UNIT 11 MULTIPLE CORRELATION

Structure

11.1 INTRODUCTION

In Unit 9, you have studied the concept of regression and linear regression. Regression coefficient was also discussed with its properties. You learned how to determine the relationship between two variables in regression and how to predict value of one variable from the given value of the other variable. Plane of regression for trivariate, properties of residuals and variance of the residuals were discussed in Unit 10 of this block, which are basis for multiple and partial correlation coefficients. In Block 2, you have studied the coefficient of correlation that provides the degree of linear relationship between the two variables.

If we have more than two variables which are interrelated in someway and our interest is to know the relationship between one variable and set of others. This leads us to multiple correlation study.

In this unit, you will study the multiple correlation and multiple correlation coefficient with its properties .To understand the concept of multiple correlation you must be well versed with correlation coefficient. Before starting this unit, you go through the correlation coefficient given in Unit 6 of the Block 2. You should also clear the basics given in Unit 10 of this block to understand the mathematical formulation of multiple correlation coefficients.

Section 11.2 discusses the concept of multiple correlation and multiple correlation coefficient. It gives the derivation of the multiple correlation coefficient formula. Properties of multiple correlation coefficients are described in Section 11.3

Objectives

After reading this unit, you would be able to

- describe the concept of multiple correlation;
- define multiple correlation coefficient;
- derive the multiple correlation coefficient formula; and
- explain the properties of multiple correlation coefficient.

11.2 COEFFICIENT OF MULTIPLE CORRELATION

If information on two variables like height and weight, income and expenditure, demand and supply, etc. are available and we want to study the linear relationship between two variables, correlation coefficient serves our purpose which provides the strength or degree of linear relationship with direction whether it is positive or negative. But in biological, physical and social sciences, often data are available on more than two variables and value of one variable seems to be influenced by two or more variables. For example, crimes in a city may be influenced by illiteracy, increased population and unemployment in the city, etc. The production of a crop may depend upon amount of rainfall, quality of seeds, quantity of fertilizers used and method of irrigation, etc. Similarly, performance of students in university exam may depend upon his/her IQ, mother's qualification, father's qualification, parents income, number of hours of studies, etc. Whenever we are interested in studying the joint effect of two or more variables on a single variable, multiple correlation gives the solution of our problem.

In fact, multiple correlation is the study of combined influence of two or more variables on a single variable.

Suppose, X_1, X_2 and X_3 are three variables having observations on N individuals or units. Then multiple correlation coefficient of X_1 on X_2 and X_3 is the simple correlation coefficient between X_1 and the joint effect of X_2 and X_3. It can also be defined as the correlation between X_1 and its estimate based on X_2 and X_3.

Multiple correlation coefficient is the simple correlation coefficient between a variable and its estimate.

Let us define a regression equation of X_1 on X_2 and X_3 as

$$X_1 = a + b_{12.3}X_2 + b_{13.2}X_3$$

Let us consider three variables x_1, x_2 and x_3 measured from their respective means. The regression equation of x_1 depends upon x_2 and x_3 is given by

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3 \qquad \ldots (1)$$

Where $X_1 - \overline{X}_1 = x_1, X_2 - \overline{X}_2 = x_2$ and $X_3 - \overline{X}_3 = x_3$

$$\therefore \sum x_1 = \sum x_2 = \sum x_3 = 0$$

Right hand side of equation (1) can be considered as expected or estimated value of x_1 based on x_2 and x_3 which may be expressed as

$$x_{1.23} = b_{12.3}x_2 + b_{13.2}x_3 \qquad \ldots (2)$$

Residual $e_{1.23}$ (see definition of residual in Unit 5 of Block 2 of MST 002) is written as

$$e_{1.23} = x_1 - b_{12.3}x_2 - b_{13.2}x_3 = x_1 - x_{1.23}$$

$$\Rightarrow e_{1.23} = x_1 - x_{1.23}$$

$$\Rightarrow x_{1.23} = x_1 - e_{1.23} \qquad \ldots (3)$$

The multiple correlation coefficient can be defined as the simple correlation coefficient between x_1 and its estimate $e_{1.23}$. It is usually denoted by $R_{1.23}$ and defined as

$$R_{1.23} = \frac{Cov(x_1, x_{1.23})}{\sqrt{V(x_1)V(x_{1.23})}} \qquad \ldots (4)$$

