PROBABILITY and Statistics - 2

Page No. Date

Assignment - 2

Q.1] X - Amount withdrawn

y - no. of hours (until next visit)

i) $n = 10$

$\Sigma x_i = 850$ $\qquad$ $\Sigma y_i = 1737$

$\Sigma x_i^2 = 92500$; $\qquad$ $\Sigma y_i^2 = 372717$

$\bar{x} = 85$ $\qquad$ $\bar{y} = 173.7$

$\Sigma x_i y_i = 182560$.

$$S_{xx} = \Sigma x_i^2 - n(\bar{x})^2 = 92500 - (10 \times 85^2) = 20250$$

$$S_{xy} = \Sigma x_i y_i - n(\bar{x})(\bar{y}) = 182560 - (10 \times 85 \times 173.7) = 34915.$$

$\cancel{\beta}$ $\quad \bar{y} = \alpha_i + \hat{\beta}\bar{x}$.

$$\therefore \hat{\beta} = \frac{S_{xx}}{S_{xy}} = \frac{20250}{34915} = 0.57998$$

$$\hat{\alpha} = \hat{y} - \hat{\beta}\bar{x} = 173.7 - (0.57998 \times 85) = 124.4$$

$y_i = \alpha_i + \beta x_i$

$\boxed{\therefore y_i = 124.4 + 0.57998 x_i}$ is the equation of the regression line of y on x.

ii) $\hat{\beta}$ — slope of the line

$\hat{\beta} = 0.57998$

as this is not zero, x_i i.e ~~no~~ Amount withdrawn affects our dependent variable.

Page No.
Date

Q.2]

X - Dell
Y - IBM.

$\Sigma x = 385.2$

$\Sigma x^2 = 12666.58$

$\Sigma y = 1162.5$ $\quad n = 12,$

$\Sigma y^2 = 119026.9$ $\quad \bar{x} = 32.1$

$\Sigma xy = 38191.41,$ $\quad \bar{y} = 96.88$

i)

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

$= \cancel{385.2} - (12 \times 32.1^2)$
12666.58

$= 301.66$

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

$= 119026.9 - (12 \times 96.88^2)$

$= 6398.1$

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

$= 38191.41 - (12 \times 32.1 \times 96.88)$

$= 873.23$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \cancel{\frac{301.66}{873.23}} \frac{873.23}{301.66} = \cancel{0.345}\ 2.895$

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

$= 96.88 - (\cancel{0.345}\ 2.895 \times 32.1)$

$= \cancel{85.81}\ 3.95$

∴ The least squares regression line where Y - IBM - dependent variable & X - Dell (independent)

is $\boxed{y_i = 85.81 + 0.345 x_i}$ // $y_i = 3.95 + 2.895 x_i$

2) $E[\hat{\beta}] = \beta = \cancel{0.345}\ 2.895$

$V[\hat{\beta}] = \frac{\sigma^2}{S_{xx}} = \frac{387.03}{301.66} = 1.283$

$\sigma^2 = \frac{S_{yy} - \frac{S_{xy}^2}{S_{xx}}}{n-2} = \frac{6398.1 - \frac{(873.23)^2}{301.66}}{10}$

$= 387.03$

Page No. Date

$\therefore \dfrac{\hat{\beta}-\beta}{S_E[\hat{\beta}]} \sim t_{n-2}$

$\therefore$ 95% Symmetrical ~~inter~~ Confidence interval is given by from the table book.

$0.95 = P(-2.228 < t_{10} < 2.228)$.

$\therefore 0.95 = P\left(-2.228 < \dfrac{\hat{\beta}-\beta}{S_E(\hat{\beta})} < 2.228\right)$

$\therefore$ 95% Confidence Interval

$\hat{\beta} \pm 2.228 \cdot S_E(\hat{\beta})$

$\therefore \left(\cancel{0.345}\ 2.895 - 2.228 \cdot \sqrt{1.283}\ ,\ \cancel{0.345}\ 2.895 + 2.228 \cdot \sqrt{1.283}\right)$

$\left(\cancel{-2.18}\ 0.371,\ \cancel{2.87}\ 5.42\right)$.

is the 95% Confidence Interval for the slope.

