Lecture Notes on Multiple and Partial Correlation

Multiple Correlation Coefficient : Study of the degree of (extent of) linear relationship of X_1 on X_2 and X_3 is one of the important issue. Degree of the linear relationship is obtained by computing the correlation between X_1 and $e_{1.23}$ (an estimated value of X_1 based on X_2 and X_3 by using linear regression method) which is called **Multiple Correlation Coefficient of** X_1 on X_2 and X_3 , it is denoted by $R_{1.23}$ and is given by

$$R_{1.23} = \frac{Cov(X_1, e_{1.23})}{SD(X_1)SD(e_{1.23})} = \frac{Cov(X_1, e_{1.23})}{|\sqrt{Var(X_1)}||\sqrt{Var(e_{1.23})}|} \quad ...(1)$$

An estimated value of X_1 by using linear regression on X_2 and X_3:

$$e_{1.23} = \bar{X}_1 + b_{12.3}(X_2 - \bar{X}_2) + b_{13.2}(X_3 - \bar{X}_3). \quad ...(2)$$

An error (residual) in predictor: $X_{1.23} = X_1 - e_{1.23}$ . ...(3)

$$Cov(X_1, e_{1.23}) = Cov(X_{1.23} + e_{1.23},\ e_{1.23}) \quad \text{(from (3))}$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(X_{1.23,j} + e_{1.23,j} - \bar{X}_{1.23} - \bar{e}_{1.23}\right)\left(e_{1.23,j} - \bar{e}_{1.23}\right)$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(e_{1.23,j} - \bar{e}_{1.23}\right)^2 + \frac{1}{N}\sum_{j=1}^{N}\left(X_{1.23,j} - \bar{X}_{1.23}\right)\left(e_{1.23,j} - \bar{e}_{1.23}\right)$$

$$\boldsymbol{Cov(X_1, e_{1.23}) = Var(e_{1.23}) + Cov(X_{1.23}, e_{1.23})} \quad \textbf{...(4)}$$

Note that:

$$\bar{X}_{1.23} = \frac{1}{N}\sum_{j=1}^{N} X_{1.23,j} = \frac{1}{N}\sum_{j=1}^{N}(X_{1j} - e_{1.23,j})$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(X_{1j} - \bar{X}_1 - b_{12.3}(X_{2j} - \bar{X}_2) - b_{13.2}(X_{3j} - \bar{X}_3) \right) \quad \text{(from (2))}$$

= 0 (Since sum of deviation of observations about their arithmetic mean is zero)

$$\boldsymbol{\bar{X}_{1.23} = 0} \quad \textbf{...(5)}$$

That is, arithmetic mean of residuals (errors) is zero.

$$\bar{e}_{1.23} = \frac{1}{N}\sum_{j=1}^{N} e_{1.23,j} = \frac{1}{N}\sum_{j=1}^{N}(\bar{X}_1 + b_{12.3}(X_{2j} - \bar{X}_2) + b_{13.2}(X_{3j} - \bar{X}_3)) = \bar{X}_1 \text{ (from (2))}$$

$$\boldsymbol{\bar{e}_{1.23} = \bar{X}_1} \quad \textbf{...(6)}$$

(By using property that sum of deviation of observations about their arithmetic mean is zero)

That is, **arithmetic mean of estimated values of X_1** based on observations on X_2 and X_3 is simply an **arithmetic mean of X_1**.

Now consider,

$Cov(X_{1.23}, e_{1.23})$

$= \frac{1}{N}\sum_{j=1}^{N}\left(X_{1.23,j} - \bar{X}_{1.23}\right)\left(e_{1.23,j} - \bar{e}_{1.23}\right)$

$= \frac{1}{N}\sum_{j=1}^{N} X_{1.23,j}(e_{1.23,j} - \bar{X}_1)$ (from (5) and (6))

$= \frac{1}{N}\sum_{j=1}^{N} X_{1.23,j}e_{1.23,j}$ (from (5))

$= \frac{1}{N}\sum_{j=1}^{N}(X_{1j} - e_{1.23,j})e_{1.23,j}$ (from (3))

$= \frac{1}{N}\sum_{j=1}^{N}\left\{X_{1j} - \bar{X}_1 - b_{12.3}(X_{2j} - \bar{X}_2) - b_{13.2}(X_{3j} - \bar{X}_3)\right\}\left\{\bar{X}_1 + b_{12.3}(X_{2j} - \bar{X}_2) + b_{13.2}(X_{3j} - \bar{X}_3)\right\}$ (from (2))

$$= \bar{X}_1 \frac{1}{N}\sum_{j=1}^{N}\left\{X_{1j} - \bar{X}_1 - b_{12.3}(X_{2j} - \bar{X}_2) - b_{13.2}(X_{3j} - \bar{X}_3)\right\}$$
$$+ b_{12.3}\ \frac{1}{N}\sum_{j=1}^{N}\left(X_{2j} - \bar{X}_2\right)\left\{X_{1j} - \bar{X}_1 - b_{12.3}(X_{2j} - \bar{X}_2) - b_{13.2}(X_{3j} - \bar{X}_3)\right\}$$
$$+ b_{13.2}\ \frac{1}{N}\sum_{j=1}^{N}\left(X_{3j} - \bar{X}_3\right)\left\{X_{1j} - \bar{X}_1 - b_{12.3}(X_{2j} - \bar{X}_2) - b_{13.2}(X_{3j} - \bar{X}_3)\right\}$$

= 0

(By using property that sum of deviation of observations about their arithmetic mean is zero and normal equations (VI))

That is

$$\boldsymbol{Cov(X_{1.23}, e_{1.23}) = 0.} \qquad \textbf{...(7)}$$

Covariance between residual (error) of X_1 and its estimated value (while estimating X_1 based on X_2, X_3) is zero.