Now,

$$Cov(x_1, x_{1.23}) = \frac{1}{N}\sum(x_1 - \overline{x}_1)(x_{1.23} - \overline{x}_{1.23})$$

(By the definition of covariance)

Since, x_1, x_2 and x_3 are measured from their respective means, so

$$\sum x_1 = \sum x_2 = \sum x_3 = 0 \Rightarrow \overline{x}_1 = \overline{x}_2 = \overline{x}_3 = 0$$

and consequently

$$\overline{x}_{1.23} = b_{12.3}\overline{x}_2 + b_{13.2}\overline{x}_3 = 0 \qquad \text{(From equation (2))}$$

Thus,

$$\begin{aligned}
Cov(x_1, x_{1.23}) &= \frac{1}{N}\sum x_1 x_{1.23} \\
&= \frac{1}{N}\sum x_1(x_1 - e_{1.23}) && \text{(From equation (3))} \\
&= \frac{1}{N}\sum x_1^2 - \frac{1}{N}\sum x_1 e_{1.23} && \text{(By third property of residuals)} \\
&= \frac{1}{N}\sum x_1^2 - \frac{1}{N}\sum e_{1.23}^2 \\
&= \sigma_1^2 - \sigma_{1.23}^2 && \text{(From equation (29) of Unit10)}
\end{aligned}$$

$$\begin{aligned}
\text{Now} \quad V(x_{1.23}) &= \frac{1}{N}\sum(x_{1.23} - \overline{x}_{1.23})^2 \\
&= \frac{1}{N}\sum(x_{1.23})^2 && \text{(Since } \overline{x}_{1.23} = 0\text{)} \\
&= \frac{1}{N}\sum(x_1 - e_{1.23})^2 && \text{(From equation (3))} \\
&= \frac{1}{N}\sum(x_1^2 + e_{1.23}^2 - 2x_1 e_{1.23})
\end{aligned}$$

$$= \frac{1}{N}\sum x_1^2 + \frac{1}{N}\sum e_{1.23}^2 - 2\frac{1}{N}\sum x_1 e_{1.23}$$

$$= \frac{1}{N}\sum x_1^2 + \frac{1}{N}\sum e_{1.23}^2 - 2\frac{1}{N}\sum e_{1.23}^2$$

(By third property of residuals)

$$= \frac{1}{N}\sum x_1^2 - \frac{1}{N}\sum e_{1.23}^2$$

$$V(x_{1.23}) = \sigma_1^2 - \sigma_{1.23}^2 \qquad \text{(From equation (29) of Unit 10)}$$

Substituting the value of $Cov(x_1, x_{1.23})$ and $V(x_{1.23})$ in equation (4), we have

$$R_{1.23} = \frac{\sigma_1^2 - \sigma_{1.23}^2}{\sqrt{\sigma_1^2(\sigma_1^2 - \sigma_{1.23}^2)}}$$

$$R_{1.23}^2 = \frac{(\sigma_1^2 - \sigma_{1.23}^2)^2}{\sigma_1^2(\sigma_1^2 - \sigma_{1.23}^2)}$$

$$R_{1.23}^2 = \frac{\sigma_1^2 - \sigma_{1.23}^2}{\sigma_1^2} = 1 - \frac{\sigma_{1.23}^2}{\sigma_1^2}$$

here, $\sigma_{1.23}^2$ is the variance of residual, which is

$$\sigma_{1.23}^2 = \frac{\sigma_1^2}{1 - r_{23}^2}(1 - r_{23}^2 - r_{12}^2 - r_{13}^2 + 2r_{12}r_{23}r_{13})$$

(From equation (30) of unit 10)

Then,

$$R_{1.23}^2 = 1 - \frac{\sigma_1^2(1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})}{\sigma_1^2(1 - r_{23}^2)}$$

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

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

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

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} \qquad \ldots (5)$$

which is required formula for multiple correlation coefficient.

where, r_{12} is the total correlation coefficient between variable X_1 and X_2,

r_{23} is the total correlation coefficient between variable X_2 and X_3,

r_{13} is the total correlation coefficient between variable X_1 and X_3.