It is assumed that $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

iii)

The We know that

$\dfrac{\hat{\mu}_0 - \mu_0}{S_E(\hat{\mu}_0)} \sim t_{n-2}$

$\mu_0 = E[Y / X = x_0]$.

$= \alpha + \beta x_0$

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

We know that

$\hat{\alpha} = 85.81 \qquad x_0 = 40$

$\hat{\beta} = \cancel{0.345}\ \cancel{3.95}\ 2.895$

$\therefore \hat{\mu}_0 = \cancel{85.81}\ 85.81 + (\cancel{0.345}\ 2.895 \times 40) =$

$= \cancel{99.61}\ 199.75$

∴ from the table book,

$$0.95 = P\left(-2.228 < t_{10} < 2.228\right)$$

The 95% Confidence Interval for the mean response is given by

$$0.95 = P\left(-2.228 < \frac{\hat{\mu}_0 - \mu_0}{SE(\hat{\mu}_0)} < 2.228\right)$$

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

$$= \left(\left[\frac{1}{12} + \frac{(40 - 32.1)^2}{301.66}\right] \times 387.03\right)^{1/2}.$$

$$= 10.6.$$

$$\therefore\ 0.95 = P\left(\hat{\mu}_0 - 2.228 \cdot SE(\hat{\mu}_0) < \mu_0 < \hat{\mu}_0 + 2.228 \cdot SE(\hat{\mu}_0)\right)$$

∴ 95% Confidence Interval for mean response is

$$\left[199.675 - (2.228 \times 10.6),\ 199.675 + (2.228 \times 10.6)\right]$$

i.e ~~[75.99, 123.23]~~ $[96.13,\ 143.36]$

Q.3]

ii) Assumptions for carrying out ANOVA test.

1) The population must be a normal distribution
2) The observations taken should be independent
3) The variance should be equal.

iii) Ho:- The mean difference among all the cities is same
H₁:- The mean difference of atleast one city is significantly different.

$$SS_{TOT} = S_{yy}$$
$$= \sum y^2 - n(\bar{y})^2$$
$$= \cancel{1495 - (7 \times 14.71^2)}$$
$$= 1495 - (28 \times 3.68^2)$$
$$= 1115.81$$

$\bar{y} = \frac{\sum y}{n} = \frac{103}{28} = \cancel{14.71}\ 3.68$

$n = 28$

$$SS_{REG} = \frac{1}{7}\left[64^2 + 44^2 + 8^2 + (-13)^2\right] - n(\bar{y}^2)$$
$$= 515.81$$

$\therefore SS_{RES} = SS_{TOT} - SS_{REG}$
$$= 1115.81 - 515.81$$
$$= 600$$

∴ The Anova table is created as follows

	Degrees of freedom	Sum of Squares	Mean sum of squares
Regression	3	515.81	171.94
Residual	24	600	25
Total	27	1115.81	

∴ Degrees of freedom are calculated as follows

$F_{3,24}$ — 3: no. of cities − 1; 24: no. of observations − no. of cities

Under Ho

$$F_{3,24} = \frac{MSS_{Reg}}{MSS_{RES}} = \frac{171.94}{25} = 6.8776 \approx 6.88$$

from the table book, $F_{3,24}$ at 5% LOS is 3.009.
As value of test statistic is greater than tabulated value. Reject Ho.

Page No.
Date

We can conclude that mean difference among cities are not same.

Q4.]

a) The mathematical model that describes the relationship between the response and treatment for the one way ANOVA is given by

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Y_{ij} – the j^{th} observation on th ith treatment.
$(j = \{1, 2, 3, 4, 5 \ldots n_i)$
$(i = 1, 2, 3, 4, 5,$ k levels$)$

τ_i - represents the ith treatment effect
ϵ_{ij} - random error present in the j^{th} observation on the ith treatment.

