Multiple and Partial Correlation

$\boxed{\text{Correlation}}$ → Mutiple Correlation.
→ Partial Correlation.

Multiple correlation :-

To measure a variable x_1 which dependent on more than one variables say x_2 and x_3.

Partial Correlation:-

To examine the influence of one variable upon another, after eliminating the effects of third (or) other variables.

Formulas:

Multiple Correlation Co-efficient

If x_1 depends on x_2 and x_3 then,

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

If x_2 depends on x_1 and x_3 then,

$$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{21}r_{23}r_{13}}{1 - r_{13}^2}}$$

If x_3 depends on x_1 and x_2

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{31}r_{32}r_{12}}{1 - r_{12}^2}}$$

Properties:

1) $0 \leq R_{1.23} \leq 1$

2) $R_{1.23} \geq r_{12}$ and $R_{1.23} \geq r_{13}$.

3) If $R_{1.23} = 0$ then $r_{12} = 0$ & $r_{13} = 0$.

4) $R_{1.23} = R_{1.32}$

but $R_{1.23} \neq R_{2.13}$.

5) $r_{11} = r_{22} = r_{33} = 1$.

Partial Correlation Coefficient

*) Correlation btwn x_1 and x_2, x_3 kept constant

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

*) Correlation btwn x_1 and x_3, x_2 kept constant.

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{32}}{\sqrt{(1-r_{12}^2)(1-r_{32}^2)}}$$

*) Correlation b/w x_2 and x_3, x_1 kept constant.

$$r_{23.1} = \frac{r_{23} - r_{21} \cdot r_{31}}{\sqrt{(1-r_{21}^2)(1-r_{31}^2)}}$$

Note: $-1 \le r_{12.3} \le 1$, similarly for $r_{13.2}$ and $r_{23.1}$.

1. Given, $r_{12} = 0.70$, $r_{13} = 0.61$, $r_{23} = 0.40$, calculate: $r_{23.1}$, $r_{13.2}$, and $r_{12.3}$.

Solution: We know that

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

$$\therefore \quad r_{23.1} = \frac{0.4-(0.7\times 0.61)}{\sqrt{1-(0.7)^2}\,\sqrt{1-(0.61)^2}} = \frac{0.4 - 0.427}{\sqrt{0.51}\sqrt{0.6279}} = \frac{-0.027}{0.714\times 0.792} = -\,0.048$$

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

$$\therefore \quad r_{13.2} = \frac{0.61-(0.7\times 0.4)}{\sqrt{1-(0.7)^2}\,\sqrt{1-(0.4)^2}} = \frac{0.61-0.28}{\sqrt{1-0.49}\sqrt{1-0.16}} = \frac{0.33}{\sqrt{0.51}\sqrt{0.84}} = 0.504$$

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

$$\therefore \quad r_{12.3} = \frac{0.7-(0.61\times 0.4)}{\sqrt{(1-(0.61)^2)}\sqrt{1-(0.4)^2}} = \frac{0.7-0.244}{\sqrt{1-0.3721}\sqrt{1-0.16}} = \frac{0.456}{\sqrt{0.6279}\sqrt{0.84}}$$

$$= \frac{0.456}{0.726} = 0.628.$$

2. The following zero-order correlation coefficients are given:

$$r_{12} = 0.98,\ r_{13} = 0.44, \text{ and } r_{23} = 0.54$$

Calculate the partial correlation coefficient between first and the third variables keeping the effect of second variable constant. Also, calculate multiple correlation coefficient.