Now let us solve a problem on multiple correlation coefficients.

Example 1: From the following data, obtain $R_{1.23}$ and $R_{2.13}$

X_1	65	72	54	68	55	59	78	58	57	51
X_2	56	58	48	61	50	51	55	48	52	42
X_3	9	11	8	13	10	8	11	10	11	7

Solution: To obtain multiple correlation coefficients $R_{1.23}$ and $R_{2.13}$, we use following formulae

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

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

We need r_{12}, r_{13} and r_{23} which are obtained from the following table:

S. No.	X_1	X_2	X_3	$(X_1)^2$	$(X_2)^2$	$(X_3)^2$	X_1X_2	X_1X_3	X_2X_3
1	65	56	9	4225	3136	81	3640	585	504
2	72	58	11	5184	3364	121	4176	792	638
3	54	48	8	2916	2304	64	2592	432	384
4	68	61	13	4624	3721	169	4148	884	793
5	55	50	10	3025	2500	100	2750	550	500
6	59	51	8	3481	2601	64	3009	472	408
7	78	55	11	6084	3025	121	4290	858	605
8	58	48	10	3364	2304	100	2784	580	480
9	57	52	11	3249	2704	121	2964	627	572
10	51	42	7	2601	1764	49	2142	357	294
Total	617	521	98	38753	27423	990	32495	6137	5178

Now we get the total correlation coefficient r_{12}, r_{13} and r_{23}

$$r_{12} = \frac{N(\sum X_1X_2) - (\sum X_1)(\sum X_2)}{\sqrt{\{N(\sum X_1^2) - (\sum X_1)^2\}\{N(\sum X_2^2) - (\sum X_2)^2\}}}$$

$$r_{12} = \frac{(10 \times 32495) - (617) \times (521)}{\sqrt{\{(10 \times 38753) - (617) \times (617)\}\{(10 \times 27423) - (521) \times (521)\}}}$$

$$r_{12} = \frac{3493}{\sqrt{\{6841\}\times\{2789\}}} = \frac{3493}{4368.01} = 0.80$$

$$r_{13} = \frac{N(\sum X_1X_3) - (\sum X_1)(\sum X_3)}{\sqrt{\{N(\sum X_1^2) - (\sum X_1)^2\}\{N(\sum X_3^2) - (\sum X_3)^2\}}}$$

$$r_{13} = \frac{(10\times 6137) - (617)\times(98)}{\sqrt{\{(10\times 38753) - (617\times 617)\}\{(10\times 990) - (98\times 98)\}}}$$

$$r_{13} = \frac{904}{\sqrt{\{6841\}\{296\}}} = \frac{904}{1423.00} = 0.64$$

and

$$r_{23} = \frac{N(\sum X_2X_3) - (\sum X_2)(\sum X_3)}{\sqrt{\{N(\sum X_2^2) - (\sum X_2)^2\}\{N(\sum X_3^2) - (\sum X_3)^2\}}}$$

$$r_{23} = \frac{(10\times 5178) - (521)\times(98)}{\sqrt{\{(10\times 27423) - (521\times 521)\}\{(10\times 990) - (98\times 98)\}}}$$

$$r_{23} = \frac{722}{\sqrt{\{2789\}\{296\}}} = \frac{722}{908.59} = 0.79$$

Now, we calculate $R_{1.23}$

We have, $r_{12} = 0.80$, $r_{13} = 0.64$ and $r_{23} = 0.79$, then

$$R^2_{1.23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2_{23}}$$

$$= \frac{0.80^2 + 0.64^2 - 2\times 0.80\times 0.64\times 0.79}{1 - 0.79^2}$$

$$= \frac{0.64 + 0.41 - 0.81}{1 - 0.62}$$

$$R^2_{1.23} = \frac{0.24}{0.38} = 0.63$$

Then

$$R_{1.23} = 0.79.$$

$$R^2_{2.13} = \frac{r^2_{12} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{13}}$$

$$= \frac{0.80^2 + 0.79^2 - 2\times 0.80\times 0.64\times 0.79}{1 - 0.64^2}$$

$$= \frac{0.64 + 0.62 - 0.81}{1 - 0.49}$$

$$= \frac{0.45}{0.51} = 0.88$$

Thus,

$$R_{2.13} = 0.94$$

Example 2: From the following data, obtain $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$