Assumptions -

1) The errors ~~e~~ ϵ_{ij} are assumed to be normally and independently distributed with mean 0 and variance σ^2_ϵ.

2) μ is always a fixed parameter, and $\tau_1, \tau_2 \ldots \tau_k$ are considered to be fixed paramete, ~~if the the~~

3) It is assumed ~~$\sum \tau_i = 0$, $i = 1, k$ Assumed~~ μ is taken such that $\sum \tau_i = 0$, $i = 1, 2, \ldots k$

i) H_0: mean claim amounts of all companies are same

H_1: mean claim amounts of all companies are not same.

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

$\Sigma y = 1856$

$\Sigma y^2 = 174,316.$

$C.F = \dfrac{(\Sigma y)^2}{n} = 104385.94$

$SS_{TOT} = \Sigma y^2 - C.F$

$= 174316 - 104385.94$

$= 69930.06$

$SS_{REG} = \sum^{K} \left(\sum_{i=j} y_{ij} \right)^2 / n_i - C.F$

$= \left[\dfrac{354^2}{7} + \dfrac{386^2}{8} + \dfrac{87^2}{6} + \dfrac{645^2}{7} + \dfrac{384^2}{5} \right] - 104385.94$

$= 22325.69$

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

$= 69930.06 - 22325.69$

$= 47604.37$

The ANOVA table is constructed as follows:-

	Degrees of freedom	Sum of Squares	Mean Sum of Squares
Regression	4	22325.69 ~~22325.69~~	5581.42
Residual	28	47604.37	1700.16
	32	69930.06	

$F_{4,28} = \dfrac{MSSREG}{MSSRES} = \cancel{\dfrac{22325.69}{47604.37}} = \dfrac{5581.42}{1700.16}$

$= 3.283$

from the table book,

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

$\therefore$ We have sufficient evidence to reject Ho.

iii).

~~Ho~~ ~~Sed~~ Ho: Salary is independent

H1:

Observed frequency (Oi),

Papers cleared.	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

The Expected frequency table

Papers cleared	Salary			
	3-5	5-8	8-10	10-12
0-3	27.418	23.089	14.43	11.06
4-6	15.152	12.78	7.97	6.11
7-9	14.43	12.15	7.8	5.82.

$$\therefore T.S = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

$$= \cancel{11.2}\ 11.27 + 0.414 + 4.92 + 3.32 + 4.2439 + 4.11 + 0.13183 + 0.00212 + 6.16284 + 1.4185 + 7.211 + 6.553$$

$$= 49.911$$

degrees of freedom $= \cancel{A} (r-1)(c-1)$
$= 2 \times 3$
$= 6.$

$\chi^2_6 = 12.59$
at 5% LOS.

Hence, we have sufficient evidence to reject H_0.
$\therefore$ Salaries are dependent on no. of actuarial papers cleared.

Page No.
Date

Q.6.] $\Sigma x = 39$ $\quad n = 10$

$\Sigma x^2 = 207$

$\Sigma y = 562$ $\quad \bar{x} = 3.9$

$\Sigma y^2 = 40508$ $\quad \bar{y} = 56.2$

$\Sigma xy = 2853$

$S_{xx} = \Sigma x_i^2 - n(\bar{x})^2$
$= 207 - (10 \times 3.9^2)$
$= 54.9$

$S_{yy} = \Sigma y_i^2 - n(\bar{y})^2$
$= 40508 - (10 \times 56.2^2)$
$= 8923.6$

$S_{xy} = \cancel{2}\,\Sigma xy - n(\bar{x})(\bar{y})$
$= 2853 - (10 \times 3.9 \times 56.2)$
$= 661.2$

