Partial and Part correlation

1. Partial correlation

Sometimes it is desirable to know the relationship between two variables with the effects of a third variable held constant. As an example suppose that it has been demonstrated that both intelligence and number of hours worked are correlated with exam marks, and further that intelligence and number of hours worked are also correlated. All of these correlations are positive. The more intelligent tend to obtain higher exam scores and tend to work harder, those who work harder tend to be more intelligent and obtain higher exam marks. In a situation like this a straight forward correlation between intelligence and exam marks will also reflect the effect of hours worked on both intelligence and exam marks. Clearly it would be useful for us to be able to find the 'pure' correlation between intelligence and exam marks with hours worked held constant. 'Holding constant' in this situation is known as partialling out, and the technique for partialling out the effects of one or more variables from two others, in order to find the relationship between them is called partial correlation.

For this chapter and the next one a change in subscripts is desirable. Instead of using letters as subscripts with correlation coefficients it will be more useful to refer to the variables being correlated as 1, 2, 3, etc. r_{12} will be the correlation between variables 1 and 2, r_{14} between the first and fourth variable and r_{1n} between the first and nth variable.

Suppose that we have three variables 1, 2, and 3 and we wish to find the relationship between 1 and 2, with the effects of 3 partialled out from both. In fact what we want to do is correlate the residual scores of 1 and 2, after the parts of 1 and 2 predictable from 3 have been subtracted. It has been shown previously that the predicted Z_1 will be $r_{13}Z_3$. The symbol for Z_1 with the effects of 3 partialled out will be $Z_{1.3}$, and generally Z on variable X with Y partialled out will be $Z_{X.Y.}$ The residual Z score on 2 will be $Z_2 - r_{23}Z_3 = Z_{2.3}$. (Note that in all of these cases the subscript of the variable partialled out comes after the dot.)

The partial correlation coefficient $r_{12.3}$ will be:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \qquad (7:1)$$

Proof

It is desired to correlate the residual scores $(Z_1 - r_{13}Z_3)$ and $(Z_2 - r_{23}Z_3)$ and the formula for the correlation coefficient is

$$\frac{\sum xy}{N\sigma_x\sigma_y}$$

So it will be necessary to work out

(a) the covariance of the residual scores and

(b) the standard deviation of the residual scores.

The covariance will be worked out first.

(1) $$\sum(Z_1 - r_{13}Z_3)(Z_2 - r_{23}Z_3)/N = \sum\left(Z_1Z_2 + r_{13}r_{23}Z_3{}^2 - r_{23}Z_1Z_3 - r_{13}Z_2Z_3\right)/N$$

(2) $$= \frac{\sum Z_1Z_2}{N} + r_{13}r_{23}\frac{\sum Z^2{}_3}{N} - r_{23}\frac{\sum Z_1Z_3}{N} - r_{13}\frac{\sum Z_2Z_3}{N}$$

(3) Recalling that $\frac{\sum Z_1 Z_2}{N} = r_{12} etc.$, and that $\frac{\sum Z_3{}^2}{N} = \sigma_Z^2 =$ 1.0 etc. (2) becomes: $r_{12} + r_{13}r_{23} - r_{23}r_{13} - r_{13}r_{23}$

(4) this equals $r_{12} - r_{13}r_{23}$

Turning now to the standard deviation of the residual scores we have

(5) $\sigma_{Z1.3} = \sqrt{\frac{\sum(Z_1 - r_{13}Z_3)^2}{N}}$

(6) $= \sqrt{\frac{\sum Z_1{}^2}{N} + r_{13}^2 \frac{\sum Z_3^2}{N} - 2r_{13} \frac{\sum Z_1 Z_3}{N}}$

(7) $\frac{\sum Z_1^2}{N} = 1.0; \frac{\sum Z_2^3}{N} = 1.0; \textit{ and } \frac{\sum Z_1 Z_3}{N} = r_{13}$

