
IT Helloprogrammers
Note by: Roshan Bist

Chapter-4
Multiple Correlation and Regression

⊛ **Multiple Correlation:-** In multiple correlation we study about the association (or relation) between three or more than three variables.

Let x_1, x_2 and x_3 be three variables where x_1 is dependent variable and x_2 & x_3 are independent variable. The correlation coefficient between dependent variable x_1 and joint effect of independent variable x_2 and x_3 is called multiple correlation coefficient between x_1 and joint effect of x_2 and x_3 and it is denoted by $R_{1.23}$.

Also denoted by $R_{2.13}$ → if x_2 is dependent variable and x_1 & x_3 are independent variables.

$R_{3.12}$ → if x_3 is dependent variable and x_1 & x_2 are independent variables.

Multiple Correlation is given by

$$\boxed{R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}}$$

Similarly for $R_{2.13}$ and $R_{3.12}$.

For two variables

$$\boxed{\text{Correlation coefficient } r = \frac{n\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}}$$

Properties of multiple Correlation Coefficient:

i) Multiple correlation coefficient lies between 0 to 1.

i.e, $0 \le R_{1.23} \le 1$

$0 \le R_{2.13} \le 1$

$0 \le R_{3.12} \le 1$

ii) Here, $R_{1.23} = R_{1.32}$

$R_{2.13} = R_{2.31}$

$R_{3.12} = R_{3.21}$

(i.e, Correlation between x and y is similar to correlation between y and x)

Coefficient of Multiple determination:

Coefficient of multiple determination is the square of coefficient of multiple correlation so, $R^2_{1.23}$, $R^2_{2.13}$ and $R^2_{3.12}$ are the coefficients of multiple determination.

Let $R_{1.23}=0.9$

then, Coefficient of multiple determination $(R^2_{1.23}) = (0.9)^2$
$= 0.81$.

Interpretation of $(R_{1.23})^2$ → This means total variation on dependent variable x_1 is 81% that is explained by indepent variables x_2 and x_3 and remaning $(100-81)\% = 19\%$ variation on depending variable is due to the effect of other independent variables other than x_2 and x_3.

Numerical Problems:

Q1. If $r_{12}=0.6$, $r_{13}=0.4$, $r_{23}=0.35$ then,

(i) Find the multiple correlation coefficient between x_1 and joint effect of x_2 and x_3.

(ii) Find the multiple correlation coefficient between x_2 and joint effect of x_1 and x_3.

Solution

Given, $r_{12}=0.6$
$r_{13}=0.4$
$r_{23}=0.35$

(i) Here, Multiple correlation coefficient betoeen x_1 and joint effect of x_2 and x_3 is

$$R_{1.23} = \sqrt{\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$$

$$= \sqrt{\frac{(0.6)^2+(0.4)^2-2\times0.6\times0.4\times0.35}{1-(0.35)^2}}$$

$$= 0.63$$

(ii). Multiple correlation coefficient between x_2 and joint effect of x_1 and x_3 is

$$R_{2.13} = \sqrt{\frac{r_{12}^2+r_{23}^2-2r_{12}\cdot r_{23}\cdot r_{13}}{1-r_{13}^2}}$$

$$= \sqrt{\frac{(0.6)^2+(0.35)^2-2\times0.6\times0.4}{1-(0.4)^2}} \qquad = 0.003$$

Note:- If $(R_{1.23}$ OR $R_{2.13}$ OR $R_{3.12}) > 1$ then r_{12}, r_{13} and r_{23} are inconsistent.

Correlation Coefficients

Q2. A sample of 10 values of three variables x_1, x_2 and x_3 were obtained as $\Sigma x_1 = 10$, $\Sigma x_2 = 20$, $\Sigma x_3 = 30$, $\Sigma x_1 x_2 = 10$, $\Sigma x_1 x_3 = 15$, $\Sigma x_2 x_3 = 64$, $\Sigma x_1^2 = 20$, $\Sigma x_2^2 = 68$, $\Sigma x_3^2 = 170$.

i) Find the partial correlation coefficient between x_1 and x_3 eliminating the effect of x_2.

ii) Find the multiple correlation coefficient of x_1 with x_2 and x_3.