$\therefore$ i) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \dfrac{661.2}{\sqrt{54.9 \times 8923.6}} = \cancel{0.944662}$ 0.944662

ii) $\hat{\beta} = \dfrac{S_{xy}}{S_{xx}} = \dfrac{\cancel{54.9}\ 661.2}{\cancel{8923.6}\ 54.9} = \cancel{0.0741}$ 12.044

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$
$= 56.2 - (\cancel{0.0741}\ 12.044 \times 3.9)$
$= \cancel{55.91}$ 9.23

$\therefore$ The fitted linear regression equation of y on x is

$\boxed{\cancel{y_i = 55.91 + 0.0741\, x_i}\ 12.044}$ $\quad \boxed{y_i = 9.23 + 12.044x_i}$

iii) SS_{TOT} $= S_{yy} = 8923.6.$
(total sum of squares

$SS_{REG} = \dfrac{S_{xy}^2}{S_{xx}} = \dfrac{661.2^2}{54.9} = 7963.3$
Regression sum of squares

Page No.
Date

SSRES = SSTOT − SSREG
residual sum of squares

$= 8923.6 - 7963.3$

$= 960.3.$

iv) $R^2 = \frac{S^2xy}{Sxx \cdot Syy} = \frac{SSREG}{SSTOT} = \frac{S^2xy}{Sxy/Syy} = \frac{7963.3}{8923.6}$

$\boxed{\therefore R^2 = 0.8924}$

$R^2 = r^2 = \left(\frac{Sxy}{\sqrt{Sxx \cdot Syy}}\right)^2 = 0.944662^2 = 0.8924$ //

Q.7.]

x - Mortality Rating
y - Premium Rate.

$\Sigma x = 174.13 \quad \Sigma y = 197.84 \qquad n = 11$

$\Sigma x^2 = 7545.90 \quad \Sigma y^2 = 4747.45$

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

$\bar{x} = 15.83 \quad \bar{y} = 17.99$

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

$Sxx = \Sigma x^2 - n\bar{x}^2$

$= 7545.90 - (11 \times 15.83^2)$

$= 4789.42$

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

$= 4747.45 - (11 \times 17.99^2)$

$= 1187.41.$

$y_i = \hat{\alpha} + \hat{\beta} x_i$ - linear regression equation of Premium rates y on x.

$\therefore \hat{\beta} = \frac{Sxy}{Sxx} = \frac{2176.84}{4789.42} = 0.4545$

Date

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$
$$= 17.99 - (0.4545 \times 15.83)$$
$$= 10.8$$

$\therefore$ linear regression equation of Y on X is

$$\boxed{y_i = 10.8 + 0.4545 x_i}$$

ii) $H_0: \beta = 0$ - No linear relationship between X and Y

$H_1: \beta \neq 0$ - There exists a linear relationship

We know that,

$$\frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \sim t_{n-2}$$

$n = 11$

$\hat{\beta} = 0.4545$

$$SE(\hat{\beta}) = \left[\frac{\sigma^2}{S_{xx}}\right]^{1/2}$$

$$= \left(\frac{22.002}{4789.42}\right)^{1/2}$$

$$= 0.0678.$$

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

$$= \left\{\frac{1187.41 - \frac{(2176.84)^2}{4789.42}}{9}\right.$$

$$= 22.002$$

$$\therefore TS = \frac{0.4545 - 0}{0.0678} = 6.7035$$

from the table book, at 5% level of significance the lower and upper values from t_9 distribution are -2.262 and 2.262

As the value of our test statistic does not lie in this interval, we have sufficient evidence to reject H_0.

$\therefore$ There exists a linear relationship between x and y.

iii) $\hat{\beta} \neq 0$ – it is assumed.

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{2176.84}{\sqrt{4789.42 \times 1187.41}} = 0.91275.$$

iv) $x^3 = 25$

~~By the linear transformation~~

$\cancel{x = 25^{1/3} = 2.92}$

$\cancel{x = 2.92}$

~~i.e. x_0~~

$\therefore$ 1) individual response.

