Frequently Asked Questions

1. When do we use Partial or Multiple correlations?
   Answer:
   Partial correlation and multiple correlations are used when we have more than two variables.

2. What do you mean by partial correlation coefficient?
   Answer:
   Sometimes the correlation between two variables X_1 and X_2 may be partly due to the correlation of a third variable, X_3 with both X_1 and X_2. In such a situation, one may want to know what the correlation between X_1 and X_2 would be if the effect of X_3 on each of X_1 and X_2 were eliminated. This correlation is called the partial correlation and the correlation coefficient between X_1 and X_2 the linear effect of X_3 on each of them has been eliminated is called the partial correlation coefficient.

3. What do you mean by multiple correlation coefficient?
   Answer:
   In tri-variate distribution in which each of the variables X_1, X_2, and X_3 has N observations, the multiple correlation coefficient between X_1 on X_2 and X_3 is, usually denoted by $R_{1.23}$ is the simple correlation coefficient between X_1 and the joint effect of X_2 and X_3 on X_1. In other words $R_{1.23}$ is the correlation coefficient between X_1 and its estimated value as given by the plane of regression of X_1 on X_2 and X_3

4. Write the formula for finding the partial correlation coefficient.
   Answer:
   The formula for the partial correlation coefficient is given by,
   $$r_{12.3} = \frac{(r_{12} - r_{13}r_{23})}{\sqrt{(1-r_{13}^{\ 2})(1-r_{23}^{\ 2})}}$$
   Similarly we can obtain $r_{13.2}$ and $r_{23.1}$

5. Write the formula for obtaining the multiple correlation coefficient.

   Answer:
   The formula for finding the multiple correlation coefficient is given by,
   $$R_{1.23}^{\ 2} = \frac{r_{12}^{\ 2} + r_{13}^{\ 2} - 2r_{12}r_{13}r_{23}}{1 - r_{23}^{\ 2}}$$
   Similarly we can obtain the formula for $R_{2.13}^2$ and $R_{3.12}^2$.

   Square root the expression gives the multiple correlation coefficient.

6. Write the formula for obtaining the partial regression coefficient.

Answer:
The formula for finding the partial regression coefficient is given by,

$$b_{12.3} = -\frac{\sigma_1 \omega_{12}}{\sigma_2 \omega_{11}}$$

7. Write the expression which gives the relation between partial and multiple correlation coefficient.
Answer:
$1-R_{1.23}^2 = (1-r_{12}^2)(1-r_{13.2}^2)$

8. If $r_{12}=0.28$, $r_{23}=0.49$ and $r_{13}=0.51$, find $r_{23.1}$ and $r_{13.2}$
Answer:

$$r_{23.1} = \frac{(r_{23} - r_{21}r_{31})}{\sqrt{(1-r_{21}^2)(1-r_{31}^2)}} = \frac{(0.49 - 0.28\times 0.51)}{\sqrt{(1-0.28^2)(1-0.51^2)}} = 0.42$$

$$r_{13.2} = \frac{(r_{13} - r_{12}r_{32})}{\sqrt{(1-r_{12}^2)(1-r_{32}^2)}} = \frac{(0.51 - 0.28\times 0.49)}{\sqrt{(1-0.28^2)(1-0.49^2)}} = 0.446$$

9. From the data relating to the yield of dry bark (X_1), height (X_2) and girth (X_3) for 18 cinchona plants, the following correlation coefficients are obtained.
$r_{12} = 0.77$, $r_{13} = 0.72$, $r_{23} = 0.52$
Find the partial correlation coefficient $r_{12.3}$ and multiple correlation coefficient $R_{1.23}$.

Answer:
We know that, partial correlation coefficient

$$r_{12.3} = \frac{(r_{12} - r_{13}r_{23})}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{(0.77 - 0.72\times 0.52)}{\sqrt{(1-0.72^2)(1-0.52^2)}} = 0.62$$

To find multiple correlation coefficient, we have,

$$R_{1.23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2} = \frac{0.77^2 + 0.72^2 - 2\times 0.77\times 0.72\times 0.52}{1-0.52^2} = 0.7334$$

$R_{1.23} = +0.8564$ (since multiple correlation coefficient is non-negative)

10. Find r_{12} if $r_{13.2}=0.6$ and $R_{1.23}=0.8$
Answer:
We know that $1-R_{1.23}^2 = (1-r_{12}^2)(1-r_{13.2}^2)$
By substitution, we get,
$1-(0.8)^2 = (1-r_{12}^2)(1-(0.6)^2)$
On simplification, we get, $r_{12}^2=0.4375$ or $r_{12}=\pm 0.6614$

11. Is it possible to get the following set of experimental data?
$r_{23}=0.8$, $r_{31}=-0.5$, $r_{12}=0.6$
Answer:

$$r_{12.3} = \frac{(r_{12} - r_{13}r_{23})}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{(0.6 - (-0.5\times 0.8))}{\sqrt{(1-(-0.5^2)(1-0.8^2)}} = 1.9246 > 1$$

Hence the given set of experimental data is wrong.

