# Frequently Asked Questions

1. Can we use multiple correlation when we have two variables?

**Answer:**

Multiple correlation is used when there is tri-variate or multivariate distribution. When there are two variables, we use total correlation coefficient given by Karl Pearson.

2. Give an example for multiple correlation coefficient.

**Answer:**

If there is interrelation between many variables and the value of one variable may be influence by many others, example, the yield of crop per acre, say $X_1$ depends upon quality of seed say $X_2$, fertility of soil, say $X_3$ , fertilizer used say $X_4$, irrigation facilities say $X_5$, weather conditions say $X_6$ and so on, then we use multiple 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. Give an expression for multiple correlation coefficient in terms of total correlation coefficients.

**Answer:**

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

5. Derive an expression for 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 $X_3$ namely,

$e_{1.23} = b_{12.3} X_2 + b_{13.2} X_3$.

We have $X_{1.23} = X_1 - b_{12.3} X_2 - b_{13.2} X_3 = X_1 - e_{1.23}$

Implies, $e_{1.23} = X_1 - X_{1.23}$.

Since $X_i$'s are measured from their respective means, we have,

$E(X_{1.23}) = 0$ and $E(e_{1.23}) = 0$. (Since $E(X_i = 0; i = 1, 2, 3)$.

By definition, $R_{1.23} = \dfrac{Cov(X_1, e_{1.23})}{\sqrt{V(X_1)V(e_{1.23})}}$

Cov( X1, $e_{1.23}$ )= E[{X1 – E(X1)}{$e_{1.23}$ – E($e_{1.23}$)}]

= E(X1 $e_{1.23}$ ) $= \frac{1}{N}\Sigma X_1 e_{1.23} = \frac{1}{N}\Sigma X_1 (X_1 - X_{1.23})$

$= \frac{1}{N}\Sigma X_1^2 - \frac{1}{N}\Sigma X_1 X_{1.23} = \frac{1}{N}\Sigma X_1^2 - \frac{1}{N}\Sigma X_{1.23}^2 = \sigma_1^2 - \sigma_{1.23}^2$

Also,

V($e_{1.23}$ )= E($e_{1.23}$ $^2$) $= \frac{1}{N}\Sigma e_{1.23}^2 = \frac{1}{N}\Sigma (X_1 - X_{1.23})^2 = \frac{1}{N}\Sigma (X_1^2 + X_{1.23}^2 - 2X_1 X_{1.23})$

$= \frac{1}{N}\Sigma X_1^2 + \frac{1}{N}\Sigma X_{1.23}^2 - \frac{2}{N}\Sigma X_1 X_{1.23} = \frac{1}{N}\Sigma X_1^2 + \frac{1}{N}\Sigma X_{1.23}^2 - \frac{2}{N}\Sigma X_{1.23}^2$

$= \sigma_1^2 - \sigma_{1.23}^2$

Substituting in $R_{1.23}$ we get,

$$R_{1.23} = \frac{\sigma_1^2 - \sigma_{1.23}^2}{\sqrt{\sigma_1^2(\sigma_1^2 - \sigma_{1.23}^2)}} \Rightarrow R_{1.23}^2 = \frac{\sigma_1^2 - \sigma_{1.23}^2}{\sigma_1^2} = 1 - \frac{\sigma_{1.23}^2}{\sigma_1^2} \Rightarrow 1 - R_{1.23}^2 = \frac{\sigma_{1.23}^2}{\sigma_1^2}$$

6. Give an expression for multiple correlation coefficient in terms of partial and total correlation coefficient.
   **Answer:**

