Partial and Semipartial Correlations
© Stevens, 2003

The purpose of this handout is to briefly introduce partial and semipartial correlations and describe their use in multiple regression analysis. The essential concept embodied in these coefficients is the estimation of the relationship between a predictor variable and a criterion or outcome variable *after controlling for the effects* of other predictors in the equation. This process of exercising statistical control is also known as partialing or residualization. Also note that another name for the semipartial is the part correlation.

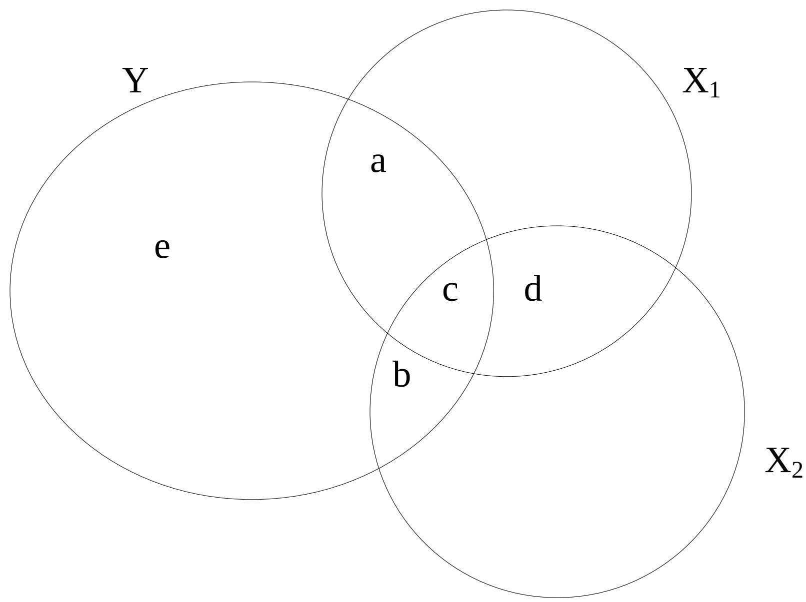
Partialing attempts to determine the degree of association between two variables that would exist if all influences of one or more other variables could be removed. Partialing represents a method of exerting statistical control over variables. It is important to distinguish statistical control from experimental control (e.g., random assignment to treatments, control by constancy, etc.). Generally, experimental control provides stronger evidence than statistical control because it is directly managed by the researcher and planned a priori.

In multiple regression, the two most commonly used coefficients are the unstandardized and standardized partial regression coefficients. As their names imply, they involve partialling and represent the unique relationship between a given predictor and the criterion while controlling for all other predictors in the equation. Unstandardized coefficients are expressed in the scale units of the predictor of interest. Standardized coefficients are expressed on a standardized scale that is expressed as a ratio of the standard deviations of Y and X. Another more general name that is used in the General Linear Model for these coefficients is function coefficients.

A <u>partial correlation coefficient</u> is another third way of expressing the unique relationship between the criterion and a predictor. Partial correlation represents the correlation between the criterion and a predictor after common variance with other predictors has been removed from <u>*both*</u> the criterion and the predictor of interest. That is, after removing variance that the criterion and the predictor have in common with other predictors, the partial expresses the correlation between the residualized predictor and the residualized criterion.

A <u>semipartial correlation coefficient</u> represents the correlation between the criterion and a predictor that has been residualized with respect to all other predictors in the equation. Note that the criterion remains unaltered in the semipartial. Only the predictor is residualized. After removing variance that the predictor has in common with other predictors, the semipartial expresses the correlation between the residualized predictor and the unaltered criterion. An important advantage of the semipartial is that the denominator of the coefficient (the total variance of the criterion, Y) remains the same no matter which predictor is being examined. This makes the semipartial very interpretable. The square of the semipartial can be interpreted as the proportion of the criterion variance associated uniquely with the predictor. It is also possible to use the semipartial to fully deconstruct the variance components in a regression analysis. Since each squared semipartial represents the unique variance of that predictor shared with the criterion, the sum of the squared semipartials can be subtracted from the overall R2 for the regression equation as a while to determine the amount of common variance in the equation shared by multiple predictors with the criterion. This common variance in nonexperimental research is likely not interpretable and of greatest use when the purpose of the research is predictive.

The diagram on the next page illustrates these relationships between a criterion variable, Y, and two predictor variables, X_1 and X_2. Areas of shared variance are labeled with lower case letters and these are then used to define coefficients.



Total variance of Y = a + b + c + e

Zero-order correlations:
r^2_{Y1} = a + c
r^2_{Y2} = b + c
r^2_{12} = c + d

Variance of Y explained by regression model = a + b + c

Variance of Y not explained by regression model = e

R^2_{Y12} = (a + b + c) / (a + b + c + e)

Multicollinearity = c + d

Semipartial for X_1 = $sr^2_1 = r^2_{Y(1.2)} = R^2_{Y12} - r^2_{Y2}$ = a / (a + b + c + e)

Semipartial for X_2 = $sr^2_2 = r^2_{Y(2.1)} = R^2_{Y12} - r^2_{Y1}$ = b / (a + b + c + e)

Partial for X_1 = $pr^2_1 = r^2_{Y1.2} = (R^2_{Y12} - r^2_{Y2}) / (1 - r^2_{Y2})$ = a / (a + e)

Partial for X_2 = $pr^2_2 = r^2_{Y2.1} = (R^2_{Y12} - r^2_{Y1}) / (1 - r^2_{Y1})$ = b / (b + e)

$pr^2 > sr^2$ except when $r^2_{12} = 0$