Part (Semi Partial) and Partial Regression Coefficients

Hervé Abdi[1]

1 overview

The *semi-partial* regression coefficient—also called *part* correlation—is used to express the specific portion of variance explained by a given independent variable in a multiple linear regression analysis (MLR). It can be obtained as the correlation between the dependent variable and the residual of the prediction of one independent variable by the other ones. The semi partial coefficient of correlation is used mainly in non-orthogonal multiple linear regression to assess the *specific* effect of each independent variable on the dependent variable.

The *partial* coefficient of correlation is designed to eliminate the effect of one variable on two other variables when assessing the correlation between these two variables. It can be computed as the correlation between the residuals of the prediction of these two variables by the first variable.

[1]In: Neil Salkind (Ed.) (2007). *Encyclopedia of Measurement and Statistics.* Thousand Oaks (CA): Sage.
Address correspondence to: Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/~herve

2 Multiple Regression framework

In MLR, the goal is to predict, knowing the measurements collected on N subjects, a dependent variable Y from a set of K independent variables denoted

$$\{X_1, \ldots, X_k, \ldots, X_K\} . \tag{1}$$

We denote by $\mathbf{X}$ the $N \times (K+1)$ augmented matrix collecting the data for the independent variables (this matrix is called augmented because the first column is composed only of ones), and by $\mathbf{y}$ the $N \times 1$ vector of observations for the dependent variable. This two matrices have the following structure.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,k} & \cdots & x_{1,K} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k} & \cdots & x_{n,K} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,k} & \cdots & x_{N,K} \end{bmatrix} \quad \text{and } \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \\ \vdots \\ y_N \end{bmatrix} \tag{2}$$

The predicted values of the dependent variable $\widehat{Y}$ are collected in a vector denoted $\hat{\mathbf{y}}$ and are obtained using MLR as:

$$\mathbf{y} = \mathbf{Xb} \quad \text{with} \quad \mathbf{b} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} . \tag{3}$$

The quality of the prediction is evaluated by computing the multiple coefficient of correlation denoted $R^2_{Y.1,\ldots,K}$. This coefficient is equal to the coefficient of correlation between the dependent variable (Y) and the predicted dependent variable ($\widehat{Y}$).

3 Partial regression coefficient as increment in explained variance

When the independent variables are pairwise orthogonal, the importance of each of them in the regression is assessed by computing the squared coefficient of correlation between each of the independent variables and the dependent variable. The sum of these

Table 1: A set of data: Y is to be predicted from X_1 and X_2 (data from Abdi et al., 2002). Y is the number of digits a child can remember for a short time (the "memory span"), X_1 is the age of the child, and X_2 is the speech rate of the child (how many words the child can pronounce in a given time). Six children were tested.

Y (Memory span)	14	23	30	50	39	67
X_1 (age)	4	4	7	7	10	10
X_2 (Speech rate)	1	2	2	4	3	6

squared coefficients of correlation is equal to the square multiple coefficient of correlation. When the independent variables are correlated, this strategy *overestimates* the contribution of each variable because the variance that they share is counted several times; and therefore the sum of the squared coefficients of correlation is not equal to the multiple coefficient of correlation anymore. In order to assess the importance of a particular independent variable, the partial regression coefficient evaluates the *specific* proportion of variance explained by this independent variable. This is obtained by computing the increment in the multiple coefficient of correlation obtained when the independent variable is added to the other variables.

For example, consider the data given in Table 1 where the dependent variable is to be predicted from the independent variables X and T. The prediction equation (using Equation 3) is

$$\widehat{Y} = 1.67 + X + 9.50T \; ; \tag{4}$$

it gives a multiple coefficient of correlation of $R^2_{Y.XT} = .9866$. The coefficient of correlation between X and T is equal to $r_{X.T} = .7500$, between X and Y is equal to $r_{Y.X} = .8028$, and between T and Y is equal to $r_{Y.T} = .9890$. The squared partial regression coefficient between X and Y is computed as

$$r^2_{Y.X|T} = R^2_{Y.XT} - r^2_{Y.T} = .9866 - .9890^2 = .0085 \; ; \tag{5}$$

This indicates that when X is entered *last* in the regression equation, it increases the multiple coefficient of correlation by .0085. In other words, X contributes a correlation of .0085 over and above the other dependent variable. As this example show the difference between the correlation and the part correlation can be very large. For T, we find that:

$$r^2_{Y.T|X} = R^2_{Y.XT} - r^2_{Y.X} = .9866 - .8028^2 = .3421 . \tag{6}$$

4 Partial regression coefficient as prediction from a residual

The partial regression coefficient can also be obtained by first computing for each independent variable the residual of its prediction from the other independent variables and then using this residual to predict the dependent variable. In order to do so, the first step is to isolate the specific part of each independent variable. This is done by first predicting a given independent variable from the other independent variables. The residual of the prediction is by definition uncorrelated with the predictors, hence it represents the *specific* part of the independent variable under consideration.

