EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT
1978, 38

# SUPPRESSOR VARIABLES AND THE SEMIPARTIAL CORRELATION COEFFICIENT

WAYNE F. VELICER
University of Rhode Island

Since Horst (1941) initially introduced the concept, a number of investigators have suggested alternative definitions that both include a broader class of situations and are more precise. Following a conceptual framework suggested by Conger (1974), the present paper discusses a definition based on the relation of the semipartial correlation to the zero order correlation. This definition possesses a number of advantages, including the fact that it can be easily extended from the two predictor case to the general $p$-predictor case.

SINCE the concept of a suppressor variable was first introduced by Horst (1941), a great deal of attention has been paid to the question of how and why a suppressor variable works. Since a suppressor variable requires no modification of the basic multiple regression model, the only practical problem is that potential suppressor variables may be ignored because of their low zero order validities. However, the concept is important for a full understanding of how a multiple regression system works. The concept of a suppressor variable has already generated a variety of empirical research in the area of response style (Wiggins, 1973).

The critical issue is a precise definition of what a suppressor is. Since Horst's (1941) rather limited definition, more general definitions which subsume the Horst suppressor have been proposed by Lubin (1957) and Darlington (1968). The best attack on this problem is Conger's recent paper (1974) which defines a regression weight in terms of its effect on the regression weight of the other predictor. Conger argues that: "the question of interest is: under what conditions . . . will the multiple regression contribution of the first predictor exceed its zero-order contribution . . ." (Conger, 1974, p. 37). This definition is operationalized mathematically in terms of the standardized regression weight ($B_1$) and the zero order validity ($r_{01}$). The second predictor is a suppressor if

$$|B_1| > |r_{01}| \tag{1}$$

or, equivalently,

$$B_1^2 > r_{01}^2 \tag{2}$$

An immediate problem with this definition is the fact that $b_1$ and $r_{01}$ are not strictly comparable; consider that $r_{01}$ is bounded above by unity and $b_1$ is not. While agreeing with Conger's approach to the problem, the present paper will consider an alternative means of operationalizing the concept in mathematical terms that possess a number of advantages, most important of which is that the definition can be directly extended to the general multiple regression situation.

First, consider the two predictor situation where $X_0$ is the criterion variable and $X_1$ and $X_2$ are two predictors. The second predictor ($X_2$) will be arbitrarily designated as the suppressor variable. If the two predictor variables are uncorrelated, the mutiple regression coefficient ($R^2$) can be expressed as

$$R_{0.12}^2 = r_{01}^2 + r_{02}^2 \tag{3}$$

where $r_{0i}$ ($i = 1, 2$) is the correlation between the $i$th predictor and the criterion. In the typical prediction situation,

$$R_{0.12}^2 < r_{01}^2 + r_{02}^2 \tag{4}$$

because of redundancy between the predictors. However, the situation where

$$R_{0.12}^2 > r_{01}^2 + r_{02}^2 \tag{5}$$

will be defined as the situation where a suppressor variable is present.

The multiple regression coefficient may also be expressed as

$$R_{0.12}^2 = r_{01}^2 + r_{0(2.1)}^2 \tag{6}$$

where $r_{0(2.1)}^2$ is the semipartial (or part) correlation squared, or

$$r_{0(2.1)}^2 = \frac{(r_{02} - r_{12}\, r_{01})^2}{1 - r_{12}^2} \tag{7}$$

and it follows directly from equation (3) that, in a suppressor situation,

$$r_{0(2.1)}^2 > r_{02}^2 \tag{8}$$

Equation (5) defines a suppressor variable in terms of the multiple correlation coefficient and Equation (6) defines a suppressor variable in terms of the squared semipartial correlation of the suppressor variable.

The reader should note that the designation of $X_2$ as the suppressor is arbitrary. The relation expressed by equation (6) holds equally well if defined on the predictor. This relationship can be stated as a theorem.

Theorem: Given a criterion variable ($X_0$) and two predictors ($X_1$, $X_2$) and the usual least squares multiple regression solution, if

$$r_{0(2.1)}^2 > r_{02}^2 \tag{9}$$

then

$$r_{0(1.2)}^2 > r_{01}^2 \tag{10}$$

Proof: Expanding the relationship in terms of the second variable yields

$$\frac{r_{02}^2 - 2\, r_{01}\, r_{02}\, r_{12} + r_{01}^2\, r_{12}^2}{1 - r_{12}^2} > r_{02}^2 \tag{11}$$

or

$$r_{02}^2 - 2\, r_{01}\, r_{02}\, r_{12} + r_{01}^2\, r_{12}^2 > r_{02}^2 - r_{02}^2\, r_{12}^2. \tag{12}$$

Then, by subtracting $r_{02}^2$ from both sides, adding $r_{01}^2$ to both sides, and rearranging terms, the inequality can be rewritten as

$$r_{01}^2 - 2\, r_{01}\, r_{02}\, r_{12} + r_{02}^2\, r_{12}^2 > r_{01}^2 - r_{01}^2\, r_{12}^2 \tag{13}$$

which immediately yields

$$r_{0(1.2)}^2 > r_{01}^2 \tag{14}$$

The method employed in the proof can also be used to demonstrate that

$$r_{0(1.2)}^2 - r_{01}^2 = r_{0(2.1)}^2 - r_{02}^2, \tag{15}$$

a relationship that holds generally.

The theorem shows that the definition of a suppressor is reciprocal, i.e., the definition works equally well if applied to the "predictor" or the "suppressor" variable. The reciprocal nature of the definition has the advantage that the method can be applied to determine if a suppressor is present without a priori identification of the suppressor. It also has the disadvantage that the method will not identify which of the two variables is the suppressor.