X_1	2	5	7	11
X_2	3	6	10	12
X_3	1	3	6	10

Solution: To obtain multiple correlation coefficients $R_{1.23}$ $R_{2.13}$ and $R_{3.12}$, we use following formulae

$$R^2_{1.23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2_{23}} \;,$$

$$R^2_{2.13} = \frac{r^2_{12} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{13}} \text{ and}$$

$$R^2_{3.12} = \frac{r^2_{13} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{12}}$$

We need r_{12} , r_{13} and r_{23} which are obtained from the following table:

S. No.	X_1	X_2	X_3	$(X_1)^2$	$(X_2)^2$	$(X_3)^2$	X_1X_2	X_1X_3	X_2X_3
1	2	3	1	4	9	1	6	2	3
2	5	6	3	25	36	9	30	15	18
3	7	10	6	49	100	36	70	42	60
4	11	12	10	121	144	100	132	110	120
Total	25	31	20	199	289	146	238	169	201

Now we get the total correlation coefficient r_{12} , r_{13} and r_{23}

$$r_{12} = \frac{N(\sum X_1X_2) - (\sum X_1)(\sum X_2)}{\sqrt{\{N(\sum X_1^2) - (\sum X_1)^2\}\{N(\sum X_2^2) - (\sum X_2)^2\}}}$$

$$r_{12} = \frac{(4 \times 238) - (25) \times (31)}{\sqrt{\{(4 \times 199) - (25) \times (25)\}\{(4 \times 289) - (31) \times (31)\}}}$$

$$r_{12} = \frac{177}{\sqrt{\{171\}\{195\}}} = \frac{177}{182.61} = 0.0.97$$

$$r_{13} = \frac{N(\sum X_1X_3) - (\sum X_1)(\sum X_3)}{\sqrt{\{N(\sum X_1^2) - (\sum X_1)^2\}\{N(\sum X_3^2) - (\sum X_3)^2\}}}$$

$$r_{13} = \frac{(4\times169)-(25)\times(20)}{\sqrt{\{(4\times199)-(25\times25)\}\{(4\times146)-(20\times20)\}}}$$

$$r_{13} = \frac{176}{\sqrt{\{171\}\{184\}}} = \frac{176}{177.38} = 0.99$$

and

$$r_{23} = \frac{N(\sum X_2X_3)-(\sum X_2)(\sum X_3)}{\sqrt{\{N(\sum X_2^2)-(\sum X_2)^2\}\{N(\sum X_3^2)-(\sum X_3)^2\}}}$$

$$r_{23} = \frac{(4\times201)-(31)\times(20)}{\sqrt{\{4\times289)-(31\times31)\}\{(4\times146)-(20\times20)\}}}$$

$$r_{23} = \frac{184}{\sqrt{\{195\}\{184\}}} = \frac{184}{189.42} = 0.97$$

Now, we calculate $R_{1.23}$

We have, $r_{12} = 0.97$, $r_{13} = 0.99$ and $r_{23} = 0.97$, then

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

$$= \frac{0.97^2 + 0.99^2 - 2\times0.97\times0.99\times0.97}{1-0.97^2}$$

$$= \frac{0.058}{0.059} = 0.98$$

Then

$$R_{1.23} = 0.99.$$

$$R^2_{2.13} = \frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1-r_{13}^2}$$

$$= \frac{0.97^2 + 0.97^2 - 2\times0.97\times0.99\times0.97}{1-0.99^2}$$

$$= \frac{0.19}{0.20} = 0.95$$

Thus,

$$R_{2.13} = 0.97$$

$$R^2_{3.12} = \frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1-r_{12}^2}$$

$$= \frac{0.99^2 + 0.97^2 - 2\times0.97\times0.99\times0.97}{1-0.97^2}$$

$$= \frac{0.58}{0.591}$$
$$= 0.981$$

Thus,

$$R_{3.12} = 0.99$$

Example 3: The following data is given:

X_1	60	68	50	66	60	55	72	60	62	51
X_2	42	56	45	64	50	55	57	48	56	42
X_3	74	71	78	80	72	62	70	70	76	65

Obtain $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$

Solution: To obtain multiple correlation coefficients $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$ we use following formulae:

$$R^2_{1.23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2_{23}},$$

$$R^2_{2.13} = \frac{r^2_{12} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{13}}$$ and

$$R^2_{3.12} = \frac{r^2_{13} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{12}}$$

