Topic for the day 29/09/2021

*PARTIAL CORRELATION

- MULTIPLE CORRELATION
- RELATION BETWEEN THEM Par. mul.
- COEFFICIENT OF DETERMINATION R^2

Partial correlation

- If there are three variables X1, X2 and X3 there will be three coefficients of partial correlation, each studying the relationship between two variables when the third is held constant. If we denote by $r_{12.3}$ ie., the coefficient of partial correlation between X1 and X2 keeping X3 constant, it is calculated as

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

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

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

X_1 X_2 X_3

r_{12} r_{13} r_{23}

Problem: In a trivariate distribution, it is found that $r_{12} = 0.7$, $r_{13} = 0.61$ and $r_{23} = 0.4$. Find the partial correlation coefficients.

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} = \frac{0.7-(0.61)(0.4)}{\sqrt{1-(0.61)^2}\sqrt{1-(0.4)^2}} = 0.628$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}} = 0.504$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}} = -0.048$$

$r_{12.3} = 1.924$

$r_{13.2}$

$r_{23.1}$

2. Is it possible to get the following from a set of experimental data?

$r_{12}=0.6$, $r_{13}=-0.5$ and $r_{23}=0.8$

$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^{2}}\sqrt{1-r_{23}^{2}}}=\frac{0.6-(-0.5)(0.8)}{\sqrt{1-(0.5)^{2}}\sqrt{1-(0.8)^{2}}}=1.923$$

Since the value of $r_{12.3}$ is greater than one, there is some inconsisting in the given data.

x_1 x_2 x_3

$r_{12} = -$

$r_{13} = -$ $r_{23} = -$

Illustration 2. On the basis of observation made on agricultural production (X_1) the use of fertilizers (X_2) and the use of irrigation (X_3), the following zero order correlation coefficients were obtained :

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

x_2 x_3

Compute the partial correlation between agricultural production and the use of fertilizers eliminating the effect of irrigation.

Solution. We have to calculate the value of $r_{12.3}$

x_1 $r_{12.3}$ const

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

Substituting the values $r_{12.3} = 0.8,\ r_{13} = 0.65$ and $r_{23} = 0.7$

$$r_{12.3} = \frac{0.8 - (0.65 \times 0.7)}{\sqrt{1 - (0.65)^2}\ \sqrt{1 - (0.7)^2}} = \frac{0.8 - 0.455}{\sqrt{1 - 0.4225}\ \sqrt{1 - 0.49}} = 0.636.$$

Multiple correlation

- The coefficients of multiple correlation with three variables X1 x_2, x_3 .. $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$ re

$R_{1.23}$ is the coefficient of multiple correlation related to X1 as a dependent variable X2 and X3 as two independent variables and it can be expressed in terms of r_{12}, r_{23} and r_{13} as

$R_{1.23}$

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

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

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

Problems

Problem: The following Zero –order correlation coefficients are given: $r_{12} = 0.98$, $r_{13} = 0.44$ and $r_{23} = 0.54$. Calculate multiple correlation coefficient treating first variable as dependent and second & third variables as independent.

$R_{1.23}$

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{23}^2}} = 0.986$$

Calculate (a) $R_{1.23}$, (b) $R_{3.12}$ and (c) $R_{2.13}$ for the following data :

$r_{12} = 0.6 \quad r_{13} = 0.7 \quad r_{23} = 0.65$

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

$$= \sqrt{\frac{(0.6)^2 + (0.7)^2 - (2 \times 0.6 \times 0.7 \times 0.65)}{1-(0.65)^2}}$$

$$= \sqrt{\frac{0.36 + 0.49 - 0.546}{0.5775}} = \sqrt{0.526} = 0.725$$

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

$$= \sqrt{\frac{(0.7)^2 + (0.65)^2 - 2(0.6 \times 0.7 \times 0.65)}{1-(0.6)^2}}$$

$$= \sqrt{\frac{0.49 + 0.4225 - 0.546}{1-0.36}} = \sqrt{0.573} = 0.757$$

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

$$= \sqrt{\frac{(0.6)^2 + (0.65)^2 - (2 \times 0.6 \times 0.7 \times 0.65)}{1-(0.7)^2}}$$

$$= \sqrt{\frac{0.36 + 0.4225 - 0.546}{0.51}} = \sqrt{0.464} = 0.681.$$

$r_{12} = 0.6$

$r_{13} = 0.7$

$r_{23} = 0.65$

Relationship between Partial and Multiple Correlation Coefficients

Interesting results connecting the multiple correlation coefficients and the various partial correlation coefficients can be found. For example, we find :

$$1 - R^2_{1.23} = (1 - r^2_{12})(1 - r^2_{13.2})$$

$$1 - R^2_{1.234} = (1 - r^2_{13})(1 - r^2_{13.2})(1 - r^2_{14.23})$$

4 var.

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

$(r_{23})^2$

- **Coefficient of Determination, r^2 or R^2:**
- The *coefficient of determination*, r^2, is useful because it gives the **proportion of the variance** (fluctuation) of one variable that is predictable from the other variable.
- The *coefficient of determination* is the ratio of the explained variation to the total variation
- The *coefficient of determination* is such that $0 \le r^2 \le 1$, and denotes the strength of the linear association between x and y.
- It represents the percent of the data that is the closest to the line of best fit.

$(0.92)^2$

- For example, if $r = 0.922$, then $r^2 = 0.850$, which means that 85% of the total variation in y can be explained by the linear relationship between x and y (as described by the regression equation). The other 15% of the total variation in y remains unexplained.

Note: The *coefficient of determination* is a measure of how well the regression line represents the data. If the regression line passes exactly

100% through every point on the scatter plot, it would be able to explain all of the variation. The further the line is away from the points, the less it is able to explain.