Solution,

Here,

$$r_{12} = \frac{n\Sigma x_1 x_2 - \Sigma x_1 \Sigma x_2}{\sqrt{n\Sigma x_1 - (\Sigma x_1)^2}\sqrt{n\Sigma x_2 - (\Sigma x_2)^2}}$$

$$= \frac{10\times10 - 10\times20}{\sqrt{10\times20 - 100}\sqrt{10\times68 - 400}}$$

$$= -0.59$$

$$r_{13} = \frac{n\Sigma x_1 x_3 - \Sigma x_1 . \Sigma x_3}{\sqrt{n\Sigma x_1^2 - (\Sigma x_1)^2}\sqrt{n\Sigma x_3^2 - (\Sigma x_3)^2}}$$

$$= \frac{10\times15 - 10\times30}{\sqrt{10\times20 - (10)^2}\sqrt{10\times170 - (30)^2}}$$

$$= -0.53$$

$$r_{23} = \frac{n\Sigma x_2 x_3 - \Sigma x_2 \Sigma x_3}{\sqrt{n\Sigma x_2^2 - (\Sigma x_2)^2}\sqrt{n\Sigma x_3^2 - (\Sigma x_3^2)}}$$

$$= \frac{10\times64 - 20\times30}{\sqrt{10\times68 - (20)^2}\sqrt{10\times170 - (30)^2}}$$

$$= 0.085$$

i) Partial correlation coefficient between x_1 and x_2 eliminating the effect of x_2 is $r_{13.2} = \dfrac{r_{13} - r_{12} r_{32}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{32}^2}}$

$$= \frac{(-0.53) - (-0.598)\times0.085}{\sqrt{1 - (-0.598)^2}\sqrt{1 - (0.085)^2}}$$

$$= 0.729$$

ii> Multiple correlation coefficient of x_1 with x_2 and x_3 is,

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2 r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(-0.598)^2 + (-0.53)^2 - 2 \times (-0.598) \times (-0.53) \times 0.085}{1 - (0.085)^2}}$$

$$= 0.767$$

Partial Correlation Coefficient:-

Let x_1, x_2 and x_3 are three variables then partial correlation coefficient between x_1 and x_2 when x_3 is taken as constant is ~~giv~~ denoted by $R_{12.3}$ and given as:-

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2}\ \sqrt{1 - r_{23}^2}}$$

Similarly for $r_{13.2}$ and $r_{23.1}$

Properties:

i> Partial correlation coefficient lies between -1 to $+1$.

i.e, $-1 \leq r_{12.3} \leq +1$

$-1 \leq r_{13.2} \leq +1$

$-1 \leq r_{23.1} \leq +1$

ii> Here, $r_{21.3} = r_{12.3}$

$r_{31.2} = r_{13.2}$

$r_{32.1} = r_{23.1}$

Coefficient of partial determination:- Coefficient of partial determination is the square of coefficient of partial correlation. So, $r_{12.3}^2, r_{13.2}^2$ and $r_{23.1}^2$ are the coefficient of partial determination.

Interpretation → Let coefficient of partial correlation $(r_{13.2}) = 0.8$

Then coefficient of partial determination is $(r_{13.2}^2) = (0.8)^2 = 0.64$

⇒ This means the total variation on dependent variable is 64% that is explained by independent variable x_3 when the independent variable x_2 is taken as constant and remaining 36% variation on x_1 is due to the effect of other independent variation.

Q3. From the data given below find $r_{12\cdot3}$, $R_{1\cdot23}$, $r_{23\cdot1}$ and $R_{2\cdot13}$

$\Sigma x_1x_2 = 40$, $\Sigma x_1x_3 = 55$, $\Sigma x_2x_3 = 35$

$\Sigma x_1^2 = 90$, $\Sigma x_2^2 = 60$, $\Sigma x_3^2 = 50$

$n = 6.$

where x_1, x_2 and x_3 are variables measured from their mean.