This reciprocal property will also hold in most cases for the Conger index. If,

$$b_1^2 > r_{01}^2 \tag{16}$$

then, generally,

$$b_2^2 > r_{02}^2 \tag{17}$$

The relations developed above can be used to assess the situation where (16) does not imply (17). First of all

$$b_1^2 > r_{01}^2 \tag{18}$$

implies

$$r_{0(1.2)}^2 - r_{01}^2 + r_{01}^2 r_{12}^2 > 0 \tag{19}$$

and also

$$r_{0(2.1)}^2 - r_{02}^2 + r_{01}^2 r_{12}^2 > 0 \tag{20}$$

If we assume the converse to (17), i.e.,

$$b_2^2 < r_{02}^2 \tag{21}$$

then

$$r_{0(2.1)}^2 - r_{02}^2 < 0 \tag{22}$$

and also

$$r_{0(1.2)}^2 - r_{01}^2 < 0 \tag{23}$$

and using both (16) and (21) results in the following inequality

$$r_{01}^2 r_{12}^2 > r_{01}^2 - r_{0(1.2)}^2 = r_{02}^2 - r_{0(2.1)}^2 > r_{02}^2 r_{12}^2 \tag{24}$$

which defines the situation where the Conger definition is not reciprocal.

The advantage of defining a suppressor situation in terms of the squared semipartial correlation of the suppressor variable, rather than in terms of the regression weight of the predictor variable is that this formulation can be immediately extended to the general case of the $p$ predictors.

*Definition.* The $p$th predictor is a suppressor variable if and only if

$$R_{0.12\cdots p}^2 > R_{0.12\cdots p-1}^2 + r_{0p.}^2 \tag{25}$$

This definition treats the linear combination of the first $p - 1$ predictors as a single variable, and reduces the general $p$-predictor problem to a two predictor situation. This definition could also be formulated in terms of the semipartial correlation of the $p$th predictor as

$$r_{0(p.12\cdots p-1)} > r_{0p.}^2 \tag{26}$$

The definition provided by Equation (25) (or alternately (26)) yields a simple, precise, practical means of determining if suppressor variable is present. Note again that the definition is reciprocal. The linear combination represented by the first $p - 1$ variable could also be viewed as a "suppressor" for the $p$th predictor. As an example of how this definition applies, consider Darlington's (1968) example for three predictors where

$$r_{01} = r_{02} = .15,\ r_{03} = .20 \text{ and } r_{12} = .00,\ r_{13} = r_{23} = .70.$$

The third predictor ($X_3$) was designated as the suppressor. Since

$$R_{0.12}^2 = .045,$$

$$R_{0.12}^2 + r_{03}^2 = .085,$$

and

$$R_{0.123}^2 = .25,$$

the example meets the requirement of equation (25) that

$$R_{0.123}^2 > R_{0.12}^2 + r_{03}^2.$$

Defining a suppressor in terms of the semipartial correlations is consistent with some previous discussions of the problem. Conger and Jackson (1972) discuss the effect of a Horst suppressor on the semipartial and Darlington (1968) suggests, but does not develop, a definition in terms of variance accounted for. The present definition possesses a number of advantages: (1) The definition applies to the general case of $p$-predictors. (2) The definition is general and can be applied to each of the three types of suppressor variables discussed by Conger (1974; see Table 1 for numeric examples). (3) The definition is consistent with stepwise regression procedures. (4) The definition is consistent with the usual operational interpretation for predictors, i.e., proportion of variance accounted for. (5) The definition does not depend on an a priori identification of a suppressor variable. A dis-

TABLE 1
*Numeric Examples of Suppressor Situations*

| Type<br>Source | Classical<br>McNemar (1945) | Reciprocal<br>Conger (1974) | Negative |
|---|---|---|---|
| $r_{01}$ | .40 | .50 | .50 |
| $r_{02}$ | .00 | .30 | .10 |
| $r_{12}$ | .707 | −.27 | .71 |
| $r_{0(2.1)}^2$ | .16 | .204 | .13 |
| $r_{02}^2$ | .00 | .09 | .01 |
| $r_{0(1.2)}^2$ | .32 | .364 | .37 |
| $r_{01}^2$ | .16 | .25 | .25 |

advantage is that the definition will identify when a suppressor variable is present, but not specifically which variable is the suppressor. Designation of a variable as the "suppressor" would require knowledge of how a suppressor variable works. Readers interested in a discussion of how suppressor variables work are referred to the work of Horst (1941) Meehl (1945), Lubin (1957), Darlington (1968), Lord and Novick (1968), Conger and Jackson (1972), and, especially, Conger (1974).

## REFERENCES

Conger, A. A revised definition for suppressor variables: A guide to their interpretation and inter prediction. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1974, 34, 35–46.

Conger, A. J. and Jackson, D. N. Suppressor variables, prediction, and the interpretation of psychological relationships. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENTS, 1972, 32, 579–599.

Darlington, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin,* 1968, 69, 161–182.

Horst, P. The prediction of personal adjustment. *Social Science Research Council Bulletin,* No. 48. New York: 1941.

Lord, F. M. and Novick, M. R. *Statistical Theories of Mental Test Scores.* Reading, Mass., Addison-Wesley, 1968.

Lubin, A. Some formulas for use with suppressor variables. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1957, 17, 286–296.

Meehl, P. E. A simple algebraic development of Horst's suppressor variables. *American Journal of Psychology,* 1945, 58, 550–554.

Wiggins, J. S. *Personality and prediction: principles of personality assessment.* Reading, Mass.: Addison-Wesley, 1973.