Solution: The partial correlation coefficient between first and the third variables keeping the effect of second variable constant is given by $r_{13.2}$. Therefore, we have

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

$$= \frac{0.44 - (0.98)(0.54)}{\sqrt{1 - (0.98)^2}\sqrt{1 - (0.54)^2}} = \frac{0.44 - 0.5292}{\sqrt{1 - 0.9604}\sqrt{1 - 0.2916}}$$

$$= \frac{-0.0892}{\sqrt{0.0396}\sqrt{0.7084}} = \frac{-0.0892}{0.199 \times 0.842} = \frac{-0.0892}{0.1676} = -\,0.5322$$

Multiple correlation coefficient, treating first variable as dependent and second and third as independent is given by

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

$$= \sqrt{\frac{(0.98)^2 + (0.44)^2 - 2(0.98)(0.44)(0.54)}{1-(0.54)^2}} = \sqrt{\frac{0.9604 + 0.1936 - 0.4657}{1 - 0.2916}}$$

$$= \sqrt{\frac{1.154 - 0.4657}{0.7084}} = \sqrt{\frac{0.6883}{0.7084}} = \sqrt{0.9716} = 0.9857.$$

Problem 3:

$$\sigma_1 = 3,\ \sigma_2 = \sigma_3 = 5,\ r_{12} = 0.6, \text{ and } r_{23} = r_{31} = 0.8$$

Find (a) $r_{23.1}$ and (b) $R_{1.23}$.

Solution:

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

$$= \frac{0.8-(0.6)(0.8)}{\sqrt{1-(0.6)^2}\,\sqrt{1-(0.8)^2}} = \frac{0.8-0.48}{\sqrt{0.64}\,\sqrt{0.36}} = \frac{0.32}{0.48} = 0.667.$$

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

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

$$= \sqrt{\frac{0.36+0.64-0.768}{0.36}} = \sqrt{\frac{0.232}{0.36}} = \sqrt{0.6444} = 0.8028.$$

Problem 4:

Is there any consistency in the following data.

$$r_{12} = 0.6,\ r_{13} = -0.4,\ r_{23} = 0.8.$$

Solution: We know that

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

$$\therefore \quad r_{12.3} = \frac{0.6 - (-0.4) \times 0.8}{\sqrt{\{1 - (-0.4)^2\}\{1 - (0.8)^2\}}} = \frac{0.6 + 0.32}{\sqrt{(1 - 0.16)(1 - 0.64)}} = \frac{0.92}{0.549} = 1.6757$$

Since, $r_{12.3} > 1$, there is inconsistency in the data.

Problem 5:

. On the basis of observations made on 39 cotton plants, the total correlation of yield of cotton (Y_1), number of bolls, *i.e.*, seed vessels (Y_2) and height (Y_3) are found to be:

$$r_{12} = 0.8, \quad r_{13} = 0.65, \quad \text{and} \quad r_{23} = 0.7$$

Comment on the partial correlation between yield of cotton and the number of bolls, eliminating the effect of height.

Solution. We are given

$$r_{12} = 0.8, \quad r_{13} = 0.65, \quad \text{and} \quad r_{23} = 0.7$$

We have to find the partial correlation between yield of cotton and the number of bolls, eliminating the effect of height, *i.e.*, in terms of symbols, we have to calculate $r_{12.3}$, which is given by

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

$$\therefore \quad r_{12.3} = \frac{0.8 - (0.65 \times 0.7)}{\sqrt{1 - (0.65)^2}\sqrt{1 - (0.7)^2}} = \frac{0.8 - 0.455}{\sqrt{1 - 0.4225}\sqrt{1 - 0.49}}$$

$$= \frac{0.345}{\sqrt{0.5775}\,\sqrt{0.51}} = \frac{0.345}{0.543} = 0.635.$$

Multiple Regression

Multiple regression:

To estimate the value of one variable from those of several others.

Formulas:

Multiple regression equations

*) The multiple regression equation of y_1 on y_2 and y_3 is given by,

$$\boxed{y_1 = a_{1.23} + b_{12.3}\, y_2 + b_{13.2}\, y_3}$$

The values of $b_{12.3}$, $b_{13.2}$ and $a_{1.23}$ are determined by two methods.

Method 1:-

By solving three normal equts. ①

Method 2:-

By using mean.

