PARTIAL CORRELATION

Course Instructor: Santosh Chhatkuli

Definition

Partial correlation measures the relationship between two variables while controlling for the influence of one or more other variables. It helps isolate the direct association between two variables by removing the effect of the confounding variable(s). It helps study of relationship between two variables after removing the overlap with remaining variables completely from two variables.

Types of partial correlation:

1. First order partial correlation: Controls the effect of one variable
2. Second order partial correlation: Controls the effect of two variables
3. Higher order partial correlation: Controls the effect of three or more variables

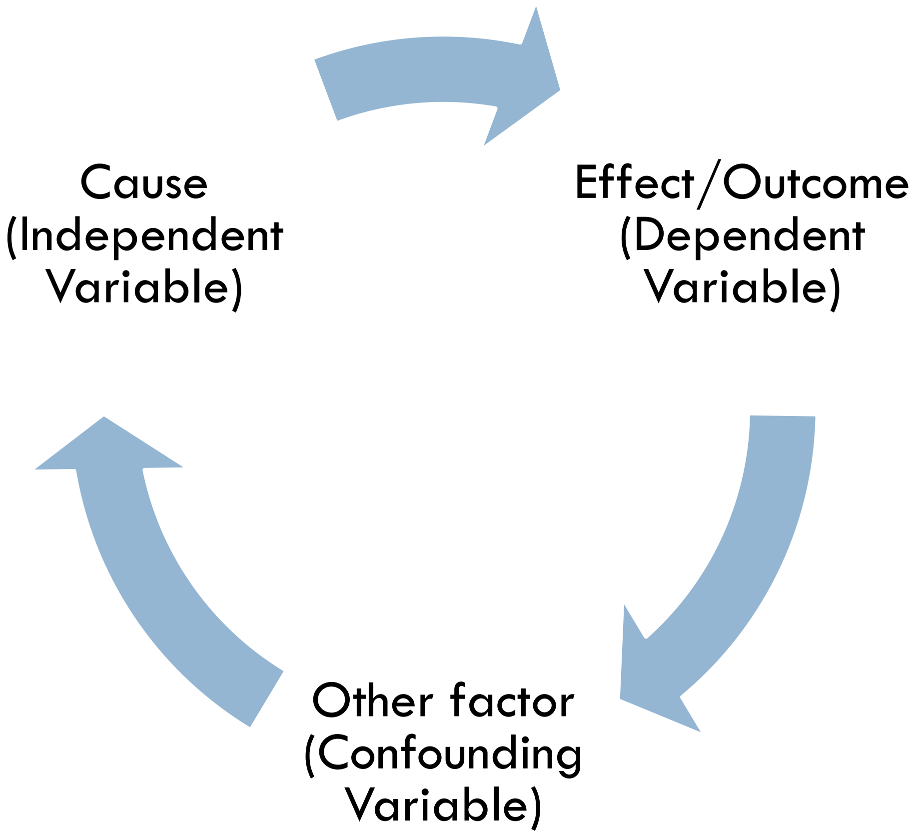
A first order partial correlation is the statistical technique that quantifies the degree of association between a dependent variable and an independent variable by removing the association of a third variable with both of the correlating variables.

Purpose of studying partial correlation coefficient

To find an association between two variables after controlling for the effect of one or more potentially confounding variables.

1. In the study of systolic blood pressure (SBP) and stress level, the age could be potential confounder.
2. In the study of exercise and weight loss, the diet could be potential confounding variable.
3. In the study of education and salary, the work experience could be potential confounding variable.
4. In the study of screen time and sleep quality, the stress level could be potential confounder
5. In the study of study hours and exam scores, the number of hours of sleep is a potential confounder
6. In the study of temperature and energy consumption, the household size could be potential confounder.
7. In the study of advertisement and product sales, the price of the price of the product is potential confounder

Confounding Effect



Computation of partial correlation coefficient

We first calculate the zero-order correlation coefficients between all possible pairs of variables. A zero-order correlation or simple correlation is the relationship between two variables but without controlling the effect of other or additional variables on both correlating variables.

Consider the study of dependent variable Y and two independent variables X_1 and X_2.

The zero order correlation coefficients are:

- r_{Y1} (Simple correlation coefficients between Y and X_1)
- r_{Y2} (Simple correlation coefficients between Y and X_2)
- r_{12} (Simple correlation coefficients between X_1 and X_2)

Notation and Formula

$r_{Y1.2}$ = Partial correlation coefficient between Y and X_1 after controlling for the effect of X_2

$$= \frac{r_{Y1} - r_{Y2}\, r_{12}}{\sqrt{1 - r_{Y2}^2}\sqrt{1 - r_{12}^2}}$$

$r_{Y2.1}$= Partial correlation coefficient between Y and X_2 after controlling for the effect of X_1

$$= \frac{r_{Y2} - r_{Y1}\, r_{12}}{\sqrt{1 - r_{Y1}^2}\sqrt{1 - r_{12}^2}}$$

$r_{12.Y}$= Partial correlation coefficient between X1 and X_2 after controlling for the effect of Y

$$= \frac{r_{12} - r_{Y1}\, r_{Y2}}{\sqrt{1 - r_{Y1}^2}\sqrt{1 - r_{Y2}^2}}$$

Range for the partial correlation coefficient

All partial correlation coefficients range from -1 to +1

So,

-1 ≤ $r_{Y1.2}$ ≤ +1

-1 ≤ $r_{Y2.1}$ ≤ +1

-1 ≤ $r_{12.Y}$ ≤ +1

The general formula for partial correlation coefficient:

$$r_{AB.C} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{1 - r_{AC}^2}\sqrt{1 - r_{BC}^2}}$$

$r_{AB.C}$ follows t distribution with n – 2 – k degrees of freedom, where n being sample size and k being order of partial correlation.

