[Mid-Western Educational Researcher](https://scholarworks.bgsu.edu/mwer)

Volume 9
Issue 4 *Regression Analyses*

Article 4

1996

Semi-Partial Correlations: I Don't Need Them; You Can Have Them

Thomas R. Knapp
The Ohio State University

Semi-Partial Correlations: I Don't Need Them; You Can Have Them

Thomas R. Knapp, The Ohio State University

Prologue

I have been teaching statistics and associated topics (measurement, research design) for 37 years and have contributed to the methodological literature on such matters. During that time I have managed to get along without knowing or caring very much about a variety of techniques, most notably exploratory data analysis, Bayesian inference, expected values of mean squares, and item response theory. In the essay that follows I talk about another one: semi-partial correlations.

What are semi-partial correlations?

As explained very nicely by Cohen and Cohen (1983), Darlington (1990), and others, a semi-partial correlation between an independent variable X and a dependent variable Y, is the correlation between Y and the "residualized" variable X.W for which the effect of a covariate W on X has been removed (partialled out, statistically controlled, etc.) from X (but not from Y). This semi-partial correlation (called a "part" correlation by some authors, e.g., McNemar, 1962) and its square are said to be the best indicators of an independent variable's "unique" contribution to the prediction or explanation of the dependent variable. Darlington lists five ways for determining the relative order of importance of independent variables in a multiple regression analysis, and he comes down in favor of focussing on semi-partial correlations.

Cohen and Cohen include in their text several Venn diagrams, or "ballantines" (named after the logo for a beer that is no longer brewed), that are alleged to be helpful in determining "variance accounted for" and in distinguishing semi-partial correlations from partial correlations.

Why I don't need them

There are several reasons why I have little or no interest in semi-partial correlations. First and foremost, an independent variable's semi-partial correlation with a dependent variable can be shown to be mathematically identical to the square root of the difference in R-squares for the "full model" hierarchical regression in which the variable under consideration is entered last and the "reduced model" that includes all of the other variables (covariates) that are to be statistically controlled. (See, for example, Pedhazur (1982), pp. 119-123.) I am a strong advocate of hierarchical regression analysis. I believe that the most interesting educational research questions are of the form: "What is the effect of _____ over and above the effect of _____?". I accordingly find change in R-square to be a more intuitively appealing notion than a squared semi-partial correlation coefficient, and since the two <u>are</u> mathematically identical I prefer the former.

Another reason I don't like to emphasize semi-partial correlations has to do with the concept of a residualized variable. Intellectualizing raw variables, deviation variables, standardized variables, log-transformed variables, etc. is difficult enough. The notion of "X without W", which underlies the proper interpretation of a semi-partial correlation, boggles my mind.

A third reason why I don't get excited about semi-partial correlations is that unlike partial correlations they seem to be useful only in regression analyses. Everybody cares about the partial correlation between, say, height and reading achievement, with age partialled out from both variables, as a device for detecting spurious relationships. But semi-partial correlations only arise in a regression context where one of the variables is a response and all of the others are explanatory.

A fourth reason has to do with those "ballantines". The variance of Y is not a thing that can be sliced up; it is an abstract statistical entity. By depicting overlapping circles with one piece "accounted for" by this and another piece "accounted for" by that, there is a serious danger of imputing causality that may not be warranted (in most non-experimental research, for example).

A fifth reason, associated with my second reason, is that I already have enough important statistical concepts to clutter up my brain without adding another one unless it is absolutely necessary. As Yogi Berra once said, a baseball player has just so many hits coming to him each year, so why waste any of them in spring training.

You can have them (if you want them): An example

Consider the simple hypothetical example exploited by Darlington in Chapter 2 of his excellent regression text. A researcher is interested in the effect of exercise on weight loss over and above (statistically controlling for) food intake. The sample data for 10 subjects are displayed in Table 1.

Table 1.

Some data for illustrating semi-partial correlation (Darlington, 1990, p. 33)

ID	Exercise (X)*	Food intake (W)*	Weight loss (Y)
1	0	2	6
2	0	4	2
3	0	6	4
4	2	2	8
5	2	4	9
6	2	6	8
7	2	8	5
8	4	4	11
9	4	6	13
10	4	8	9

* Darlington calls these two variables X_1 and X_2, respectively, but I prefer the X and W notation.

Darlington provides the reader with eight informative plots of the data. The first is a simple scatter-plot of Y against W. The next two are the conventional three-dimensional plots, one without and one with the best-fitting plane superimposed. The fourth figure again plots Y against W, with the X values shown by parallel lines in the body of the figure; the fifth figure plots Y against X, with the W values shown by parallel lines. The sixth figure is a simple scatter-plot of X against W. The seventh figure plots Y against W.X; and the eighth figure, which is the key figure regarding the semi-partial correlation for the research question, plots Y against X.W.

I need only the first three figures to make geometrical sense of what is going on. (How about you?) And algebraically (or arithmetically) I need the Pearson r and its square for the reduced-model first plot (they are .047 and .002, respectively—the covariate W actually had very little effect) and the multiple R and its square for the full-model second and third plots (they are .915 and .838, respectively). Therefore the magnitude of the effect of X on Y over and above W is given by the difference between the .838 and the .002 (.836, which is equal to the squared semi-partial correlation) and/or the square root of that difference (.914, which is the semi-partial correlation itself).

At the end of Chapter 2 Darlington gives equivalent formulas for semi-partial correlations, as well as formulas for partial correlations and beta weights. I don't need those formulas for semi-partial correlations (do you?), but in fairness to those who do, it is instructive to note the similarities among the formulas for semi-partial correlations, partial correlations, and beta weights (same numerators, different denominators).

Later in his text Darlington discusses hypothesis testing and estimation for various regression statistics (Chapter 5) and provides his argument for preferring semi-partial correlations (Chapter 9) as indicators of the relative importance of independent variables in a regression (as opposed to change in R-square and three other methods for ranking regressors). Reference to semi-partial correlations per se is interestingly absent in Chapter 5, but he does provide the formula for testing the significance of change in R-square (which implicitly tests the significance of a semi-partial correlation).

Epilogue

The probability is very small that the foregoing remarks will sway all readers of this journal to my point of view. I am a methodological loner (you should hear me expound on my idiosyncratic notion of validity!) and I rather enjoy being in that position. But if nothing else I hope that this essay may serve to generate some interesting discussion. Isn't that what it's all about?

References

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd. ed.). Hillsdale, NJ: Erlbaum.

Darlington, R.B. (1990). *Regression and linear models*. New York: McGraw-Hill.

McNemar, Q. (1962). *Psychological statistics* (3rd. ed.). New York: Wiley.

Pedhazur, E.J. (1982). *Multiple regression in behavioral research* (2nd. ed.). New York: Holt, Rinehart, and Winston.