Y_1 on Y_2 and Y_3

The multiple regression equation of y_1 on y_2 and y_3 is given by,

$$\boxed{y_1 = a_{1.23} + b_{12.3}\,y_2 + b_{13.2}\,y_3}$$

Method 1: Normal equations

The values of $b_{12.3}$ and $b_{13.2}$ are determined by solving simultaneously the following three normal equations.

$$\Sigma Y_1 = n a_{1.23} + b_{12.3}\,\Sigma Y_2 + b_{13.2}\,\Sigma Y_3$$

$$\Sigma Y_1 Y_2 = a_{1.23}\,\Sigma Y_2 + b_{12.3}\,\Sigma Y_2^2 + b_{13.2}\,\Sigma Y_2 Y_3$$

$$\Sigma Y_1 Y_3 = a_{1.23}\,\Sigma Y_3 + b_{12.3}\,\Sigma Y_2 Y_3 + b_{13.2}\,\Sigma Y_3^2$$

Y_2 on Y_1 and Y_3

The multiple regression equation of Y_2 on Y_1 and Y_3 is given by

$$Y_2 = a_{2.13} + b_{21.3}Y_1 + b_{23.1}Y_3$$

The values of $b_{21.3}$ and $b_{23.1}$ are determined by solving simultaneously the following three normal equations.

$$\Sigma Y_2 = n\,a_{2.13} + b_{21.3}\,\Sigma Y_1 + b_{23.1}\,\Sigma Y_3$$

$$\Sigma Y_1 Y_2 = a_{2.13}\,\Sigma Y_1 + b_{21.3}\,\Sigma Y_1^2 + b_{23.1}\,\Sigma Y_1 Y_3$$

$$\Sigma Y_2 Y_3 = a_{2.13}\,\Sigma Y_3 + b_{21.3}\,\Sigma Y_1 Y_3 + b_{23.1}\,\Sigma Y_3^2$$

Y_3 on Y_1 and Y_2

The multiple regression equation of Y_3 on Y_1 and Y_2 is given by

$$Y_3 = a_{3.12} + b_{31.2} Y_1 + b_{32.1} Y_2$$

The values of $b_{31.2}$ and $b_{32.1}$ are determined by solving simultaneously the following three normal equations.

$$\Sigma Y_3 = n a_{3.12} + b_{31.2} \Sigma Y_1 + b_{32.1} \Sigma Y_2$$

$$\Sigma Y_1 Y_3 = a_{3.12} \Sigma Y_1 + b_{31.2} \Sigma Y_1^2 + b_{32.1} \Sigma Y_1 Y_3$$

$$\Sigma Y_2 Y_3 = a_{3.12} \Sigma Y_2 + b_{31.2} \Sigma Y_1 Y_2 + b_{32.1} \Sigma Y_2^2$$

1. We are given

$$\bar{Y}_1 = 28.02,\ \bar{Y}_2 = 4.91,\ \bar{Y}_3 = 594$$

$$\sigma_1 = 4.4, \sigma_2 = 1.1, \sigma_3 = 80$$

$$r_{12} = 0.8, r_{23} = -0.56, r_{31} = -0.4$$

Find the correlation coefficients $r_{23.1}$ and $R_{1.23}$.

Solution. We know that

$$r_{23.1} = \frac{r_{23} - r_{13}r_{12}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{12}^2}} \text{ and } R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$$

Substituting the values, we get

$$r_{23.1} = \frac{-0.56-(-0.4)(0.8)}{\sqrt{1-(-0.4)^2}\sqrt{1-(0.8)^2}} = \frac{-0.56+0.32}{\sqrt{0.84}\sqrt{0.36}} = \frac{-0.24}{0.55} = -0.44$$

$$R_{1.23} = \sqrt{\frac{(0.8)^2 + (-0.4)^2 - 2(0.8)(-0.4)(-0.56)}{1-(-0.56)^2}}$$

$$= \sqrt{\frac{0.64+0.16-0.3584}{1-0.3136}} = \sqrt{\frac{0.4416}{0.6864}} = \sqrt{0.6433} = 0.80$$

