yy} = \Sigma y^2 - n(\bar{y})^2 \Rightarrow 71000.1$

$S_{xy} = \Sigma xy - n(\bar{x}\bar{y}) \Rightarrow 34915$.

(a) $\beta = \dfrac{S_{xy}}{S_{xx}} = 1.7242$.

$\alpha = \bar{y} - \beta\bar{x} \Rightarrow 27.143$.

So, $\boxed{\hat{y} = 27.143 + 1.7242\,x}$

(b) gradient is no. of hours he spend in Atm upon ~~no~~ amount withdrawn.

Ques 12) $\Sigma x = 385.2$ $\quad \Sigma x^2 = 12666.58$

(i) $\Sigma y = 1162.5$ $\quad \Sigma y^2 = 119026.9$

$\Sigma xy = 38191.41$

$S_{xx} = 301.66$ $\quad S_{yy} = 6409.7125$

$S_{xy} = 875.16$

$\hat{\beta} = 2.901147$ $\quad \hat{\alpha} = 3.74818$

$$\boxed{y = 3.74818 + 2.901147x}$$

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

$$= 387.0745$$

$$-2.228 \le \frac{\hat{\beta} - \beta}{\hat{\sigma} / \sqrt{S_{xx}}} \le 2.228$$

$$\hat{\beta} - 2.228 \frac{\hat{\sigma}}{\sqrt{S_{xx}}} \le \beta \le \hat{\beta} + 2.288 \frac{\hat{\sigma}}{\sqrt{S_{xx}}}.$$

$$(0.377, 5.425).$$

(iii) $\hat{\mu}_0 = 119.794$

$Se(\hat{\mu}_0) = 10.5989$

95% confidence Interval of μ_0

$\Rightarrow 119.794 \pm 2.228(10.598)$

$\Rightarrow 119.794 \pm 23.6144$

$\Rightarrow (96.17956, 143.4084)$

Sol ③ (i) Comments on Plot :

- The centres of distributions differ for all four cities. Thus there's a prima facie case for suggesting that underlying means are different.
- The difference b/w mean time taken to communicate to office in peak hours & non peak hours are in order City A, City D [highest, lowest].

(ii) Assumptions

→ The Population must be normal.

→ The Population must have a common variance.

→ The observations are independent.

(iii) H_0 : The mean diff is same for each city

H_a : The mean diff are not same

$$SS_T = 1495 - \frac{103^2}{28} = 1116.11$$

$$SS_B = \frac{1}{7}\left(64^2 + 44^2 + 8^2 + (-13)^2\right) - \frac{103^2}{28} = 516.11$$

$$SS_R = SS_T - SS_B = 600.00$$

ANOVA

Source	df	SS	MS	F
Treatments	3	516.11	172.04	6.88
Residual	24	600.00	25.00	
Total	27	1116.11		

Under H_0 : $f = \frac{172.04}{25} = 6.88$ using $F_{3,24}$ distribution

The 5% critical point is 3.009, so we have sufficient evidence to reject H_0.

(iv) Since : $\bar{y}_{1*} = 9.14$ $\quad \bar{y}_{2*} = 6.29$ $\quad \bar{y}_{3*} = 1.14$ $\quad \bar{y}_{4*} = -1.86$

$\bar{y}_1 > \bar{y}_2 > \bar{y}_3 > \bar{y}_4$; $\hat{\sigma}^2 = \frac{SS_R}{n-k} = 25$.

$$t_{24, 0.025} * \hat{\sigma} \sqrt{\left(\frac{1}{7} + \frac{1}{7}\right)} = 2.064 * \sqrt{25} * \sqrt{\frac{1}{7} + \frac{1}{7}} = 5.52$$

$\bar{y}_1 - \bar{y}_2 = 2.85$, $\bar{y}_2 - \bar{y}_3 = 5.15$, $\bar{y}_3 - \bar{y}_4 = 3$

observing that all 3 differency are less than 5.52, we can say they have no significant difference.

Examining $\bar{y}_1 - \bar{y}_3 = 8$

hence there is a significant diff. b/w means 1 & 3

so.

Sol (4) (i) Already in sol (3),(ii)

(iii) $H_0 : \mu_1 = \mu_2 \cdots \mu_5$

H_a : atleast one of the mean is n't same.

$$\frac{T^2}{N} = \frac{(\Sigma y_1 + \Sigma y_2 + \Sigma y_3 + \Sigma y_4 + \Sigma y_5)^2}{N} = 104385.9394$$

$$SS_{Re} = \frac{(354)^2}{7} + \frac{(386)^2}{8} + \frac{(87)^2}{6} + \frac{(645)^2}{7} + \frac{(384)^2}{5} - \frac{T^2}{N}$$

$$= 22325.68917$$

$$SS_{TOT} = 69930.0606.$$

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

$$= 47604.37143$$

ANOVA

Sources of variation	df	SS	MS	F
Reg.	4	22325.69	5581.4225	3.2829
Res	28	47604.3143	1700.154	
Total	33	69930.0606		

$\frac{3.19}{10^3}$

So, $F_{4,28}$ tabulated is : 2.21

and our calculated value is 3.28

So, we have sufficient evidences to reject Ho.