So (6) becomes $\sqrt{1 + r_{13}^2 - 2r_{13}^2} = \sqrt{1 - r_{13}^2}$

(8) Repeating these steps for $Z_{2.3} = \sqrt{\frac{\sum(Z_2 - r_{23}Z_3)^2}{N}}$ gives

$\sqrt{1 - r_{23}^2}$

(9) Putting (4) as the numerator and the products of (7) and (8) as the denominator gives:-

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

The other formulae for the 3 variable case are:

(a) $$r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}} \qquad (7:2)$$

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

In effect the partial correlation coefficient $r_{12.3}$ tells us what the relationship between variables 1 and 2 would be if everyone obtained the same score on variable 3.

Problems

A. Calling exam marks (1), intelligence (2) and hours worked (3), and given $r_{12} = .50$, and r_{13}=.40, and r_{23} of .40 work out the value of $r_{12.3}$.

B. Given three variables (1) prognosis in terms of weeks to recover, (2) an anxiety questionnaire, (3) a physiological measure, and $r_{12} = .40; r_{13} = .30,\ and\ r_{23} = .80$, what is the correlation of the physiological measure with prognosis with the anxiety questionnaire results partialled out from both variables?

Answers

A. $r_{12.3} = \dfrac{.50 - (.40 \times .40)}{\sqrt{1 - .40^2}\sqrt{1 - .40^2}} = \dfrac{.34}{.86} = .396$

B. $r_{13.2} = \mathit{from}\ (7:2)\dfrac{r_{13} - r_{12} r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}}$

$$= \frac{.30 - (.40 \times .80)}{\sqrt{1 - .40^2}\sqrt{1 - .80^2}} = \frac{-.02}{.93 \times .60} = -.036$$

A partial correlation coefficient with one variable partialled out is called a first order partial *r*, with two variables partialled out a second order partial *r* and so on. The general formula for a second order *r* is:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} r_{24.3}}{\sqrt{1 - r_{14.3}^2}\sqrt{1 - r_{24.3}^2}} \qquad (7:3)$$

For an *N*th order partial the formula is:

$$r_{12.34...N} = \frac{r_{12.34...(N-1)} - r_{1N.34...(N-1)} r_{2N.34...(N-1)}}{\sqrt{1 - r_{1N.34...(N-1)}^2}\sqrt{1 - r_{2N.34...(N-1)}^2}} \qquad (7:4)$$

Partial correlation assumes linearity of regression between all variables. If there are serious departures from linearity a partial *r* will be meaningless.

2. *Part or Semi-partial Correlation*

In the case of partial correlation the variable partialled out is partialled out from both of the variables of interest. However, it is also possible to correlate partialled scores on one variable with ordinary scores on another. This type of correlation is called part or semi-partial correlation. The formula for the part correlation coefficient is:

Part correlation coefficient: $r_{1(2.3)} = \dfrac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{23}^2}}$ (7:5)

Note that the partialled variable and the variable partialled from it are put in brackets so $r_{1(2.3)}$ is the correlation between 1 and 2 with the effects of 3 partialled out from 2.

Proof

(1) $$r_{1(2,3)} = \frac{\sum Z_1\left(Z_2 - r_{23}Z_3\right)}{N\sqrt{1-r_{23}^2}}$$

(The standard deviation of the Z_1 scores will be 1 and it has been shown in 7:1 step 8 that the standard deviation of

$$\left[Z_2 - r_{23}Z_3\right] = \sqrt{1-r_{23}^2}$$)

(2) $$r_{1(2,3)} = \frac{\sum Z_1 Z_2 - r_{23}\sum Z_1 Z_3}{N\sqrt{1-r_{23}^2}}$$

(3) Dividing numerator and denominator by *N* gives:

$$r_{1(2,3)} = \frac{r_{12} - r_{23}r_{13}}{\sqrt{1 - r_{23}^2}}$$

$$\text{(because } \frac{\sum Z_1 Z_2}{N} = r_{12} \ \textit{and} \ \frac{\sum Z_1 Z_3}{N} = r_{13}\text{)}$$