12. In a tri-variate distribution, $\sigma_1 = 2$, $\sigma_2 = \sigma_3 = 3$, $r_{12} = 0.7$, $r_{23} = r_{31} = 0.5$

Find

i. $r_{23.1}$

ii. $R_{1.23}$

iii. $b_{12.3}$ and $b_{21.3}$ and hence $r_{12.3}$

iv. $\sigma_{1.23}$

Answer:

i. We know that, partial correlation coefficient

$$r_{23.1} = \frac{(r_{23} - r_{21}r_{31})}{\sqrt{(1-r_{21}{}^2)(1-r_{31}{}^2)}} = \frac{(0.5 - 0.7\times 0.5)}{\sqrt{(1-0.7^2)(1-0.5^2)}} = 0.2425$$

ii. To find multiple correlation coefficient, we have,

$$R_{1.23}{}^2 = \frac{r_{12}{}^2 + r_{13}{}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}{}^2} = \frac{0.7^2 + 0.5^2 - 2\times 0.7\times 0.5\times 0.5}{1-0.5^2} = 0.52$$

$R_{1.23}$ =+0.7211 (since multiple correlation coefficient is non-negative)

iii. We know that,

$$b_{12.3} = -\frac{\sigma_1\omega_{12}}{\sigma_2\omega_{11}}, b_{21.3} = -\frac{\sigma_2\omega_{21}}{\sigma_1\omega_{22}}$$

Where $\omega = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0.7 & 0.5 \\ 0.7 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{vmatrix}$

From above determinant, ω_{11}=1-(0.5x0.5)=0.75

ω_{12}=-(0.7-(0.5x0.5))=-0.45

ω_{21}=-(0.7-0.5x0.5)=-0.45

ω_{22}=1-0.25=0.75

Therefore

$$b_{12.3} = -\frac{2\times(-0.45)}{3\times 0.75} = 0.4, b_{21.3} = -\frac{3\times(-0.45)}{2\times 0.75} = 0.9$$

we know that form partial regression coefficients we get partial correlation coefficient as follows

$$r_{12.3} = \sqrt{b_{12.3}b_{21.3}} = \sqrt{0.4\times 0.9} = 0.6$$

iv. We have $\sigma_{1.23} = \sigma_1 \sqrt{\frac{\omega}{\omega_{11}}}$

Where ω is defined as earlier.

The value of ω is given by,

$\omega = 1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2 r_{12} r_{13} r_{23}$

by substituting and simplifying, we get

$\omega = 0.36$ and we have already found that,

$\omega_{11} = 0.75$.

Therefore, $\sigma_{1.23} = 2\sqrt{\frac{0.36}{0.75}} = 2 \times \sqrt{0.48} = 1.3856$

13. Find the regression equation of X_1 on X_2 and X_3 given the following results

Trait	Mean	S. D	r_{12}	r_{23}	r_{31}
X_1	28.02	4.42	+0.8	-	-
X_2	4.91	1.1	-	-0.56	-
X_3	594	85	-	-	-0.4

Where X_1 = seed per acre, X_2 = Rainfall in inches

X_3=Accumulated temperature above 42^0F

Answer:

Regression equation of X1 on X2 and X3 is given by,

$$\frac{\omega_{11}}{\sigma_1}(X_1-\overline{X}_1)+\frac{\omega_{12}}{\sigma_2}(X_2-\overline{X}_2)+\frac{\omega_{13}}{\sigma_3}(X_3-\overline{X}_3)=0$$

Where $\omega=\begin{vmatrix}1 & r_{12} & r_{13}\\ r_{21} & 1 & r_{23}\\ r_{31} & r_{32} & 1\end{vmatrix}$

$$\omega_{11}=\begin{vmatrix}1 & r_{23}\\ r_{32} & 1\end{vmatrix}=1-r_{23}^{\ 2}=1-(-0.56)^2=0.686$$