$\boxed{\hat{y}_0 = \hat{\alpha} + \hat{\beta} x_0}$ – unbiased estimator

$\therefore$ We know that,

$$\frac{\hat{y}_0 - y_0}{SE(\hat{y}_0)} \sim t_{n-2}$$

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

$$= \left[\left(1 + \frac{1}{11} + \frac{(22.92 - 15.83)^2}{4789.42} \right) 22.002 \right]$$

$$= \cancel{4.98} \quad 4.96$$

$$\hat{y}_0 = 10.8 + 0.4545 \times \overset{25}{\cancel{2.92}})$$

$$= 22.16$$

$$\approx 22.2$$

$\therefore$ 95% CI is given by

from the table book,

$$0.95 = P(-2.262 < t_a < 2.262)$$

$$\therefore 0.95 = P\left(-2.262 < \frac{\hat{y}_0 - y_0}{SE(\hat{y}_0)} < 2.262 \right)$$

$\therefore$ 95% C.I is given as

$$[\hat{y}_0 - 2.262 \cdot SE(\hat{y}_0), \ \hat{y}_0 + 2.262 \cdot SE(\hat{y}_0)]$$

$$[22.20 - 2.262 \times 4.98, \ 22.20 + 2.262 \times 4.96]$$

$$\cancel{(0.865, 23.395)} \quad [10.98, 33.42]$$

ii) Mean response

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

$\therefore \hat{\mu}_0 = \hat{\alpha} + \hat{\beta} x_0$ — unbiased estimator

$$= 10.8 + 0.4545 \times \cancel{2.92}\ 25$$

$$= \cancel{12.13}\ \ 22.2$$

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

$$= \left(\left[\frac{1}{11} + \frac{(22.92 - 15.83)^2}{4789.42} \right] 22.002 \right)^{1/2}$$

$$= \cancel{1.66}\ \ 1.55$$

$\therefore$ 95% Confidence Interval is given by,

$$0.95 = P(-2.262 < t_9 < 2.262)$$

$$\therefore 0.95 = P\left(-2.262 < \frac{\hat{\mu}_0 - \mu_0}{SE(\hat{\mu}_0)} < 2.262\right)$$

$$\therefore \left[\hat{\mu}_0 - 2.262 \cdot SE(\hat{\mu}_0),\ \hat{\mu}_0 + 2.262 \cdot SE(\hat{\mu}_0)\right]$$

$$\left[22.2 - 2.262 \times 1.\cancel{66}55,\ 22.20 - 2.262 \times 1.\cancel{66}55\right]$$

$$\left[\cancel{8.375, 15.885}\right],\ \left[18.69, 25.71\right]$$

Q.8]

Principal Component analysis, also known as 'factor analysis' provides a method for reducing the dimensionality of the data set, X, in other words, it seeks to identify the key components necessary to model and understand the data.

Old variables have some correlation between them.
To reduce / eliminate the redundancy in the data, new variables known as 'principle components' are created which are uncorrelated, and are a linear combination of the 'old' variables.

ii) The diagonal elements are.
PC1 - 0.456
PC2 - 0.137
PC3 - 0.08
PC4 - 0.0165
PC 5 - 0.012

The sum of these diagonal elements is 0.7015

∴ Total variance explained by the PCi is as follows
PC1 = 65.0%.
Pc2 = 19.53%.
Pc3 = 11.4%.
PC4 = 2.35%.
PC5 = 1.72

iii) As The 1st 3 PCi explain over 95% of total variance, dimensionality can be reduced to these 3 PCs.