We need r_{12} , r_{13} and r_{23} which are obtained from the following table:

S. No.	**X_1**	**X_2**	**X_3**	**$d_1 = X_1 - 60$**	**$d_2 = X_2 - 50$**	**$d_3 = X_3 - 70$**	**$(d_1)^2$**	**$(d_2)^2$**	**$(d_3)^2$**	**d_1d_2**	**d_1d_3**	**d_2d_3**
1	60	42	74	0	-8	4	0	64	16	0	0	-32
2	68	56	71	8	6	1	64	36	1	48	8	6
3	50	45	78	-10	-5	8	100	25	64	50	-80	-40
4	66	64	80	6	14	10	36	196	100	84	60	140
5	60	50	72	0	0	2	0	0	4	0	0	0
6	55	55	62	-5	5	-8	25	25	64	-25	40	-40
7	72	57	70	12	7	0	144	49	0	84	0	0
8	60	48	70	0	-2	0	0	4	0	0	0	0
9	62	56	76	2	6	6	4	36	36	12	12	36
10	51	42	65	-9	-8	-5	81	64	25	72	45	40
Total				4	15	18	454	499	310	325	85	110

Here, we can also use shortcut method to calculate r_{12}, r_{13} & r_{23},

Let $d_1 = X_1 - 60$

$d_2 = X_2 - 50$

$d_3 = X_1 - 70$

Now we get the total correlation coefficient r_{12} , r_{13} and r_{23}

$$r_{12} = \frac{N(\sum d_1 d_2) - (\sum d_1)(\sum d_2)}{\sqrt{\{N(\sum d_1^2) - (\sum d_1)^2\}\{N(\sum d_2^2) - (\sum d_2)^2\}}}$$

$$r_{12} = \frac{(10 \times 325) - (4) \times (15)}{\sqrt{\{(10 \times 454) - (4) \times (4)\}\{(10 \times 499) - (15) \times (15)\}}}$$

$$r_{12} = \frac{3190}{\sqrt{\{4524\}\{4765\}}} = \frac{3190}{4642.94} = 0.69$$

$$r_{13} = \frac{N(\sum d_1 d_3) - (\sum d_1)(\sum d_3)}{\sqrt{\{N(\sum d_1^2) - (\sum d_1)^2\}\{N(\sum d_3^2) - (\sum d_3)^2\}}}$$

$$r_{13} = \frac{(10 \times 85) - (4) \times (18)}{\sqrt{\{(10 \times 454) - (4 \times 4)\}\{(10 \times 310) - (18 \times 18)\}}}$$

$$r_{13} = \frac{778}{\sqrt{\{4524\}\{2776\}}} = \frac{778}{3543.81} = 0.22$$

and

$$r_{23} = \frac{N(\sum d_2 d_3) - (\sum d_2)(\sum d_3)}{\sqrt{\{N(\sum d_2^2) - (\sum d_2)^2\}\{N(\sum d_3^2) - (\sum d_3)^2\}}}$$

$$r_{23} = \frac{(10 \times 110) - (15) \times (18)}{\sqrt{\{(10 \times 499) - (15 \times 15)\}\{(10 \times 310) - (18 \times 18)\}}}$$

$$r_{23} = \frac{830}{\sqrt{\{4765\}\{2776\}}} = \frac{830}{3636.98} = 0.23$$

Now, we calculate $R_{1.23}$

We have, $r_{12} = 0.69$, $r_{13} = 0.22$ and $r_{23} = 0.23$, then

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

$$= \frac{0.69^2 + 0.22^2 - 2 \times 0.69 \times 0.22 \times 0.23}{1 - 0.23^2}$$

$$= \frac{0.4547}{0.9471} = 0.4801$$

Then

$$R_{1.23} = 0.69$$

$$R^2_{2.13} = \frac{r^2_{12} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{13}}$$

$$= \frac{0.69^2 + 0.23^2 - 2 \times 0.69 \times 0.22 \times 0.23}{1 - 0.22^2}$$

$$= \frac{0.4592}{0.9516} = 0.4825$$

Thus,

$$R_{2.13} = 0.69$$

$$R^2_{3.12} = \frac{r^2_{13} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{12}}$$

$$= \frac{0.22^2 + 0.23^2 - 2 \times 0.69 \times 0.22 \times 0.23}{1 - 0.69^2}$$

$$= \frac{0.0315}{0.5239} = 0.0601$$

Thus,

$$R_{3.12} = 0.25$$

Now let us solve some exercises.