Similarly, $\omega_{12}=r_{13}r_{23}-r_{21}$ = (-0.4) (-0.56)-(0.8) +=-0.576

$\omega_{13}=r_{23}r_{12}-r_{13}$ = (-0.56)(0.8)-(-0.4)=-0.048

Therefore required equation of plane of regression of X_1 on X_2 and X_3 is given by

$$\frac{0.686}{4.42}(X_1-28.02)+\frac{-0.576}{1.1}(X_2-4.91)+\frac{-0.048}{85}(X_3-594)=0$$

14. Five hundred students were examined in three subjects I, II and III, each subject carrying 100 marks.

The Coefficient of Correlation between pairs of subjects is given by,

r_{12}=0.6, r_{13}=0.7, r_{23}=0.8

Assuming that the marks are normally distributed

i. Find the correlation between marks in subjects I and II among students who scored equal marks in subject III
ii. If r_{23} was not known, obtain the limits within which it may lie from the value of r_{12} and r_{13} (ignoring sampling error)

Answer:

The correlation between marks in subjects I and II among students who scored equal marks in subject III is $r_{12.3}$ and is given by,

$$r_{12.3} = \frac{(r_{12} - r_{13}r_{23})}{\sqrt{(1 - r_{13}{}^2)(1 - r_{23}{}^2)}} = \frac{(0.6 - 0.7 \times 0.8)}{\sqrt{(1 - 0.7^2)(1 - 0.8^2)}} = 0.0934$$

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

Since r_{23} is not known, we consider it as 'a'

$$\therefore \frac{(0.6 - 0.7a)^2}{(1 - 0.49)(1 - a^2)} \leq 1$$

Implies, $0.36+0.49a^2 -0.84a \leq 0.51 (1 - a^2)$

Implies $a^2 -0.84a -0.15 \leq 0$

Thus 'a' lies between the roots of the equation $a^2 -0.84a -0.15 = 0$ which are 0.99 and -0.15.

Hence r_{23} should lie between -0.15 and 0.99

15. A teacher in mathematics wishes to determine the relationship of marks on final examination to those on two tests given during the semester. Calling X_1, X_2 and X_3, the marks of a student in 1st, 2nd and final examination respectively, he made the following computations from a total of 120 students.

$\overline{X}_1 = 6.8, \overline{X}_2 = 7.0, \overline{X}_3 = 74$

$\sigma_1=1$ $\sigma_2=0.8$ $\sigma_3=9$

r12=0.6, r13 = 0.7 and r23 = 0.65.

Find the least square regression equation of X_3 on X_1 and X_2

Estimate the final marks of two students who scored respectively 9 and 7, 4 and 8 on two tests.

Answer:
The least square regression equation of X_3 on X_1 and X_2

$X_3= a_{3.12} + b_{31.2}X_1 + b_{32.1}X_2$

The regression equation of X_3 on X_1 and X_2 in the deviation form can be written as

$x_3= b_{31.2}x_1 + b_{32.1}x_2$

Where $x_3 = X_3 - \overline{X_3}, x_1 = X_1 - \overline{X_1}, x_2 = X_2 - \overline{X_2}$

$$b_{31.2} = \frac{\sigma_3}{\sigma_1}\frac{r_{31} - r_{32}r_{12}}{1 - r_{12}{}^2} = \frac{9}{1}\frac{0.7 - (0.65)(0.6)}{1 - 0.6^2} = 4.36$$

$$b_{32.1} = \frac{\sigma_3}{\sigma_2}\frac{r_{31} - r_{32}r_{12}}{1 - r_{12}{}^2} = \frac{9}{0.8}\frac{0.65 - (0.7)(0.6)}{1 - 0.6^2} = 4.04$$

The regression equation of X_3 on X_1 and X_2 is given by

X_3 -74 = 4.36(X_1 – 6.8) + 4.04 (X_2 – 7)

Implies, X_3 =16.07 + 4.36X_1 + 4.04X_2

Final marks of students who scored 9 and 7 marks, i.e., substituting X_1=9 and X=7, in the above regression equation, we get,

X_3=16.07+ 4.36(9) +4.04(7) =83.59 or 84.

Final marks of students who scored 4 and 8 marks, i.e., substituting X_1=4 and X_2=8, in the above regression equation, we get,

X_3=16.07+ 4.36(4) +4.04(8) =65.8 or 66.