From (7) and (4), we have

$$\boldsymbol{Cov(X_1, e_{1.23}) = Var(e_{1.23})}\ (\geq \mathbf{0}) \qquad \textbf{...(8)}$$

That is, covariance between X_1 and its estimated value is always non-negative and hence multiple correlation coefficient between X_1 and its estimated value is always non-negative.

Also note that,

$Cov(X_1, e_{1.23})$

$= \frac{1}{N}\sum_{j=1}^{N}\left(X_{1j} - \bar{X}_1\right)\left(e_{1.23,j} - \bar{e}_{1.23}\right)$

$= \frac{1}{N}\sum_{j=1}^{N}\left(X_{1j} - \bar{X}_1\right)\left(\bar{X}_1 + b_{12.3}(X_{2j} - \bar{X}_2) + b_{13.2}(X_{3j} - \bar{X}_3) - \bar{X}_1\right)$ (from (2) and (6))

$= b_{12.3}\frac{1}{N}\sum_{j=1}^{N}\left(X_{1j} - \bar{X}_1\right)\left(X_{2j} - \bar{X}_2\right) + b_{13.2}\frac{1}{N}\sum_{j=1}^{N}\left(X_{1j} - \bar{X}_1\right)\left(X_{3j} - \bar{X}_3\right)$

$= b_{12.3}Cov(X_1, X_2) + b_{13.2}Cov(X_1, X_3)$

$= -\frac{\sigma_1 R_{12}}{\sigma_2 R_{11}}\sigma_1\sigma_2 r_{12} - \frac{\sigma_1 R_{13}}{\sigma_3 R_{11}}\sigma_1\sigma_3 r_{13}$ (form (VII) and (VIII))

$= -\frac{{\sigma_1}^2}{R_{11}}(r_{12}R_{12} + r_{13}R_{13}) = -\frac{{\sigma_1}^2}{R_{11}}(r_{11}R_{11} + r_{12}R_{12} + r_{13}R_{13} - r_{11}R_{11})$

$= -\frac{{\sigma_1}^2}{R_{11}}(|R| - r_{11}R_{11}) = {\sigma_1}^2(1 - \frac{|R|}{R_{11}})$.

That is,

$$\boldsymbol{Cov(X_1, e_{1.23}) = Var(e_{1.23}) = {\sigma_1}^2(1 - \frac{|R|}{R_{11}})} \quad \textbf{...(9)}$$

Hence from (1) and (9)

$$\boldsymbol{R_{1.23} = \frac{Var(e_{1.23})}{\left|\sqrt{Var(X_1)}\right|\left|\sqrt{Var(e_{1.23})}\right|} = \left|\frac{\sqrt{Var(e_{1.23})}}{\sqrt{Var(X_1)}}\right| = \left(1 - \frac{|R|}{R_{11}}\right)^{1/2}} \quad \textbf{...(10)}$$

Similarly it can be shown that**:** $\boldsymbol{R_{2.13} = \left(1 - \frac{|R|}{R_{22}}\right)^{1/2}}$ and $\boldsymbol{R_{3.12} = \left(1 - \frac{|R|}{R_{33}}\right)^{1/2}}$

Partial Correlation Coefficient:

The degree of (extent of) linear relationship between X_1 and X_2 after eliminating the linear effect of X_3 from X_1 and X_2 is called **Partial Correlation Coefficient between** X_1 on X_2. It is denoted by $r_{12.3}$. Mathematically, we obtain an expression for $\mathbf{r_{12.3}}$ as below:

We have, linear regression equation of X_1 on X_3 and that of X_2 on X_3 by least square method are respectively given by:

$X_1 = \bar{X}_1 + b_{13}(X_3 - \bar{X}_3) = e_{1.3}\ (say)$ and : $X_2 = \bar{X}_2 + b_{23}(X_3 - \bar{X}_3) = e_{2.3}\ (say)$ **..(11)**

Therefore, residual of X_1 after eliminating the effect of X_3 and residual of X_2 after eliminating the effect of X_3 are respectively $X_{1.3} = X_1 - e_{1.3}$ *and* $X_{2.3} = X_2 - e_{2.3}$. That is

$$\boldsymbol{X_{1.3} = X_1 - \bar{X}_1 - b_{13}(X_3 - \bar{X}_3)} \textbf{ and } \boldsymbol{X_{2.3} = X_2 - \bar{X}_2 - b_{23}(X_3 - \bar{X}_3)} \quad \textbf{...(12)}$$