Solution

Given, $\Sigma x_1x_2 = 40$, $\Sigma x_1x_3 = 55$, $\Sigma x_2x_3 = 35$.

$\Sigma x_1^2 = 90$, $\Sigma x_2^2 = 60$, $\Sigma x_3^2 = 50$

$n = 6$

Since, x_1, x_2 and x_3 are measured from their mean.

So,
$$r_{12} = \frac{\Sigma x_1x_2}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma x_2^2}}$$
$$= \frac{40}{\sqrt{90}\sqrt{60}}$$
$$= 0.54$$

Rough

we know that
$$r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$
$$= \frac{\frac{1}{n}\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\frac{1}{n}\Sigma(X-\bar{X})^2}\sqrt{\frac{1}{n}\Sigma(Y-\bar{Y})^2}}$$
$$= \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2}}$$
$$= \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$

$$r_{13} = \frac{\Sigma x_1x_3}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma x_3^2}}$$
$$= \frac{55}{\sqrt{90}\sqrt{50}}$$
$$= 0.639$$

$$r_{23} = \frac{\Sigma x_2x_3}{\sqrt{\Sigma x_2^2}\sqrt{\Sigma x_3^2}}$$
$$= \frac{35}{\sqrt{60}\sqrt{50}}$$
$$= 0.819$$

Check the answers error maybe in ans while using calculator consider them as normal errors and correct yourself.

Now,
$$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$
$$= \frac{0.54 - 0.639 \times 0.819}{\sqrt{1-(0.639)^2}\sqrt{1-(0.819)^2}}$$
$$= 0.038$$

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.54)^2 + (0.639)^2 - 2 \times 0.54 \times 0.639 \times 0.819}{1 - (0.819)^2}}$$

$$= 0.64$$

$$r_{23.1} = \frac{r_{23} - r_{12} \cdot r_{13}}{\sqrt{1 - r_{12}^2}\,\sqrt{1 - r_{13}^2}}$$

$$= \frac{0.819 - 0.54 \times 0.639}{\sqrt{1 - (0.54)^2}\,\sqrt{1 - (0.639)^2}}$$

$$= 0.73$$

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12} \cdot r_{23} \cdot r_{13}}{1 - r_{13}^2}}$$

$$= \sqrt{\frac{(0.54)^2 + (0.819)^2 - 2 \times 0.54 \times 0.819 \times 0.639}{1 - (0.639)^2}}$$

$$= 0.82$$

Q4. From the information given below calculate $r_{12.3}$, $r_{13.2}$ and $r_{23.1}$.

x_1	6	8	9	11	12	14
x_2	14	16	17	18	20	23
x_3	21	22	27	29	31	32

Sol^n

x_1	x_2	x_3	$u=x_1-A$ $=x_1-9$	$v=x_2-B$ $=x_2-17$	$w=x_3-C$ $=x_3-29$	u^2	v^2	w^2	uv	uw	vw
6	14	21	-3	-3	-8	9	9	64	9	24	24
8	16	22	-1	-1	-7	1	1	49	1	7	7
9	17	27	0	0	-2	0	0	4	0	0	0
11	18	29	2	1	0	4	1	0	2	0	0
12	20	31	3	3	2	9	9	4	9	6	6
14	23	32	5	6	3	25	36	9	30	15	18
			$\Sigma u=6$	$\Sigma v=6$	$\Sigma w=-12$	$\Sigma u^2=48$	$\Sigma v^2=56$	$\Sigma w^2=130$	$\Sigma uv=51$	$\Sigma uw=52$	$\Sigma vw=55$