Numerical

The salaries of workers are expected to be dependent among other factors, on the number of years they have spend in college and their work experience. The following table gives information on the annual salaries (in thousands of dollars) for 12 persons, the no. of years of each of them spend in school, and the total no. of years of experiences.

S.N.	1	2	3	4	5	6	7	8	9	10	11	12
Salary	52	44	48	77	68	48	59	83	28	61	27	69
Schooling	16	12	13	20	18	16	14	18	12	16	12	16
Experiences	6	10	15	8	11	2	12	4	6	9	2	18

Compute partial correlation coefficients and interpret the result

Computational Result

Here

Y = Salary (thousand of dollar)

X_1 = Years of schooling

X_2 = Years of experiences

$\sum Y = 664$	$\sum X_1 = 183$	$\sum X_2 = 103$
$\sum Y^2 = 40166$	$\sum X_1^2 = 2869$	$\sum X_2^2 = 1155$
$\sum YX_1 = 10576$	$\sum YX_2 = 5985$	$\sum X_1X_2 = 1569$

Computing simple correlation coefficients:

$$r_{Y1} = \frac{n\,.\sum YX_1 - \sum Y\,.\sum X_1}{\sqrt{n\sum Y^2 - (\sum Y)^2}\sqrt{n\sum X_1^{\,2} - (\sum X_1)^2}}$$

$$= \frac{12\ X\ 10576 - 664\ X\ 183}{\sqrt{12\ X\ 40166 - (\,664)^2}\sqrt{12\ X\ 2869 - (\,183)^2}} = 0.869 \approx 0.87$$

$$r_{Y2} = \frac{n\,.\sum YX_2 - \sum Y\,.\sum X_2}{\sqrt{n\sum Y^2 - (\sum Y)^2}\sqrt{n\sum X_2^{\,2} - (\sum X_2)^2}}$$

$$= \frac{12\ X\ 5985 - 664\ X\ 103}{\sqrt{12\ X\ 40166 - (\,664)^2}\sqrt{12\ X\ 1155 - (\,103)^2}} = 0.297 \approx 0.30$$

$$r_{12} = \frac{n.\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{n\sum X_1^{\,2} - (\sum X_1)^2}\sqrt{n\sum X_2^{\,2} - (\sum X_2)^2}}$$

$$= \frac{12\ X\ 1569 - 183\ X\ 103}{\sqrt{12\ X\ 2689 - (\,183)^2}\sqrt{12\ X\ 1155 - (\,103)^2}} = -\ 0.01202 \approx -\ 0.012$$

Computing partial correlation coefficients

$$r_{Y1.2} = \frac{r_{Y1} - r_{Y2}\, r_{12}}{\sqrt{1-r_{Y2}^2}\sqrt{1-r_{12}^2}} = \frac{0.869 - 0.297\ x - 0.01202}{\sqrt{1-0.297^2}\sqrt{1-(-0.01202)^2}} = 0.914$$

$$r_{Y2.1} = \frac{r_{Y2} - r_{Y1}\, r_{12}}{\sqrt{1-r_{Y1}^2}\sqrt{1-r_{12}^2}} = \frac{0.297 - 0.869\ x - 0.01202}{\sqrt{1-0.869^2}\sqrt{1-(-0.01202)^2}} = 0.695$$

$$r_{12.Y} = \frac{r_{12} - r_{Y1}\, r_{Y2}}{\sqrt{1-r_{Y1}^2}\sqrt{1-r_{Y2}^2}} = \frac{-0.01202 - 0.869\ x\ 0.297}{\sqrt{1-0.869^2}\sqrt{1-0.297^2}} = -\ 0.571$$

Comparison between simple and partial correlation

Simple correlation coefficient	Partial correlation coefficient
$r_{Y1} = 0.869$	$r_{Y1.2} = 0.914$
$r_{Y2} = 0.297$	$r_{Y2.1} = 0.695$
$r_{12} = -0.01202$	$r_{12.Y} = -0.571$

Conclusion

- There is a very strong positive correlation between annual salary and number of years of schooling after controlling for years of education. There is a slight increment in the correlation index after controlling for the confounder i.e. number of years of experience.
- There is a moderate to strong degree of correlation between annual salary and number of years of experience after controlling for the effect of schooling years i.e. people having same schooling years, experience period affecting moderately to the annual salary.
- Correlation between number of years of schooling and number of years of experience drastically change from negligible (- 0.01202) to moderate degree of negative correlation (- 0.571) after controlling for salary. It means that in order to get same salary employee with less schooling years has to work for a greater number of years than those who have higher degree.

Using notation: X_1, X_2 and X_3

First order partial correlation coefficients are:

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

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

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

Numericals

1. In tri-variate study where Y = Husband's hours of housework per week, X_1 = No. of children of dual wage earner families, and X_2 = Husband's years of education the following zero order correlation coefficient were obtained.

	Husband's Housework	No. of children	Husband's years of education
Husband's Housework	1	0.50	- 0.30
No. of children			- 0.47
Husband's years of education			1

Compute first order partial correlation coefficients. Interpret them

More problems

Suppose we have following information about three variables X1, X2 and X3.

$r_{12} = 0.91$, $r_{13} = 0.33$ and $r_{23} = 0.81$

Check whether this information is consistent or not.