Suppression

Definitions

As one can picture from the Venn diagram, the inclusion of an additional variable in a regression model will tend to decrease the slope of other variables in the model to the extent that the added variable is correlated with the other predictors and the outcome. So, we expect the standardized regression coefficient in a multiple regression to be equal to or smaller than the ("zero-order") correlation of that variable with the outcome. In this sense, we use the term partial regression coefficient, because it represents only part of the overall relationship with the outcome—the unique part that does not have common overlap with other predictors and the outcome. Occasionally, this is not the case, and the addition of a predictor has the effect of change the other predictors in a counterintuitive way.

The term suppression, originally coined by Horst (1941, 1966), was used to describe how verbal ability, which had no correlation with successful training, could increase the $R^2$ and the size of the other coefficients. Adding verbal ability increased the effect of spatial ability in predicting success in training. Verbal ability was suppressing the effects of the other variables, and, after it was taken into account, the real effects of spatial ability were seen. The explanation here is that pilots needed to have enough reading ability to understand the questions of the spatial ability test. Since Horst, others have considered the suppression effect in broader terms, and the definition has expanded over time. Originally, as in the Horst example, the idea was that the suppressor variable was a variable that was uncorrelated with the outcome. Conger (1974) described a suppressor variable as one that increases the "predictive value" of the one or more of the other independent variables in the model. Later definitions and more common use of the term, however, have tended to include cases in which one or more coefficients in the model reverse sign after the inclusion of a new predictor (see Krus & Wilkinson, 1986, for several specific suppression forms). Although any of these effects may seem counterintuitive (and they do indeed tend to be relatively rarely observed), suppression effects are not really a mystery mathematically nor is there anything wrong with the analysis.

Explanation

Suppression effect is relatively easy to explain algebraically, when you consider the equation that shows how the standardized coefficient is derived from the correlations among $Y$, $X_1$, and $X_2$. For a model with both $X_1$, and $X_2$ included as predictors.

$$\beta_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}$$

One can think of the above equation for the standardized regression coefficient for the first predictor, $\beta_1$, as a correction to the correlation between $X_1$ and $Y$ (i.e., $r_{y1}$), where the second term in the numerator, $r_{y2}r_{12}$, is subtracted. Assuming for a moment that the correlation $r_{y1}$ is positive, the value of $\beta_1$ will tend to increase if either $r_{y1}$ or $r_{12}$ are negative, because subtraction of a negative quantity is the same as adding that quantity. In such a case, $\beta_1$ will be larger than $r_{y1}$, which is not the expected direction. As a second example, assuming again that $r_{y1}$ is positive, but, this time, both $r_{y2}$ and $r_{12}$ are both positive. If their product, $r_{y2}r_{12}$, happens to be larger than $r_{y1}$, then $\beta_1$ will be negative, in essence changing an initial positive association with the dependent variable into a negative (partial) association. Note also that the correlation between $X_1$ and $X_2$ can make a difference in the estimate of $\beta_1$, because of its appearance in the denominator of the equation. Larger values of $r_{12}^2$ will tend to increase the value of $\beta_1$.

Conceptually, understanding suppression can be a bit more challenging and it is not always simple to explain the effects theoretically. One can sometimes envision the existence of another factor, as in the case of the verbal ability of pilots, that is interfering or clouding up the ability to see the relationship of the other variables in the model to the outcome; and once the suppressor is included in the model and its obscuring effects have been removed, we can more clearly observe the effect of the other predictors. If we were interested in the effects that college entrance exam scores have on mood the day they are received by the student (a positive correlation, presumably), it may be that the impact of more pleasant
weather on mood may obscure the relationship between scores and achievement to some extent. After controlling for the weather, there may be a relationship between scores and mood, even though mood is initially uncorrelated with achievement scores. But because the weather may be good on some days that students receive their scores and bad on other days that other students receive their scores, weather is a source of "error" that increases the variability of mood. If we can measure it and remove it, we may see a clearer relationship between weather and mood.