Now,

$$r_{12} = \frac{n\Sigma uv - \Sigma u \cdot \Sigma v}{\sqrt{n\Sigma u^2-(\Sigma u)^2}\sqrt{n\Sigma v^2-(\Sigma v)^2}}$$

$$= \frac{6\times 51 - 6\times 6}{\sqrt{6\times 48-(6)^2}\sqrt{6\times 56-(6)^2}}$$

$$= 0.98$$

$$r_{13} = \frac{n\Sigma uw - \Sigma u \cdot \Sigma w}{\sqrt{n\Sigma u^2-(\Sigma(u))^2}\sqrt{n\Sigma w^2-(\Sigma w)^2}}$$

$$= \frac{6\times 52 - 6\times(-12)}{\sqrt{6\times 48-(6)^2}\sqrt{6\times 130-(-12)^2}}$$

$$= 0.95$$

$$r_{23} = \frac{n\Sigma vw - \Sigma v \cdot \Sigma w}{\sqrt{n\Sigma v^2-(\Sigma v)^2}\sqrt{n\Sigma w^2-(\Sigma w)^2}}$$

$$= \frac{6\times 55 - 6\times(-12)}{\sqrt{6\times 56-(6)^2}\sqrt{6\times 130-(-12)^2}}$$

$$= 0.92$$

Now,

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

$$= \frac{0.98 - 0.95 \times 0.92}{\sqrt{1-(0.95)^2}\sqrt{1-(0.92)^2}}$$

$$= 0.87$$

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}}$$

$$= \frac{0.95 - 0.98 \times 0.92}{\sqrt{1-(0.98)^2}\sqrt{1-(0.92)^2}}$$

$$= 0.65$$

$$r_{23.1} = \frac{r_{23} - r_{12} \cdot r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}}$$

$$= \frac{0.92 - 0.98 \times 0.95}{\sqrt{1-(0.98)^2}\sqrt{1-(0.95)^2}}$$

$$= -0.18$$

Q.5. Are the following data consistent;
$r_{23} = 0.8$, $r_{31} = -0.5$, $r_{12} = 0.6$.

Soln
For testing it is consistency we need to find multiple correlation coefficient $R_{1.23}$. (We take $R_{1.23}$, also we can take $R_{2.13}$ or $R_{3.12}$ for testing)

Now,
$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.6)^2 + (-0.5)^2 - 2 \times 0.6 \times (-0.5) \times 0.8}{1-(0.8)^2}}$$

$$= 1.118$$

Since $R_{1.23} \pm 1.118 > 1$. (Not in the range of 0 to 1). So, the data are inconsistent.

Multiple Regression:-

Multiple Regression is the functional relationship between three or more than three variables where one variable is dependent and remaning are independent variable. By the use of regression model we can be able to estimate the value of dependent variable with the help of independent variable.

Let X_1, X_2 and X_3 are three variables, if X_1 is dependent and X_2 and X_3 are independent then, the multiple regression equation is,

also we can take Y in place of X_1

$$X_1 = a + b_1 X_2 + b_2 X_3$$

where, $a \rightarrow X_1$ intercept

$b_1 \rightarrow$ regression coeff. of X_1 on X_2 when X_3 is taken as constant.

$b_2 \rightarrow$ regression coeff. of X_1 on X_3 when X_3 is taken as constant.

For finding the values of a, b_1 and b_2 we have,

$$X_1 = a + b_1 X_2 + b_2 X_3 \text{ --- (i)}$$

By using least square method the normal equations are:-

$$\Sigma X_1 = na + b_1 \Sigma X_2 + b_2 \Sigma X_3 \text{ --- (ii)}$$

On multiplying both sides by X_2 of eqⁿ (i)

$$\Sigma X_1 X_2 = a \Sigma X_2 + b_1 \Sigma X_2^2 + b_2 \Sigma X_2 X_3 \text{ --- (iii)}$$

On multiplying both sides by X_3 of eqⁿ (i)

$$\Sigma X_1 X_3 = a \Sigma X_3 + b_1 \Sigma X_2 X_3 + b_2 \Sigma X_3^2 \text{ --- (iv)}$$

Solving these three equations (ii), (iii) and (iv) we get the value of a, b_1 & b_2. Finally substituting values of a, b_1 & b_2 in eqⁿ (i) we get the solution.

Q.N.6 The table shows the corresponding values of the three variables X_1, X_2 and X_3.