Other formulae for the three variable case include:

(a) $$r_{2(1.3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}} \quad (7:6)$$

(b) $$r_{3(1.2)} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1 - r_{13}^2}}$$

Part or semi-partial correlation has the effect of reducing the correlation between the partialled variable and the variable partialled from it to zero.

$$r_{3(1.3)} = r_{1(2.1)} = r_{2(1.2)} \ etc. = 0 \quad (7:7)$$

Proof

(1) $$r_{1(2.1)} = r_{12} - r_{12}r_{11}$$

(2) $$r_{11} = \frac{\sum Z_1 Z_1}{N} = \frac{\sum Z_1^{\ 2}}{N} = 1.0$$

(3) So (1) becomes

$$r_{1(2.1)} = \frac{r_{12} - r_{12}}{\sqrt{1 - r_{12}^2}} = 0$$

Both part and partial correlation are useful in prediction problems, and have a fairly straightforward relationship to multiple correlations, as will be seen in the next chapter.

Problems

A. Given the following data on the relationship between prognosis (1), anxiety questionnaire (2), and physiological measure (3), r_{12} = .40 r_{13} = .30, and r_{23} = .80. What is the correlation between the physiological measure and prognosis with anxiety questionnaire scores partialled out from the physiological measure i.e. what is the value of $r_{1(3.2)}$?

B. A performance test (1) and a verbal intelligence test (2) are used for predicting scholastic success (3). You want to know the correlation between the performance test, with verbal intelligence partialled out from it, and exam marks. If r_{12} =.60, r_{13} = .60, and r_{23} = .40, what will the correlation be?

Answers

A.
$$r_{1(3.2)} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{23}^2}} = \frac{.30-(.40\times.80)}{\sqrt{1-.80^2}} = \frac{-.02}{.60}$$
$$= -.033$$

B.
$$r_{3(1.2)} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{13}^2}} = \frac{.60-(.60\times.40)}{\sqrt{1-.60^2}} = \frac{.36}{.80}$$
$$= +.45$$

3. *The Partial Standard Deviation*

As has been stated previously, scores from which another variable has been partialled out are called partial scores or residual scores. Partial scores have been used in our discussion of the partial and part correlation. The formula for a partial score with one variable partialled out is:

$$\text{Partial score} = Z_{1.2} = Z_1 - r_{12}Z_2 \qquad (7:8)$$

The standard deviation of these partial scores has also been used and derived in 7.1, (5 to 8).

$$\text{Partial standard deviation} = \sigma_{Z1.2} = \sqrt{1-r_{12}^2} \qquad (7:9)$$

In terms of raw scores (7:8) becomes:

Partial score: $$Y.X = Y - r_{xy}\frac{\sigma_y}{\sigma_x}\left(X - M_x\right) + M_y \qquad (7:10)$$

or in terms of deviation scores:

Partial score: $$y.x = y - r_{xy}\frac{\sigma_y}{\sigma_x}x \qquad (7:11)$$

The raw score standard deviation is.

$$\sigma_{y.x} = \sigma_y\sqrt{1 - r_{xy}^2} \qquad (7:12)$$

Recalling that r_{xy}^2 is the coefficient of determination, and that the coefficient of determination is the proportion of variance accounted for, it can be seen from (7:9) that the partial standard deviation is the square root of the variance remaining after variance attributable to another variable has been subtracted. To convert this to a raw score form as in (7:12) $\sqrt{1 - r_{xy}^2}$ is multiplied by the standard deviation of the partialled variable.