In other circumstances, suppression may occur because two of correlations among the three variables are positive and one of the correlations is negative (or vice versa). Examining the correlation matrix among the three variables will be illuminating, and there may be a variety of theoretical explanations for the seemingly complex relationship among the three variables. In one type of explanation, suppression may be a result of a certain type of mediational model, known as inconsistent mediation. Mediation models involve a causal chain of events, where say \( X_1 \) causes \( X_2 \) which, in turn, causes \( Y \). MacKinnon and colleagues (MacKinnon, Krull, & Lockwood, 2000) present a hypothetical example in which intelligence leads to boredom, which, in turn, leads to more assembly line errors. Intelligence initially would be expected to be negatively related to assembly line errors overall, yet intelligence might be found to be positively associated with assembly line errors if a certain causal chain of events holds. If it is the case that intelligence is positively associated with boredom and boredom is positively associated with assembly line errors, an inconsistent mediation model would occur. Once boredom is controlled in the relationship between intelligence and assembly line errors, the partial regression coefficient between intelligence and assembly line errors will be negative as expected. Without taking boredom into account, it may appear that intelligence is positively related to assembly line errors.

Simple regression without taking boredom into account:

\[
\begin{align*}
\text{Intelligence} & \rightarrow \text{Boredom} \\
& \rightarrow \text{Assembly line errors}
\end{align*}
\]

Multiple regression reveals negative effect of intelligence on errors once boredom is taken into account:

\[
\begin{align*}
\text{Intelligence} & \rightarrow \text{Boredom} \\
& \rightarrow \text{Assembly line errors}
\end{align*}
\]

In practice, if you simply look at the correlation matrix and then observe a suppression effect upon testing a larger regression model with several predictors, it may be difficulty to know which variable is causing the suppression effect (or, in fact, whether several predictors together are causing the effect). You may have to run a series of regression models with and without different predictors to narrow down which variable is the suppressor variable.

Darlington Food Intake Example

Darlington (1990) gives another suppression example. Why might higher calorie intake be associated with decreased weight?

get file='c:\jason\spsswin\da2\food.sav'
correlations vars=food ex wt.
regression vars=wt food ex
    /descriptives=mean stdev
    /statistics=anova coeff ses r ci
    /dependent=wt
    /method=enter food /enter ex.

---

1 Based on example given by Darlington, R. B. (1990). *Regression and linear models* (pp. 292-293). New York: McGraw-Hill. Note that weightloss was changed to weight (wt) for clearer presentation.
## Correlations

<table>
<thead>
<tr>
<th></th>
<th>food intake</th>
<th>ex exercise</th>
<th>wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>food intake</td>
<td>Pearson Correlation</td>
<td>.378</td>
<td>-.047</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.282</td>
<td>.898</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>ex exercise</td>
<td>Pearson Correlation</td>
<td>.378</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.282</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>wt</td>
<td>Pearson Correlation</td>
<td>-.047</td>
<td>-.864</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.898</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

## Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>95.0% Confidence Interval for B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>5.857</td>
<td>2.923</td>
</tr>
<tr>
<td>food food intake</td>
<td>-.071</td>
<td>.541</td>
<td>-.047</td>
</tr>
<tr>
<td>2</td>
<td>(Constant)</td>
<td>7.000</td>
<td>1.725</td>
</tr>
<tr>
<td>food food intake</td>
<td>.500</td>
<td>.252</td>
<td>.326</td>
</tr>
<tr>
<td>ex exercise</td>
<td>-2.000</td>
<td>.333</td>
<td>-.987</td>
</tr>
</tbody>
</table>

*a. Dependent Variable: wt*
What do you think the mediational model behind this might be?

References and Further Reading
Horst, P. (1941). The role of predictor variables which are independent of the criterion. Social Science Research Bulletin, 48, 431-436.