E1) In bivariate distribution, $r_{12} = 0.6$, $r_{23} = r_{31} = 0.54$, then calculate $R_{1.23}$.

E2) If $r_{12} = 0.70$, $r_{13} = 0.74$ and $r_{23} = 0.54$, calculate multiple correlation coefficient $R_{2.13}$.

E3) Calculate multiple correlation coefficients $R_{1.23}$ and $R_{2.13}$ from the following information: $r_{12} = 0.82$, $r_{23} = -0.57$ and $r_{13} = -0.42$.

E4) From the following data,

X_1	22	15	27	28	30	42	40
X_2	12	15	17	15	42	15	28
X_3	13	16	12	18	22	20	12

Obtain $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$

E5) The following data is given:

X_1	50	54	50	56	50	55	52	50	52	51
X_2	42	46	45	44	40	45	43	42	41	42
X_3	72	71	73	70	72	72	70	71	75	71

By using the short-cut method obtain $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$

11.3 PROPERTIES OF MULTIPLE CORRELATION COEFFICIENT

The following are some of the properties of multiple correlation coefficients:

1. Multiple correlation coefficient is the degree of association between observed value of the dependent variable and its estimate obtained by multiple regression,
2. Multiple Correlation coefficient lies between 0 and 1,
3. If multiple correlation coefficient is 1, then association is perfect and multiple regression equation may said to be perfect prediction formula,
4. If multiple correlation coefficient is 0, dependent variable is uncorrelated with other independent variables. From this, it can be concluded that multiple regression equation fails to predict the value of dependent variable when values of independent variables are known,
5. Multiple correlation coefficient is always greater or equal than any total correlation coefficient. If $R_{1.23}$ is the multiple correlation coefficient than $R_{1.23} \geq r_{12}$ or r_{13} or r_{23}, and
6. Multiple correlation coefficient obtained by method of least squares would always be greater than the multiple correlation coefficient obtained by any other method.

11.4 SUMMARY

In this unit, we have discussed:

1. The multiple correlation, which is the study of joint effect of a group of two or more variables on a single variable which is not included in that group,
2. The estimate obtained by regression equation of that variable on other variables,
3. Limit of multiple correlation coefficient, which lies between 0 and +1,
4. The numerical problems of multiple correlation coefficient, and
5. The properties of multiple correlation coefficient.

11.5 SOLUTIONS / ANSWERS

E1) We have,

$$r_{12} = 0.6, \quad r_{23} = r_{31} = 0.54$$

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

$$= \frac{0.36 + 0.29 - 0.35}{0.71}$$

$$= \frac{0.30}{0.71} = 0.42$$

Then

$$R_{1.23} = 0.65$$

E2) We have

$$R^2_{2.13} = \frac{r^2_{12} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{13}}$$

$$= \frac{0.49 + 0.29 - 0.56}{1 - 0.55}$$

$$= \frac{0.22}{0.45} = 0.49$$

Thus

$$R_{2.13} = 0.70.$$

E3) We have

$r_{12} = 0.82$, $r_{23} = -0.57$ $r_{13} = -0.42$.

$$R^2_{1.23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2_{23}}$$

$$= \frac{0.67 + 0.18 - 0.39}{0.68}$$

$$= \frac{0.46}{0.68} = 0.68$$

Then

$$R_{1.23} = 0.82$$

$$R^2_{2.13} = \frac{r^2_{12} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{13}}$$

$$= \frac{0.67 + 0.32 - 0.39}{0.82}$$

$$= \frac{0.60}{0.82} = 0.73$$

Thus,

$$R_{2.13} = 0.85.$$

E4) To obtain multiple correlation coefficients $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$ we use following formulae:

$$R^2_{1.23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2_{23}},$$

$$R^2_{2.13} = \frac{r^2_{12} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{13}} \text{ and}$$

$$R^2_{3.12} = \frac{r^2_{13} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{12}}$$