~~So~~ (c) Ho : They are independent of each other.

Ha : They are dependent.

Expected values

Paper cleared	Salary 3-5	5-8	8-10	10-12
0-3	27.42	23.089	14.43	11.06.
4-6	15.15	12.76	7.97	6.14
7-9	14.43	12.15	7.59	5.82

$$\chi^2_{cal} = \frac{(E_i - O_i)^2}{E_i}$$

$\Rightarrow$ 11.27 + 0.4133 + 4.925 + 3.3204 + 4.384 + 4.1079 + 1.4175 + 7.234 + 6.56 + 0.133 + 0.00319 + 6.162

$\Rightarrow \boxed{39.929}$

$\chi^2_6 = 12.592$

Since $\chi^2_{cal} > \chi^2_{Tab}$; hence we have sufficient evidences to reject Ho.

Sol(5) (i) Already mentioned in Q3 (ii)

(ii) For $x = \sqrt{x}$, value of sample mean for Horse 1 is
$\frac{1}{3}(\sqrt{29} + \sqrt{18} + \sqrt{21}) = 4.52$

For $x = \log_e x$, the value of sample variance for Horse 2 is
$\frac{1}{2}[(\log_e 180 - 4.827)^2 + (\log_e 90 - 4.827)^2 + (\log_e 120 - 4.827)^2]$

$\Rightarrow \boxed{0.1213}$

(iii) The scientist was correct in asserting that the $\log_e$ transformation must be done before carrying out a one-way ANOVA as for this transformation it can be claimed the assumption of common variance for the underlying population hold.

(iv) ANOVA on $\log_e x$ data:
We assume the following model
$y_{ij} = \mu + \tau_i + e_{ij}$; $i = 1, 2, 3, 4$; $j = 1, 2, 3$

$H_0 : \tau_i = 0$, $i = 1, 2, 3, 4$ $H_1 :$ $\tau_i \neq 0$ for atleast one i

ANOVA

Source of variation	df	SS	MS	F
Pales	3	21.153	7.051	45.638
Residuals	8	1.236	0.155	
Total	11	22.389		

Sol(5) (i) Already mentioned in Q4 (iii)

(ii) for $x \to \sqrt{x}$, value of sample mean for Hose 1 is

$$\frac{1}{3}\left(\sqrt{29} + \sqrt{13} + \sqrt{21}\right) = 4.52$$

for $x \to \log_e x$, the value of sample variance for Hose 2

$$\frac{1}{2}\left[(\log_e 180 - 4.827)^2 + (\log_e 70 - 4.827)^2 + (\log_e 120 - 4.827)^2\right]$$

$$\Rightarrow \boxed{0.1213}$$

(iii) The scientist was correct in assuming that the $\log_e$ transformation must be done before carrying out a one-way ANOVA as for this transformation it can be claimed the assumption of common variance for the underlying population hold

(iv) We assume the following model: ANOVA on $\log_e x$ data

$$y_{ij} = \mu + \tau_i + e_{ij} \quad ; \quad i = 1, 2, 3, 4 \quad ; \quad j = 1, 2, 3$$

$H_0: \tau_i = 0, \; i = 1, 2, 3, 4$ $\quad H_1: \tau_i \neq 0$ for atleast one i

ANOVA

Sources of variation	df	SS	MS	F
Rates	3	21.153	7.051	45.638
Residuals	8	1.236	0.155	
Total	11	22.389		

Sol(5) (i) Already mentioned in Q3 (ii)

(ii) for $x \to \sqrt{x}$, value of sample mean for Hose 1 is

$$\frac{1}{3}(\sqrt{29}+\sqrt{13}+\sqrt{21}) = 4.52.$$

for $x \to \log_e x$, the value of sample variance for Hose 2

$$\frac{1}{2}\left[(\log_e 180-4.827)^2+(\log_e 90-4.827)^2+(\log_e 120-4.827)^2\right]$$

$\Rightarrow$ $\boxed{0.1213}$

(iii) The scientist was correct in asserting that the $\log_e$ transformation must be done before carrying out a one-way ANOVA as for this transformation it can be claimed the assumption of common variance for the underlying population hold.

(iv) we assume the following model.

ANOVA on $\log_e x$ data:

$$y_{ij} = \mu + \tau_i + e_{ij} \quad ; \quad i=1,2,3,4 \quad ; \quad j=1,2,3.$$

$H_0: \tau_i = 0, i=1,2,3,4$ $H_1:$ $\tau_i \neq 0$ for atleast one 'i'.