X_1: 5 7 8 6 10 9

X_2: 12 20 30 40 33 25

X_3: 51 55 58 60 70 66

Find the regression equation of X_1 on X_2 and X_3. Estimate X_1 when $X_2 = 50$ and $X_3 = 100$. Where X_1 represents pull length, X_2 represents wire length and X_3 represents die height.

Soln.

Since X_1 depends upon X_2 & X_3 so, the multiple regression equation is; (also we can write $\hat{b_0}$ notation in place of a)

$$X_1 = a + b_1X_2 + b_2X_3 \quad \text{—— (i)}$$

For finding the value of a, b_1 and b_2 we have the following normal equations,

$$\Sigma X_1 = na + b_1\Sigma X_2 + b_2\Sigma X_3 \quad \text{— (ii)}$$

$$\Sigma X_1X_2 = a\Sigma X_2 + b_1\Sigma X_2^2 + b_2\Sigma X_2X_3 \quad \text{— (iii)}$$

$$\Sigma X_1X_3 = a\Sigma X_3 + b_1\Sigma X_2X_3 + b_2\Sigma X_3^2 \quad \text{— (iv)}$$

for the calculation of $\Sigma X_1, \Sigma X_2, \Sigma X_3, \Sigma X_2^2, \Sigma X_3^2, \Sigma X_1X_2, \Sigma X_2X_3, \Sigma X_1X_3$ we proceed as following table.

X_1	X_2	X_3	X_2^2	X_3^2	X_1X_2	X_1X_3	X_2X_3
5	12	51	144	2601	60	255	612
7	20	55	400	3025	140	385	1100
8	30	58	900	3364	240	464	1740
6	40	60	1600	3600	240	360	2400
10	33	70	1089	4900	330	700	2310
9	25	66	625	4356	225	594	1650
$\Sigma X_1=45$	$\Sigma X_2=160$	$\Sigma X_3=360$	$\Sigma X_2^2=4758$	$\Sigma X_3^2=21846$	$\Sigma X_1X_2=1235$	$\Sigma X_1X_3=2758$	$\Sigma X_2X_3=9812$

Now, we put the values of $\Sigma X_1, \Sigma X_2, \Sigma X_3, \Sigma X_2^2, \Sigma X_3^2, \Sigma X_1 X_2, \Sigma X_2 X_3, \Sigma X_1 X_3$ and n in (ii), (iii) and (iv).

$$6a + 160 b_1 + 360 b_2 = 45 \quad \text{— (v)}$$

$$160a + 4758 b_1 + 9812 b_2 = 1235 \quad \text{— (vi)}$$

$$360a + 9812 b_1 + 21846 b_2 = 2758 \quad \text{— (vii)}$$

<u>Using Cramer's Rule</u> (<u>OR</u>, we can find a, b_1, b_2 by directly solving equations).

Coefficient of a	coefficient of b_1	coefficient of b_2	Constant
6	160	360	45
160	4758	9812	1235
360	9812	21846	2758

Now,

$$D = \begin{vmatrix} 6 & 160 & 360 \\ 160 & 4758 & 9812 \\ 360 & 9812 & 21846 \end{vmatrix}$$

$$= 6(103943268 - 96275344) - 160(3495360 - 3532320) + 360(1569920 - 1712880)$$

$$= 455544$$

<u>Similarly</u>

$$D_1 = \begin{vmatrix} 45 & 160 & 360 \\ 1235 & 4758 & 9812 \\ 2758 & 9812 & 21846 \end{vmatrix}$$

$$= 45(103943268 - 96275344) - 160(26979810 - 27061496) + 360(12117820 - 13122564)$$

$$= -3581500$$

$$D_2 = \begin{vmatrix} 6 & 45 & 360 \\ 160 & 1235 & 9812 \\ 360 & 2758 & 21846 \end{vmatrix}$$

$$= 6(26979810 - 27061496) - 45(3495360 - 3532320) + 360(441280 - 444600)$$

$$= -22116.$$

$$D_3 = \begin{vmatrix} 6 & 160 & 45 \\ 160 & 4758 & 1235 \\ 360 & 9812 & 2758 \end{vmatrix}$$