2. Find the multiple linear regression of Y_1 on Y_2 and Y_3 from the data relating to these variables given below.

Y_1:	11	17	26	28	31	35	41	49	63	69
Y_2:	2	4	6	5	8	7	10	11	13	14
Y_3:	2	3	4	5	6	7	9	10	11	13

Solution: The regression equation of Y_1 on Y_2 and Y_3 is given by

$$Y_1 = a_{1.23} + b_{12.3}Y_2 + b_{13.2}Y_3$$

The values of the constants $a_{1.23}$, $b_{12.3}$, and $b_{13.2}$ are obtained by solving the following three normal equations.

$$\Sigma Y_1 = n\,a_{1.23} + b_{12.3}\,\Sigma Y_2 + b_{13.2}\,\Sigma Y_3$$

$$\Sigma Y_1 Y_2 = a_{1.23}\,\Sigma Y_2 + b_{12.3}\,\Sigma Y_2^2 + b_{13.2}\,\Sigma Y_2 Y_3$$

$$\Sigma Y_1 Y_3 = a_{1.23}\,\Sigma Y_3 + b_{12.3}\,\Sigma Y_2 Y_3 + b_{13.2}\,\Sigma Y_3^2$$

Computation of required values

Y_1	Y_2	Y_3	Y_1Y_2	Y_1Y_3	Y_2Y_3	Y_1^2	Y_2^2	Y_3^2
11	2	2	22	22	4	121	4	4
17	4	3	68	51	12	289	9	16
26	6	4	156	104	24	676	16	36
28	5	5	140	140	25	784	25	25
31	8	6	248	186	48	961	36	64
35	7	7	245	245	49	1225	49	49
41	10	9	410	369	90	1681	81	100
49	11	10	539	490	110	2401	100	121
63	13	11	819	693	143	3969	121	169
69	14	13	966	897	182	4761	169	196
$\Sigma Y_1 = 370$	$\Sigma Y_2 = 80$	$\Sigma Y_3 = 70$	$\Sigma Y_1Y_2 = 3613$	$\Sigma Y_1Y_3 = 3197$	$\Sigma Y_2Y_3 = 687$	$\Sigma Y_1^2 = 16868$	$\Sigma Y_2^2 = 780$	$\Sigma Y_3^2 = 610$

Substituting the values in the normal equations, we have

$$370 = 10a_{1.23} + 80b_{12.3} + 70b_{13.2} \quad \text{... (i)}$$

$$3613 = 80a_{1.23} + 780b_{12.3} + 687b_{13.2} \quad \text{... (ii)}$$

$$3197 = 70a_{1.23} + 687b_{12.3} + 610b_{13.2} \quad \text{... (iii)}$$

Dividing each equation by the coefficient of $b_{12.3}$, we get

$$0.125a_{1.23} + b_{12.3} + 0.875b_{13.2} = 4.625 \quad \text{... (iv)}$$

$$0.103a_{1.23} + b_{12.3} + 0.881b_{13.2} = 4.632 \quad \text{... (v)}$$

$$0.102a_{1.23} + b_{12.3} + 0.888b_{13.2} = 4.654 \quad \text{... (vi)}$$

Substracting equation (iv) from (v) and from (vi), we get

$$-0.022a_{1.23} + 0.006b_{13.2} = 0.007 \quad \text{... (vii)}$$

$$-0.001a_{1.23} + 0.007b_{13.2} = 0.022 \quad \text{... (viii)}$$

Multiplying (vii) and (viii) by 1000, we get

$$-22a_{1.23} + 6b_{13.2} = 7 \quad \text{... (ix)}$$

$$-a_{1.23} + 7b_{13.2} = 22 \quad \text{... (x)}$$

Solving these equations, we get

$$a_{1.23} = 0.561,\ b_{12.3} = 1.735,\ \text{and}\ b_{13.2} = 3.223$$

Hence, the required equation is

$$Y_1 = 0.561 + 1.735Y_2 + 3.223Y_3$$

Problem 3:

X	11	17	26	28	31	35	41	49	63
Y	12	14	16	20	23	27	30	35	43
Z	22	35	37	46	49	52	57	60	71

Find the followings for the above data

1. Multiple Correlation between the variables
2. Partial correlation between the variables
3. Multiple regression between the variables