ANOVA

Sources of variation	df	SS	MSS	F
Rates	3	21.153	7.051	45.638
Residuals	8	1.236	0.155	
Total	11	22.389		

The 1% critical value for $f_{(3,8)}$ distribution $\boxed{7.951}$

Since ~~$R^2_{cal} \times R^2_{Tab}$~~ $F_{cal} > F_{Tab}$

hence we have insufficient evidences to reject H_0.

Sq ⑥ (i) Correlation coefficient :

$S_{yy} = 8923.6$

$S_{xx} = 54.9$

$S_{xy} = 661.2$

$$r = \frac{S_{xy}}{\sqrt{S_{xy} * S_{xx}}} = 0.94466.$$

(ii) $\hat{\beta} = 12.0437$

$\hat{\alpha} = 9.22957$

$$\boxed{y = 9.22957 + 12.0437x}$$

~~$SS_{RES} =$~~

(iii) $SS_{RES} = \left[S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right] \Rightarrow 960.2951$

$SS_{TOT} = 8923.6$

$SS_{REG} = 7963.305.$

(iv) $R^2 = \frac{SS_{REG}}{SS_{TOT}} = \frac{7963.305}{8923.6} = 0.89148$

Hence: 89% of variation is shown by our model

Sol (i) $S_{xx} = \sum x^2 - n(\bar{x})^2$

$= 4789.4221$

$S_{yy} = \sum y^2 - n(\bar{y})^2 = 1189.2077$.

$S_{xy} = \sum(x-\bar{x})(y-\bar{y}) = 2176.84$.

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

$\alpha = \bar{y} - \hat{\beta}\bar{x} = 10.79056$.

So, required eq^n : $\boxed{y = 10.79056 + 0.45451\, x}$

$\hat{\sigma}^2 = 22.20136$.

(ii) $H_0 : \beta = 0 \qquad H_a : \beta \neq 0$

$Z_c = \frac{\hat{\beta} - \beta}{\hat{\sigma} / \sqrt{S_{xx}}} = \frac{0.45451}{\sqrt{22.20136 / 4789.4221}}$

$Z_c = 6.67567$.

$Z_{tab} = 2.262$ since : $Z_{cal} > Z_{tab}$

Hence, we have insufficient evidences to reject H_0.

(iii) $r = \frac{S_{xy}}{\sqrt{S_{xx} * S_{yy}}} = \frac{2176.84}{\sqrt{4789.4221 \times 1189.2077}}$

$= 0.91213$

(iv) Given: $x^3 = 25 \Rightarrow$ $\boxed{x = 2.924}$

Now:

for mean response

$$\hat{\mu}_0 = 10.79 + 0.4545(2.924)$$

$$\boxed{\hat{\mu}_0 = 12.11773}$$

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

$$= 22.20136\left[\frac{1}{11} + \frac{(2.924 - 15.83)^2}{4789.4221}\right]$$

$$\boxed{Var(\hat{\mu}_0) = 2.790417}$$

95% CI:

$$\hat{\mu}_0 \pm 2.262\sqrt{2.790417}$$

$$\Rightarrow 12.11773 \pm 2.262\sqrt{2.790417} \Rightarrow (8.3392, 15.8963)$$

for individual response.

$$\hat{y}_0 = 12.11773.$$

$$V(\hat{y}_0) = 24.991777.$$

95% CI

$$\hat{y}_0 \pm 2.262\sqrt{24.991777}$$

$$\Rightarrow 12.11773 \pm 2.262\sqrt{24.991777}$$

$$= (0.8096, 23.42587)$$ Solⁿ.

(v) Since, there is some pattern hence we can't say it exhibits normal distribution.

Sol 9 (i) The given plot shows negative correlation b/w marks scored and hours spent per day on social media.

(ii) $S_{xx} = 75$, $S_{xy} = -296$.

$S_{yy} = 2482.4$

$$r = \frac{-296}{\sqrt{2482.4\,(75)}} \approx \boxed{-0.686}$$

This shows -ve correlation coefficient.

(iii) $H_0 : \rho = 0$ $\quad H_a : \rho < 0$

$w = \tanh^{-1}(r) = -0.84036.$

$w \sim N\left(0, \frac{1}{7}\right)$

$$z_c = \frac{-0.84036 - 0}{\sqrt{1/7}} \Rightarrow -2.2234.$$

$z_{tab} = -1.96.$

Since : $z_c < z_{tab}$..

hence we reject null hypothesis.

(iv) $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = -3.94667$

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

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

(v) $R = r^2 = 0.470596$

So, 47% of variation is shown by our model.

(vi) So, when $x = 1$

$y = 82.16 - 3.94667$

$= 78.21333$

when $x = 2$

$y = 74.26666$

for every additional 1 hour spent, expected change is $\boxed{-3.94667}$