$$= 6(13122564 - 12117820) - 160(441280 - 444600) + 45(1569920 - 1712880)$$

$$= 126464$$

Here,

b_0 OR $a = \dfrac{D_1}{D} = \dfrac{-3581500}{455544} = -7.862$

$b_1 = \dfrac{D_2}{D} = \dfrac{-22116}{455544} = -0.048$

$b_2 = \dfrac{D_3}{D} = \dfrac{126464}{455544} = 0.277$

Now putting values of a, b_1 & b_2 on 4 we get regression equation of X_1 on X_2 and X_3 as:

$$X_1 = -7.862 - 0.048 X_2 + 0.277 X_3.$$

Again, X_1 when $X_2 = 50$ and $X_3 = 100$ is,

$$X_1 = -7.862 - 0.048 \times 50 + 0.277 \times 100$$
$$= -7.862 - 2.4 + 27.7$$
$$= 27.7 - 10.262$$
$$= 17.438$$

Measure of variation:

Total variation (Total sum of square) = explained variation (Sum of square due to regression) + unexplained variation (Sum of square due to error.)

or, We can write TSS as SST

$\therefore TSS = SSR + SSE$

or, $\boxed{SSR = TSS - SSE}$

where; $TSS = \sum (Y-\bar{Y})^2$

$= \sum Y^2 - n\bar{Y}^2$

& $SSE = \sum (Y-\hat{Y})^2$

$= \sum Y^2 - b_0 \sum Y - b_1 \sum YX_1 - b_2 \sum YX_2$

derived from $\hat{Y} = b_0 + b_1X_1 + b_2X_2$
$b_0 \sum Y - b_1 \sum YX_1 - b_2 \sum YX_2$

Coefficient of Determination:

Coefficient of determination is also determined

as $(R^2) = \frac{SSR}{TSS}$

or, $R^2 = \frac{TSS - SSE}{TSS}$ $(\because SSR = TSS - SSE)$

or, $R^2 = 1 - \frac{SSE}{TSS}$

Interpretation:→ Coefficient of determination that measures the total variation on dependent variable explained by independent variable.

Standard Error of Estimate (S.E):

Standard error of estimate $(S.E) = \sqrt{\frac{SSE}{n-k-1}}$

where, n = no. of observations.

k = no. of independent variable.

Q.N.7 From following information of variables X_1, X_2 and X_3

$\Sigma X_1 = 13$, $\Sigma X_2 = 11$, $\Sigma X_3 = 51$, $\Sigma X_1^2 = 63$, $\Sigma X_2^2 = 95$, $\Sigma X_1 X_3 = 77$,
$\Sigma X_2 X_3 = 136$, $\Sigma X_1 X_2 = -240$, $\Sigma X_3^2 = 450$, $n = 10$.

i) Find the regression equation of X_3 on X_1 and X_2 and interpret the regression coefficients.

ii) Predict X_3 when $X_1 = 1$ and $X_2 = 4$.

iii) Compute TSS, SSR and SSE.

iv) Compute standard error of estimate.

v) Compute the coefficient of multiple determination and interpret.

Solution:

Given, The regression equation of X_3 on X_1 and X_2 is

$$X_3 = b_0 + b_1 X_1 + b_2 X_2 \quad \text{--- (i)}$$

For finding b_0, b_1 & b_2 we need to solve the following normal equations:-

$$\Sigma X_3 = nb_0 + b_1 \Sigma X_1 + b_2 \Sigma X_2 \quad \text{--- (ii)}$$
$$\Sigma X_1 X_3 = b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 \quad \text{--- (iii)}$$
$$\Sigma X_2 X_3 = b_0 \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2 \quad \text{--- (iv)}$$

Now, putting the given values in eqn (ii), (iii) and (iv),

$$10b_0 + 13b_1 + 11b_2 = 51.$$
$$13b_0 + 63b_1 - 240b_2 = 77$$
$$\& \quad 11b_0 - 240b_1 + 95b_2 = 136$$

Using Cramers Rule

Coeff. b_0	Coeff. b_1	Coeff. b_2	Constants
10	13	11	51
13	63	-240	77
11	-240	95	136.