Q.9

$\Sigma x = 45$ $\Sigma y = 644$ $\Sigma xy = 2602$

$\Sigma x^2 = 277.5$ $\Sigma y^2 = 43956$

X - Hours $n = 10$

Y - marks $\therefore \bar{x} = 4.5$ $\bar{y} = 64.4$.

i) By observing the scatterplot we can observe there is a weak negative correlation between marks scored and hours spent per day on social media.

ii) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$

$\therefore S_{xy} = \Sigma xy - n(\bar{x})(\bar{y}) = 2602 - (10 \times 4.5 \times 64.4)$

$= -296.$

-2898

~~SS~~ $S_{xx} = \sum x^2 - n(\bar{x})^2$
$= 277.5 - (10 \times 4.5^2)$
$= 75$

$S_{yy} = \sum y^2 - n(\bar{y})^2$
$= 43956 - (10 \times 64.4^2)$
$= 2482.4$

$r = \frac{-296}{\sqrt{75 \times 2482.4}} = -0.686$

This indicates there is a negative correlation between marks scored and hour spent on social media.

ii) Using fisher's transformation

$W = \tanh^{-1} r = -0.8404$

$H_0: \rho = 0$
$H_1: \rho < 0$

$$W \sim N\left(\underbrace{\tanh^{-1}\rho}_{\text{mean}}, \underbrace{\frac{1}{\sqrt{n-3}}}_{\text{s.d}}\right)$$

$\therefore$ ~~The test statistic is given by~~.

$$\cancel{\frac{W - \tanh^{-1}\rho}{\frac{1}{\sqrt{n-3}}} \sim N(0,3)}$$

$\therefore$ ~~TS =~~

$\tanh^{-1}\rho = \tanh^{-1}(0) = 0$ $\quad \therefore W \sim N(0, 0.378)$

$\frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{7}} = 0.378$

$\therefore$ 95% Confidence Interval

~~from the table book~~

5: $T.S = \frac{W - 0}{0.378} = \frac{-0.8404}{0.378} = -2.223$

This is a one sided test.

$\therefore$ The lower 5% value from $N(0,1)$ distribution is -1.6449

As the value of our test statistic is lower than our lower 5% value, we have sufficient evidence to reject H_0.

$\therefore$ We can conclude that $\rho < 0$.

iv) Y - Marks $\rightarrow$ response variable X - Hours $\rightarrow$ explanatory variable.

$$\therefore \hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{-296}{75} = -3.95$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$
$$= 64.4 - (-3.95 \times 4.5)$$
$$= 82.175$$

$\therefore$ The linear regression line of y on x is as follows:-

$$\boxed{y_i = 82.175 - 3.95x_i}$$

v) Coefficient of Determination - R^2

$$R^2 = \frac{SS_{REG}}{SS_{TOT}} = \frac{S_{xy}^2/S_{xx}}{S_{yy}} = \frac{S^2_{xy}}{S_{xx} \cdot S_{yy}} = \frac{(-296)^2}{75 \times 2482.4}$$

$$= 0.4706$$

We can see that very less proportion of error is explained by the model.

vi) On the basis of the regression line obtained in part iv). we can say that on an average the marks reduce by 3.95 ~4 for every additional hour spent.

Page No.
Date

Q.10.

H_0: The mean ~~resig~~ attrition rate is same for all Industries
H_1: The mean attrition rate of atleast one industry is significantly different.

Mean of Resignation = $\frac{(27+36+30)}{3} = 31$.

$\rightarrow \Sigma(x-\bar{x})^2$

$\therefore SS_{TOT} = 20\left[(27-31)^2 + (36-31)^2 + (30-31)^2\right]$

$= 840$

$SS_R = 19(5^2 + 10^2 + 8^2)$
$= 3591.$

$$\underbrace{F}_{\substack{\text{No. of categories}\\ -1}}{}_{2,\underbrace{57}_{\substack{\text{No. of observation}\\ -\text{ No. of categories.}}}} = \frac{840/2}{3591/57} = 6.667$$

from the table book,
at 5% Level of Significance ~~the $F_{2,57}$ value~~ is.
$F_{2,60}$ value is 2.393
As our test statistic is higher than this,
We reject H_0.