We illustrate the procedure by showing how to compute the semi partial coefficient between X and Y after the effect T has been partialed out. We denote by $\widehat{X}_T$ the prediction of X from T.

The equation for predicting X from T is given by

$$\widehat{X}_T = a_{X.T} + b_{X.T} T , \tag{7}$$

where $a_{X.T}$ and $b_{X.T}$ denote the intercept and slope of the regression line of the prediction of X from T.

Table 2 gives the values of the sums of squares and sum of cross-products needed to compute the prediction of X from T.

- We find the following values for predicting X from T:

$$b_{X.T} = \frac{SCP_{XT}}{SS_T} = \frac{18}{16} = 1.125 ; \tag{8}$$

$$a_{X.T} = M_X - b_{X.T} \times M_T = 7 - 1.125 \times 3 = 3.625 . \tag{9}$$

Table 2: The different quantities needed to compute the values of the parameters $a_{X.T}$, $b_{X.T}$. The following abbreviations are used: $x = (X - M_X)$, $t = (T - M_T)$,

	X	x	x^2	T	t	t^2	$x \times t$
	4	−3	9	1	−2	4	6
	4	−3	9	2	−1	1	3
	7	0	0	2	−1	1	0
	7	0	0	4	1	1	0
	10	3	9	3	0	0	0
	10	3	9	6	3	9	9
$\sum$	42	0	36	18	0	16	18
			SS_X			SS_T	SCP_{XT}

So, the first step is to predict one independent variable from the other one. Then, by subtracting the predicted value of the independent variable from its actual value, we obtain the residual of the prediction of this independent variable. The residual of the prediction of X by T is denoted $e_{X.T}$, it is computed as

$$e_{X.T} = X - \widehat{X}_T \ . \tag{10}$$

Table 3 gives the quantities needed to compute $r^2_{Y.X|T}$. It is obtained as

$$r^2_{Y.X|T} = r^2_{Y.e_{X.T}} = \frac{(SCP_{Ye_{X.T}})^2}{SS_Y SS_{e_{X.T}}} \ . \tag{11}$$

In our example, we find

$$r^2_{Y.X|T} = \frac{15.75^2}{1{,}846.83 \times 15.75} = .0085 \ .$$

Table 3: The different quantities to compute the semi-partial coefficient of correlation between Y and X after the effects of T have been partialed out of X. The following abbreviations are used: $y = Y - M_Y$, $e_{X.T} = X - \widehat{X}_T$.

	Y	y	y^2	X	$\widehat{X}_T$	$e_{X.T}$	$e^2_{X.T}$	$y \times e_{X.T}$
	14	−23.1667	536.69	4	4.7500	−0.7500	0.5625	17.3750
	23	−14.1667	200.69	4	5.8750	−1.8750	3.5156	26.5625
	30	−7.1667	51.36	7	5.8750	1.1250	1.2656	−8.0625
	50	12.8333	164.69	7	8.1250	−1.1250	1.2656	−14.4375
	39	1.8333	3.36	10	7.0000	3.0000	9.0000	5.5000
	67	29.8333	890.03	10	10.3750	−0.3750	0.1406	−11.1875
$\sum$	223	0	1,846.83	42	42.0000	0	15.7500	15.7500
			SS_Y				$SS_{e_{X.T}}$	$SCP_{Ye_{X.T}}$

5 F and t tests for the partial regression coefficient

The partial regression coefficient can be tested by using a standard F-test with the following degrees of freedom $\nu_1 = 1$ and $\nu_2 = N - K - 1$ (with N being the number of observations and K being the number of predictors). Because ν_1 is equal to 1, the square root of F gives a Student-t test. The computation of F is best described with an example: The F for the variable X in our example is obtained as:

$$F_{Y.X|T} = \frac{r^2_{Y.X|T}}{1 - R^2_{Y.XT}} \times (N - 3) = \frac{.0085}{1 - .9866} \times 3 = 1.91 .$$

The relations between the partial regression coefficient and the different correlation coefficients are illustrated in Figure 1.