Now,

$$D = \begin{vmatrix} 10 & 13 & 11 \\ 13 & 63 & -240 \\ 11 & -240 & 95 \end{vmatrix}$$

$$= 10(5985 - 57600) - 13(1235 + 2640) + 11(-3120 - 693)$$

$= -516150 - 50375 - 41943$
$= -608468$

$$D_1 = \begin{vmatrix} 51 & 13 & 11 \\ 77 & 63 & -240 \\ 136 & -240 & 95 \end{vmatrix}$$

$= 51(5985 - 57600) - 13(7315 + 32640) + 11(-18480 - 8568)$
$= -2632365 - 519415 - 297528$
$= -3449308$

$$D_2 = \begin{vmatrix} 10 & 51 & 11 \\ 13 & 77 & -240 \\ 11 & 136 & 95 \end{vmatrix}$$

$= 10(7315 + 32640) - 51(1235 + 2640) + 11(1768 - 847)$
$= 399550 - 197625 + 10131$
$= 212056$

$$D_3 = \begin{vmatrix} 10 & 13 & 51 \\ 13 & 63 & 77 \\ 11 & -240 & 136 \end{vmatrix}$$

$= 10(8568 + 18480) - 13(1768 - 847) + 51(-3120 - 693)$
$= 270480 - 11973 - 194463$
$= 64044$

Now,

$$b_0 = \frac{D_1}{D} = \frac{-3449308}{-608468} = 5.66$$

$$b_1 = \frac{D_2}{D} = \frac{212056}{-608468} = -0.348$$

$$b_2 = \frac{D_3}{D} = \frac{64044}{-608468} = -0.105$$

i) On putting values of b_0, b_1 & b_2 in (i) we get regression equation of X_3 on X_1 and X_2 as: $X_3 = 5.7 - 0.348X_1 - 0.105X_2$.

ii) X_3 when $X_1 = 1$ and $X_2 = 4$ is:

$X_3 = 5.7 - 0.348 \times 1 - 0.105 \times 4$
$= 4.932$

ii) (Continue) part:

Interpretation → Since, $b_1 = -0.348$, this means the value of dependent variable is decreased by -0.348 as per unit change in the value of X_1 and $b_2 = -0.105$, this means value of independent variable is decreased by -0.105 as per unit change in value of X_2.

iii) Here,

$$TSS = \Sigma X_3^2 - n\bar{X}_3^2 \qquad \left(\because \bar{X}_3^2 = \frac{X_3^2}{X_3}\right)$$
$$= 450 - 10 \times (8.82)^2$$
$$= 450 - 778.5467$$
$$= -328.5467$$

$$SSE = \Sigma X_3^2 - b_0 \Sigma X_3 - b_1 \Sigma X_3X_1 - b_2 \Sigma X_3X_2$$
$$= 450 - 5.66 \times 51 + 0.348 \times 77 + 0.105 \times 136$$
$$= 450 - 288.66 + 26.796 + 14.28$$
$$= 202.416$$

& $SSR = TSS - SSE$
$$= -328.5467 - 202.416$$
$$= -530.9627$$

iv) Standard error of estimate $(S.E)_{3.12} = \sqrt{\frac{SSE}{n-k-1}}$
$$= \sqrt{\frac{202.416}{10-2-1}}$$
$$= 5.377$$

v) Coefficient of multiple determination is given by,

$$(R_{3.12})^2 = \frac{SSR}{TSS}$$
$$= \frac{-530.9627}{-328.5467}$$
$$= 1.616$$

This question is from kec book 222 page Q.N.20

यो 1 भन्दा सानो आउनु पर्ने Question मा value wrong दिइएको छ। [Solving method यही हो यो जान्ने]

Interpretation → It means ___ % of total variation on dependent variable X_3 is explained by independent variable X_1 & X_2.