Higher order partial standard deviations are also equal to the square root of the variance remaining after the effects of the other variables have been partialled out. Let us consider the case of $\sigma_{1.23}$, which is the partial standard deviation of variable 1 with variables 2 and 3 partialled out. The proportion of variance in variable 1 accounted for by variable 2 will be r_{12}^2. Thus after variable 2 has been partialled out the proportion of variance remaining will equal $1 - r_{12}^2$. Of this remainder some will be accounted for by variable 3. If variable 2 and variable 3 were not related the variance attributable to variable 3 would be r_{13}^2. The variance remaining after the partialling out of variables 2 and 3 would, therefore, be $1 - r_{12}^2 - r_{13}^2$. The square root of this would be the partial standard deviation.

The situation becomes more complicated when variables 2 and 3 correlated with one another. Starting as before, the proportion of variance accounted for by variable 2 will be r_{12}^2, and as before the remaining variance will equal $1 - r_{12}^2$.

However, because 2 and 3 are correlated it will not be possible merely to subtract r_{13}^2 as the proportion of variance attributable to variable 3. As variables 2 and 3 are correlated some of the variance in 1 accounted for by 2, will be shared with variable 3. To obtain the proportion of variance in 1 accounted for by 3, from which variance also accounted for by 2 is excluded, it is necessary to use the partial correlation coefficient $r_{13.2}$. The square of this will give the proportion of variance in 1 which is attributable to 3 after the effects of 2 have been excluded from 1 and 3.

Therefore:

(a) the variance remaining after variable 2 has been partialled out will be:-

$$1 - r_{12}^2$$

(b) of this remainder, variable 3 will account for a proportion of $r_{13.2}^2$, leaving a remainder of

$$1 - r_{13.2}^2$$

Multiplying (a) and (b) together will give the variance remaining after both 2 and 3 have been partialled out. So the formula for the variance not accounted for by the partialling out of the two variables will be:

Partial variance with two variables partialled out

$$(Z \text{ score form}) = \sigma^2_{Z1.23} = \left(1-r^2_{12}\right)\left(1-r^2_{13.2}\right) \tag{7:13}$$

$$\sigma_{Z1.23} = \sqrt{\left(1-r^2_{12}\right)\left(1-r^2_{13.2}\right)} \tag{7:14}$$

Raw score form

$$\sigma_{1.23} = \sigma_1\sqrt{\left(1-r^2_{12}\right)\left(1-r^2_{13.2}\right)} \tag{7:15}$$

Suppose that $r_{12} = .71,$ and $r_{13.2} = .50.$ The proportion of Variance accounted for by $r_{12} = r^2_{12} = .50.$ Fifty per cent of the variance is accounted for which leaves 50 per cent not accounted for. Of this remaining 50 per cent, 25 per cent can be accounted for by variable 3 with 2 partialled out from it. Thus 75 per cent of the 50 per cent remaining after the effects of variable 2 have been allowed for will still remain unaccounted for after variable 3 has been partialled out.

The partial standard deviation will therefore be $\sqrt{.50 \times .75} = \sqrt{.375.}$ This is the value which would also be obtained by use of formula (7:13).

By similar reasoning it can be shown that the partial standard deviation of variable 1 with variables 234...*N* partialled out is:

$$\sigma_{Z1.234...N} = \sqrt{\left(1-r^2_{12}\right)\left(1-r^2_{13.2}\right)\left(1-r^2_{14.23}\right)\cdots\left[1-r^2_{1N.234...(N-1)}\right]} \tag{7:16}$$

The raw score equivalent can be obtained by multiplying by σ_1.

Problem

Given $r_{12} = .40, r_{13} = .50,$ and $r_{23} = .60,$ what is the value of $\sigma_{Z1.23}$?

Answer

(a) Two correlation coefficients are needed r_{12} and $r_{13.2}$. The formula for the latter will be:

$$\frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}} = \frac{.50-(.40\times.60)}{\sqrt{1-.16}\sqrt{1-.36}} = \frac{.26}{.92\times.80} = .35$$

(b) The formula for the partial standard deviation thus becomes:

$$\sqrt{(1-.40^2)(1-.35^2)} = \sqrt{.74} = .86$$