We need r_{12}, r_{13} and r_{23} which are obtained from the following table:

S. No.	X_1	X_2	X_3	$(X_1)^2$	$(X_2)^2$	$(X_3)^2$	X_1X_2	X_1X_3	X_2X_3
1	22	12	13	484	144	169	264	286	156
2	15	15	16	225	225	256	225	240	240
3	27	17	12	729	289	144	459	324	204
4	28	15	18	784	225	324	420	504	270
5	30	42	22	900	1764	484	1260	660	924
6	42	15	20	1764	225	400	630	840	300
7	40	28	12	1600	784	144	1120	480	336
Total	204	144	113	6486	3656	1921	4378	3334	2430

Now, we get the total correlation coefficient r_{12}, r_{13} and r_{23}

$$r_{12} = \frac{N(\sum X_1X_2) - (\sum X_1)(\sum X_2)}{\sqrt{\{N(\sum X_1^2) - (\sum X_1)^2\}\{N(\sum X_2^2) - (\sum X_2)^2\}}}$$

$$r_{12} = \frac{(7 \times 4378) - (204) \times (144)}{\sqrt{\{(7 \times 6486) - (204) \times (204)\}\{(7 \times 3656) - (144) \times (144)\}}}$$

$$r_{12} = \frac{1270}{\sqrt{\{3786\}\{4856\}}}$$

$$= \frac{1270}{4287.75} = 0.30$$

$$r_{13} = \frac{N(\sum X_1X_3) - (\sum X_1)(\sum X_3)}{\sqrt{\{N(\sum X_1^2) - (\sum X_1)^2\}\{N(\sum X_3^2) - (\sum X_3)^2\}}}$$

$$r_{13} = \frac{(7 \times 3334) - (204) \times (113)}{\sqrt{\{(7 \times 6486) - (204 \times 204)\}\{(7 \times 1921) - (113 \times 113)\}}}$$

$$r_{13} = \frac{286}{\sqrt{3786 \times 678}}$$

$$= \frac{286}{1602.16} = 0.18$$

and

$$r_{23} = \frac{N(\sum X_2X_3) - (\sum X_2)(\sum X_3)}{\sqrt{\{N(\sum X_2^2) - (\sum X_2)^2\}\{N(\sum X_3^2) - (\sum X_3)^2\}}}$$

$$r_{23} = \frac{(7 \times 2430) - (144) \times (113)}{\sqrt{\{7 \times 3656) - (144 \times 144)\}\{(7 \times 1921) - (113 \times 113)\}}}$$

$$r_{23} = \frac{738}{\sqrt{\{4856\}\{678\}}}$$

$$= \frac{738}{1814.49} = 0.41$$

Now, we calculate $R_{1.23}$

We have, $r_{12} = 0.30$, $r_{13} = 0.18$ and $r_{23} = 0.41$, then

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

$$= \frac{0.30^2 + 0.18^2 - 2 \times .30 \times 0.18 \times 0.41}{1 - (0.41)^2}$$

$$= \frac{0.0781}{0.8319} = 0.9380$$

Then

$$R_{1.23} = 0.30.$$

$$R^2_{2.13} = \frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}$$

$$= \frac{0.30^2 + 0.41^2 - 2 \times .30 \times 0.18 \times 0.41}{1 - 0.18^2}$$

$$= \frac{0.2138}{0.9676} = 0.221$$

Thus,

$$R_{2.13} = 0.47$$

$$R^2_{3.12} = \frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}$$

$$= \frac{0.18^2 + 0.41^2 - 2 \times 0.30 \times 0.18 \times 0.41}{1 - 0.30^2}$$

$$= \frac{0.1562}{0.9100} = 0.1717$$

Thus,

$$R_{3.12} = 0.41$$

E5) To obtain multiple correlation coefficients $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$ we use following formulae

$$R^2_{1.23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2_{23}},$$

$$R^2_{3.12} = \frac{r^2_{13} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{12}}$$

and

$$R^2_{2.13} = \frac{r^2_{12} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{13}}$$

We need r_{12}, r_{13} and r_{23} which are obtained from the following table:

S. No.	X_1	X_2	X_3	$d_1 = X_1 - 50$	$d_2 = X_2 - 40$	$d_3 = X_3 - 70$	$(d_1)^2$	$(d_2)^2$	$(d_3)^2$	d_1d_2	d_1d_3	d_2d_3
1	50	42	72	0	2	2	0	4	4	0	0	4
2	54	46	71	4	6	1	16	36	1	24	4	6
3	50	45	73	0	5	3	0	25	9	0	0	15
4	56	44	70	6	4	0	36	16	0	24	0	0
5	50	40	72	0	0	2	0	0	4	0	0	0
6	55	45	72	5	5	2	25	25	4	25	10	10
7	52	43	70	2	3	0	4	9	0	6	0	0
8	50	42	71	0	2	1	0	4	1	0	0	2
9	52	41	75	2	1	5	4	1	25	2	10	5
10	51	42	71	1	2	1	1	4	1	2	1	2
Total				20	30	17	86	124	49	83	25	44

Now, we get the total correlation coefficient r_{12}, r_{13} and r_{23}

$$r_{12} = \frac{N(\sum d_1d_2) - (\sum d_1)(\sum d_2)}{\sqrt{\left\{N(\sum d_1^2) - (\sum d_1)^2\right\}\left\{N(\sum d_2^2) - (\sum d_2)^2\right\}}}$$

$$r_{12} = \frac{(10 \times 83) - (20) \times (30)}{\sqrt{\{(10 \times 86) - (20) \times (20)\}\{(10 \times 124) - (30) \times (30)\}}}$$

$$r_{12} = \frac{230}{\sqrt{460 \times 340}}$$

$$= \frac{230}{395.47} = 0.58$$

$$r_{13} = \frac{N(\sum d_1 d_3) - (\sum d_1)\ (\sum d_3)}{\sqrt{\left\{N(\sum d_1^2) - (\sum d_1)^2\right\}\left\{N(\sum d_3^2) - (\sum d_3)^2\right\}}}$$

$$r_{13} = \frac{(10 \times 25) - (20) \times (17)}{\sqrt{\{(10 \times 86) - (20 \times 20)\}\ \{(10 \times 49) - (17 \times 17)\}}}$$

$$r_{13} = \frac{-90}{\sqrt{\{460\}\{201\}}}$$

$$= \frac{-90}{304.07} = -\ 0.30$$

and

$$r_{23} = \frac{N(\sum X_2 X_3) - (\sum X_2)(\sum X_3)}{\sqrt{\left\{N(\sum X_2^2) - (\sum X_2)^2\right\}\left\{N(\sum X_3^2) - (\sum X_3)^2\right\}}}$$

$$r_{23} = \frac{(10 \times 44) - (30) \times (17)}{\sqrt{\{(10 \times 124) - (20 \times 20)\}\{(10 \times 49) - (17 \times 17)\}}}$$

$$r_{23} = \frac{-70}{\sqrt{\{340\}\{201\}}}$$

$$= \frac{-70}{261.42} = -\ 0.27$$

Now, we calculate $R_{1.23}$

We have, $r_{12} = 0.58$, $r_{13} = -\ 0.30$ and $r_{23} = -\ 0.27$, then

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

$$= \frac{0.58^2 + (-\ 0.30)^2 - 2 \times 0.58 \times (-\ 0.30) \times (-0.27)}{1 - (-\ 0.27)^2}$$

$$= \frac{0.3324}{0.9271} = 0.36$$

Then

$$R_{1.23} = 0.60\ .$$

$$R^2_{2.13} = \frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}$$

$$= \frac{0.58^2 + (-0.27)^2 - 2 \times 0.58 \times (-0.30) \times (-0.27)}{1 - (-0.30)^2}$$

$$= \frac{0.3153}{0.9100} = 0.35$$

Thus,

$$R_{2.13} = 0.59$$

$$R^2_{3.12} = \frac{r^2_{13} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{12}}$$

$$= \frac{(-0.30)^2 + (-0.27)^2 - 2 \times 0.58 \times (-0.30) \times (-0.27)}{1 - (0.58)^2}$$

$$= \frac{0.0689}{0.6636} = 0.10$$

Thus,

$$R_{3.12} = 0.32$$