$$1 - R_{1.23}^2 = (1 - r_{12}^2)(1 - r_{13.2}^2)$$

7. Write the properties of Multiple Correlation coefficient.
   **Answer:**
   1. Multiple correlation coefficient measures the closeness of the association between the observed values and the expected values of a variable obtained from the multiple linear regression of that variable on other variables.
   2. Multiple correlation coefficient between observed values and expected values when the expected values are calculated form a linear relation of the variables determined by the method of least square, is always greater than that where expected values are calculated from any other linear combination of the variables.
   3. Since $R_{1.23}$ is the simple correlation between $X_1$ and $e_{1.23}$, it must lie between - 1 and +1. But $R_{1.23}$ is a non-negative quantity, $0 \le R_{1.23} \le 1$.
   4. If $R_{1.23}=1$ then association is perfect and all the regression residuals are zero, and as such $\sigma_{1.23}^2 = 0$. In this case, since X1 = $e_{1.23}$, the predicted value of X1, the multiple linear regression equation of X1 on X2 and X3 may be said to be a perfect prediction formula.
   5. If $R_{1.23}=0$ then all total and partial correlations involving X1 are zero. So X1 is completely uncorrelated with all the other variables in this case and the multiple regression equation fails to throw any light on the value of X1 when X2 and X3 are known.
   6. $R_{1.23}$ is not less than any total correlation coefficient. ie., $R_{1.23} > r_{12}$ or $r_{13}$ or $r_{23}$.

8. Show that $1-R_{1.23}{}^2=(1-r_{12}{}^2)(1-r_{13.2}{}^2)$.

**Answer:**

We know that $1-R_{1.23}{}^2=1-\dfrac{r_{12}{}^2+r_{13}{}^2-2r_{12}r_{13}r_{23}}{1-r_{23}{}^2}=\dfrac{1-r_{12}{}^2-r_{23}{}^2-r_{13}{}^2+2r_{12}r_{13}r_{23}}{1-r_{23}{}^2}$

Also $1-r_{13.2}{}^2=1-\dfrac{(r_{13}{}^2+r_{12}r_{23})^2}{(1-r_{12}{}^2)(1-r_{23}{}^2)}=\dfrac{1-r_{12}{}^2-r_{23}{}^2-r_{13}{}^2+2r_{12}r_{13}r_{23}}{(1-r_{12}{}^2)(1-r_{23}{}^2)}$

$\Rightarrow 1-R_{1.23}{}^2=(1-r_{12}{}^2)(1-r_{13.2}{}^2)$

9. From result in question No. 7, deduce that, $R_{1.23} \geq r_{12}$

**Answer:**

Since $|r_{13.2}| \leq 1$, we have,

$1-R_{1.23}{}^2 \leq 1-r_{12}{}^2$ implies, $R_{1.23} \geq r_{12}$

10. From result in question No. 7, deduce that $R_{1.23}{}^2 = r_{12}{}^2 + r_{13}{}^2$ if $r_{23}=0$.

**Answer:**

We have $r_{13.2}{}^2=\dfrac{(r_{13}{}^2+r_{12}r_{23})^2}{\sqrt{(1-r_{12}{}^2)(1-r_{23}{}^2)}}=\dfrac{r_{13}{}^2}{\sqrt{(1-r_{12}{}^2)}}$, if $r_{23}=0$.

Therefore $1-R_{1.23}{}^2=(1-r_{12}{}^2)\left[1-\dfrac{r_{13}{}^2}{1-r_{12}{}^2}\right]=1-r_{12}{}^2-r_{13}{}^2$

Hence $R_{1.23}{}^2=r_{12}{}^2+r_{13}{}^2$, if $r_{23}=0$

11. Using expression for multiple correlation coefficient in terms of partial and total correlation coefficients, deduce that $1-R_{1.23}{}^2=\dfrac{(1-\rho)(1+2\rho)}{(1+\rho)}$ provided all coefficients of zero order are equal to ρ.

**Answer**

Here we have given, $r_{12} = r_{13} = r_{23} = \rho$. Therefore,

$r_{13.2}=\dfrac{\rho-\rho^2}{\sqrt{(1-\rho^2)(1-\rho^2)}}=\dfrac{\rho(1-\rho)}{(1-\rho^2)}=\dfrac{\rho}{1+\rho^2}$. Hence we get,

$$1-R_{1.23}{}^2=(1-\rho^2)\left[1-\frac{\rho^2}{(1+\rho^2)}\right]=\frac{(1-\rho)(1+2\rho)}{(1+\rho)}$$

12. If $R_{1.23}=0$, X1 is uncorrelated with any of other variables, ie., $r_{12} = r_{13}=0$.

**Answer:**

Let us consider the expression for multiple correlation coefficient in terms of partial and total correlation coefficient If $R_{1.23}=0$, we get,

$1=(1-r_{12}{}^2)(1-r_{13.2}{}^2)$

Since $0 \leq r_{12}^2 \leq 1$ and $0 \leq r_{13.2}^2 \leq 1$, will hold if and only if $r_{12} = 0$ and $r_{13.2} = 0$.