Therefore, $\boldsymbol{r_{12.3}} = \frac{\boldsymbol{Cov(X_{1.3}, X_{2.3})}}{\boldsymbol{SD(X_{1.3})SD(X_{2.3})}}$ **...(13)**

$Cov(X_{1.3}, X_{2.3}) = \frac{1}{N}\sum_{j=1}^{N}(X_{1.3,j} - \bar{X}_{1.3})(X_{2.3,j} - \bar{X}_{2.3}) = \frac{1}{N}\sum_{j=1}^{N}(X_{1.3,j})(X_{2.3,j})$

(by definition of covariance and from (12), $\bar{X}_{2.3} = \bar{X}_{1.3} = 0$)

$= \frac{1}{N}\sum_{j=1}^{N}(X_{1j} - \bar{X}_1 - b_{13}(X_{3j} - \bar{X}_3))(X_{2j} - \bar{X}_2 - b_{23}(X_{3j} - \bar{X}_3))$ from (12)

$= \frac{1}{N}\sum_{j=1}^{N}(X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2) - \frac{b_{23}}{N}\sum_{j=1}^{N}(X_{1j} - \bar{X}_1)(X_{3j} - \bar{X}_3)$

$- \frac{b_{13}}{N}\sum_{j=1}^{N}(X_{3j} - \bar{X}_3)(X_{2j} - \bar{X}_2) + \frac{b_{13}b_{23}}{N}\sum_{j=1}^{N}(X_{3j} - \bar{X}_3)^2$

$= r_{12}\sigma_1\sigma_2 - r_{23}\frac{\sigma_2}{\sigma_3} \times r_{13}\sigma_1\sigma_3 - r_{13}\frac{\sigma_1}{\sigma_3} \times r_{23}\sigma_2\sigma_3 + r_{13}\frac{\sigma_1}{\sigma_3} \times r_{23}\frac{\sigma_2}{\sigma_3} \times \sigma_3^2$

(by property of covariance and regression coefficients)

$= \sigma_1\sigma_2(r_{12} - r_{13}r_{23})$ That is,

$$\boldsymbol{Cov(X_{1.3}, X_{2.3}) = \sigma_1\sigma_2(r_{12} - r_{13}r_{23})} \quad \textbf{...(14)}$$

$Var(X_{1.3}) = \frac{1}{N}\sum_{j=1}^{N}(X_{1.3,j} - \bar{X}_{1.3})^2 = \frac{1}{N}\sum_{j=1}^{N}(X_{1.3,j})^2$

(by definition of variance and from (12), $\bar{X}_{1.3} = 0$)

$= \frac{1}{N}\sum_{j=1}^{N}(X_{1j} - \bar{X}_1 - b_{13}(X_{3j} - \bar{X}_3))^2$

$= \frac{1}{N}\sum_{j=1}^{N}(X_{1j} - \bar{X}_1)^2 + \frac{{b_{13}}^2}{N}\sum_{j=1}^{N}(X_{3j} - \bar{X}_3)^2 - 2b_{13}\frac{1}{N}\sum_{j=1}^{N}(X_{1j} - \bar{X}_1)(X_{3j} - \bar{X}_3)$

$= \sigma_1^2 + r_{13}^2\frac{\sigma_1^2}{\sigma_3^2} \times \sigma_3^2 - 2r_{13}\frac{\sigma_1}{\sigma_3} \times r_{13}\sigma_1\sigma_3 = \sigma_1^2 + r_{13}^2\sigma_1^2 - 2r_{13}^2\sigma_1^2 = \sigma_1^2(1 - r_{13}^2)$. That is,

$$\boldsymbol{S.D.(X_{1.3}) = \sigma_1\sqrt{(1 - r_{13}^2)}} \quad \textbf{...(15)}$$

Similarly,

$$\boldsymbol{S.D.(X_{2.3}) = \sigma_2\sqrt{(1 - r_{23}^2)}} \quad \textbf{...(16)}$$

Put (14) , (15) , (16) in (13) we get,

Therefore, $\boldsymbol{r_{12.3} = \frac{(r_{12} - r_{13}r_{23})}{\sqrt{(1-r_{13}^2)}\sqrt{(1-r_{23}^2)}} = \frac{-R_{12}}{\sqrt{R_{11}}\sqrt{R_{22}}}}$ **...(17)**

Similarly, $\boldsymbol{r_{13.2} = \frac{(r_{13} - r_{12}r_{32})}{\sqrt{(1-r_{12}^2)}\sqrt{(1-r_{32}^2)}} = \frac{-R_{13}}{\sqrt{R_{11}}\sqrt{R_{33}}}}$ and

$$\boldsymbol{r_{23.1} = \frac{(r_{23} - r_{21}r_{31})}{\sqrt{(1-r_{21}^2)}\sqrt{(1-r_{31}^2)}} = \frac{-R_{23}}{\sqrt{R_{22}}\sqrt{R_{33}}}}$$