5.1 Alternative formulas for the semi-partial correlation coefficients

The semi-partial coefficient of correlation can also be computed directly from the different coefficients of correlation of the independent variables and the dependent variable. Specifically, we find that the semi-partial correlation between Y and X can be computed as

$$r^2_{Y.X|T} = \frac{(r_{Y.X} - r_{Y.T}r_{X.T})^2}{1 - r^2_{X.T}} \tag{12}$$

For our example, taking into account that

- $r_{X.T} = .7500$
- $r_{Y.X} = .8028$
- $r_{Y.T} = .9890$

we find that

$$r^2_{Y.X|T} = \frac{(r_{Y.X} - r_{Y.T}r_{X.T})^2}{1 - r^2_{X.T}} = \frac{(.8028 - .9890 \times .7500)^2}{1 - .7500^2} \approx .0085 . \tag{13}$$

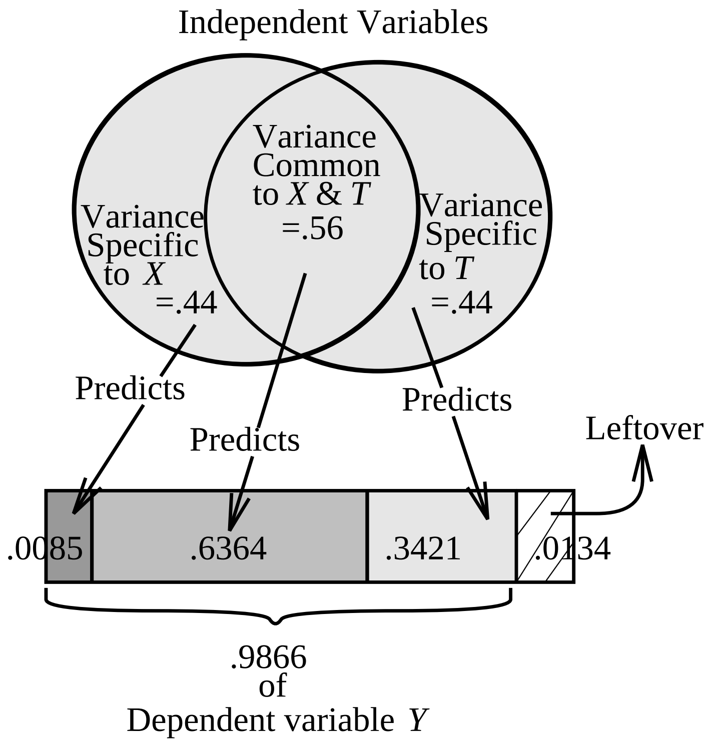

Figure 1: Illustration of the relationship of the independent variables with the dependent variable showing what part of the independent variables explains what proportion of the dependent variable. The independent variables are represented by a Venn diagram, and the dependent variable is represented by a bar.

6 Partial correlation

When dealing with a set of dependent variables, we sometimes want to evaluate the correlation between two dependent variables after the effect of a third dependent variable has been removed from *both* dependent variables. This can be obtained by computing the coefficient of correlation between the residuals of the prediction of each of the first two dependent variables by the third dependent variable (*i.e.,* if you want to eliminate the effect of say variable Q from variables Y and W, you, first, predict Y from Q and W from Q, and then you compute the residuals and correlate them).

This coefficient of correlation is called a *partial coefficient* of correlation. It can also be computed directly using a formula involving only the coefficients of correlation between pairs of variables. As an illustration, suppose that we want to compute the square partial coefficient of correlation between Y and X after having eliminated the effect of T from both of them (this is done only for illustrative purposes because X and T are independent variables, not dependent variable). This coefficient is noted $r^2_{(Y.X)|T}$ (read "r square of Y and X after T has been partialed out from Y and X"), it is computed as

$$r^2_{(Y.X)|T} = \frac{(r_{Y.X} - r_{Y.T}r_{X.T})^2}{(1 - r^2_{Y.T})(1 - r^2_{X.T})} . \tag{14}$$

For our example, taking into account that

- $r_{X.T} = .7500$,
- $r_{Y.X} = .8028$, and
- $r_{Y.T} = .9890$,

we find the following values for the partial correlation of Y and X:

$$r^2_{(Y.X)|T} = \frac{(r_{Y.X} - r_{Y.T}r_{X.T})^2}{(1 - r^2_{Y.T})(1 - r^2_{X.T})} = \frac{(.8028 - .9890 \times .7500)^2}{(1 - .9890^2)(1 - .7500^2)} \approx .3894 . \tag{15}$$

References

[1] Abdi, H., Dowling, W.J., Valentin, D., Edelman, B., & Posamentier M. (2002). *Experimental Design and research methods. Unpublished manuscript.* Richardson: The University of Texas at Dallas, Program in Cognition.