Now, $r_{13.2} = 0$, $\Rightarrow \frac{r_{13} - r_{12}r_{32}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}} = 0 \Rightarrow r_{13} = 0$

If $R_{1.23} = 0$ then $r_{12} = r_{13} = 0$. ie., X1 is uncorrelated with any of other variables

13.Give generalised form of multiple correlation coefficient.

**Answer:**

In the case of n variables $X_1, X_2, \ldots X_n$, the multiple correlation coefficients of $X_1$ on $X_2, X_3 \ldots X_n$ is usually denoted by $R_{1.23\ldots n}$ , is the correlation coefficient between X1 and $e_{1.23\ldots n} = X_1 - X_{1.23\ldots n}$

$$\therefore R_{1.23\ldots n} = \frac{Cov(X_1, e_{1.23\ldots n})}{\sqrt{V(X_1)V(e_{1.23\ldots n})}}$$

$$Cov(X_1, e_{1.23\ldots n}) = \frac{1}{N}\Sigma X_1 e_{1.23\ldots n} = \frac{1}{N}\Sigma X_1 (X_1 - X_{1.23\ldots n})$$

$$= \frac{1}{N}\Sigma X_1^2 - \frac{1}{N}\Sigma X_1 X_{1.23\ldots n} = \frac{1}{N}\Sigma X_1^2 - \frac{1}{N}\Sigma X_{1.23\ldots n}^2 = \sigma_1^2 - \sigma_{1.23\ldots n}^2$$

Also, $V(e_{1.23\ldots n}) = \frac{1}{N}\Sigma e_{1.23\ldots n}^2 = \frac{1}{N}\Sigma (X_1 - X_{1.23\ldots n})^2 = \frac{1}{N}\Sigma (X_1^2 + X_{1.23\ldots n}^2 - 2X_1 X_{1.23\ldots n})$

$$= \frac{1}{N}\Sigma X_1^2 + \frac{1}{N}\Sigma X_{1.23\ldots n}^2 - 2\frac{1}{N}\Sigma X_1 X_{1.23\ldots n}$$

$$= \frac{1}{N}\Sigma X_1^2 + \frac{1}{N}\Sigma X_{1.23\ldots n}^2 - \frac{2}{N}\Sigma X_{1.23\ldots n}^2 = \sigma_1^2 - \sigma_{1.23\ldots n}^2$$

Therefore

$$R_{1.23\ldots n} = \frac{\sigma_1^2 - \sigma_{1.23\ldots n}^2}{\sqrt{\sigma_1^2(\sigma_1^2 - \sigma_{1.23\ldots n}^2)}} = \left(\frac{\sigma_1^2 - \sigma_{1.23\ldots n}^2}{\sigma_1^2}\right)^{1/2}$$

$$R_{1.23\ldots n}^2 = 1 - \frac{\sigma_{1.23\ldots n}^2}{\sigma_1^2} = 1 - \frac{\omega}{\omega_{11}}$$

14. Write the expressions for the multiple correlation coefficient , $R_{2.13}$ and $R_{3.12}$.

**Answer:**

$R_{2.13}^2 = \frac{r_{21}^2 + r_{23}^2 - 2r_{21}r_{23}r_{13}}{1 - r_{13}^2}$ and $R_{3.12}^2 = \frac{r_{13}^2 + r_{12}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}$

15. 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 were obtained. $r_{12} = 0.77$; $r_{13} = 0.72$ and $r_{23} = 0.52$.

Find the multiple correlation coefficient $R_{1.23}^2$

**Answer:**

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

Therefore $R_{1.23}$ = square root (0.7334) = 0.8564, since multiple correlation coefficient is non-negative.