⊛. Significance test of regression Coefficient: [Harder]

Q.No.8: Given the following information from a multiple regression analysis; $n=20$, $b_1=4$, $b_2=3$, $Sb_1=1.2$, $Sb_2=0.8$. At 0.05 level of significance, determine whether each of explanatory (dependent) variable makes a significant contribution to the regression model.

Soln

Given,

$n=20$

$b_1=4$

$b_2=3$

$Sb_1=1.2$ (i.e, standard error of b_1)

$Sb_2=0.8$ (i.e, standard error of b_2).

level of significance $= \alpha = 0.05$.

Problem to test:

Null hypothesis (H_0): $\beta_1=0$ i.e, there is no linear relationship between dependent variable Y and independent variable X_1.

Alternative hypothesis (H_1): $\beta_1 \neq 0$ i.e, there is linear relationship between dependent variable Y and independent variable X_1.

Test statistics:

$$t_{cal} = \frac{b_1}{Sb_1} = \frac{4}{1.2} = 3.33$$

Critical value— the tabulated value of 't' at 0.05 level of significant with n-k-1 degree of freedom is $(t_{0.05}, 20-2-1)$

$$= t_{0.05, 17}$$
$$= 2.110$$

value of $t_{0.05, 17}$ from table given back of book in page no. 318

Decision:→ Since $t_{cal} = 3.33 > t_{tab} = 2.110$

So, H_0 is rejected i.e, H_1 is accepted.

Conclusion → Hence, there is linear relation between dependent variable Y & independent variable X_1.

Note:- We have done for X_1, Similarly we can do same for X_2.

Q.N.9: In order to establish the functional relationship between anual salaries (y), years of educated high school (x_1) and years of experience with the firm (x_2), data on these three variables were collected from a random sample of 10 persons working in a large firm. Analysis of data produces the following results. The sum of squares $\sum(Y-\bar{Y})^2 = 397.6$. Sum of squares due to error $\sum(Y-\hat{Y})^2 = 23.5$. Test the overall significance of regression coefficients at 5% level of significance.

Soln
We have,
regression model, $Y = b_0 + b_1 X_1 + b_2 X_2$.

$$n = 10.$$

$$\sum(Y-\bar{Y})^2 = TSS = 397.6$$

$$\sum(Y-\hat{Y})^2 = SSE = 235$$

So,

$$SSR = TSS - SSE$$
$$= 397.6 - 235$$
$$= 374.1.$$

Problem to test:

Null hypothesis (H_0); $\beta_1 = \beta_2 = 0$. i.e, there is no linear relationship between dependent variable Y and independent variables X_1 & X_2.

Alternative hypothesis (H_1); $\beta_1 \neq \beta_2 \neq 0$. i.e, there is linear " " " ....

Test Statistics:

$$F_{cal} = \frac{MSR}{MSE}$$

Question मा overall word आए यो Fcal method बाट गर्ने

where, MSR → Mean square due to regression.
MSE → Mean square due to error.

and $MSR = \frac{SSR}{\text{degree of freedom (d.f)}}$

$$MSE = \frac{SSE}{d.f}$$

Now, we construct ANOVA table for regression analysis:

Source of Variation	degree of freedom	Sum of Square	MSS	F_{cal}.
Regression	$F = 2$	$SSR = 374.1$	$MSR = \frac{SSR}{d.f} = \frac{374.1}{2} = 187.05$	$F_{cal} = \frac{MSR}{MSE} = \frac{187.05}{3.35} = 55.83$
Error	$n-F-1 = 7$	$SSE = 23.5$	$MSE = \frac{23.5}{7} = 3.35$	
Total	$n-1 = 10-1 = 9$	$TSS = 397.6$ (SSR+SSE)		

Critical value - the tabulated value of F at 0.05 level of significance with (2,7) d.f is $F_{0.05}(2,7)$.

$= 4.74$ ← from table given in book page no. 321

<u>Decision</u> → Since $F_{cal} = 55.835 > F_{tab} = 4.74$ So, H_0 is rejected. i.e, H_1 is accepted.

equals to or smaller अगर accept हुन्छ

<u>Conclusion</u> → Hence we can conclude that there is linear relationship between Y and independent variables.