ONE STEP AT A TIME ***(continued)***

This outcome is evidence that the control variable (*Z*) has little or no effect and that the relationship between *X* and *Y* is direct.

2. The partial correlation coefficient is much less (say, more than 0.10 less) than the bivariate correlation. This is evidence that the control variable (*Z*) changes the relationship between *X* and *Y*. The relationship between *X* and *Y* is either spurious (*Z* causes *X* and *Y*) or intervening (*X* and *Y* are linked by *Z*).

Be aware that *X*, *Y*, and *Z* may have an interactive relationship in which the relationship between *X* and *Y* changes for each category of *Z*. Partial correlation analysis cannot detect interactive relationships.

STATISTICS IN EVERYDAY LIFE

An analysis of a representative sample of Americans aged 25 to 65 found a correlation of 0.43 between years of schooling and income—a moderate positive relationship. Income increases as education increases and education explains almost 19% of the variation in income.

Is this relationship affected by age? The partial correlation between education and income while controlling for age is also 0.43, the same value as the zero-order correlation. This indicates that the relationship between education and income is direct: Regardless of age, income increases as education increases.

People with higher levels of education tend to have higher income regardless of their age—a finding that you may find reassuring as you move to the completion of your degree.

Multiple Regression: Predicting the Dependent Variable

In Chapter 13, the least-squares regression line was introduced as a way of describing the overall linear relationship between two interval-ratio variables and of predicting scores on *Y* from scores on *X*. This line was the best-fitting line to summarize the bivariate relationship and was defined by this formula:

FORMULA 14.2

$$Y = a + bX$$

where: a = the Y intercept
b = the slope

The least-squares regression line can be modified to include (theoretically) any number of independent variables. This technique is called **multiple regression**. For ease of explication, we will confine our attention to the case involving two independent variables. The least-squares multiple regression equation is:

FORMULA 14.3

$$Y = a + b_1X_1 + b_2X_2$$

where: b_1 = the partial slope of the linear relationship between the first independent variable and Y
b_2 = the partial slope of the linear relationship between the second independent variable and Y

Some new notation and some new concepts are introduced in this formula. First, while the dependent variable is still symbolized as Y, the independent variables are differentiated by subscripts. Thus, X_1 identifies the first independent variable and X_2 the second. The symbol for the slope (b) is also subscripted to identify the independent variable with which it is associated.

Partial Slopes

A major difference between the multiple and bivariate regression equations concerns the slopes (b's). In the case of multiple regression, the b's are called **partial slopes**, and they show the amount of change in Y for a unit change in the independent variable while controlling for the effects of the other independent variables in the equation. The partial slopes are thus analogous to partial correlation coefficients and represent the direct effect of the associated independent variable on Y.

Computing Partial Slopes. The partial slopes for the independent variables are determined by Formulas 14.4 and 14.5:

FORMULA 14.4

$$b_1 = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}\right)$$

FORMULA 14.5

$$b_2 = \left(\frac{s_y}{s_2}\right)\left(\frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2}\right)$$

where: b_1 = the partial slope of X_1 on Y
b_2 = the partial slope of X_2 on Y
s_y = the standard deviation of Y
s_1 = the standard deviation of the first independent variable (X_1)
s_2 = the standard deviation of the second independent variable (X_2)
r_{y1} = the bivariate correlation between Y and X_1
r_{y2} = the bivariate correlation between Y and X_2
r_{12} = the bivariate correlation between X_1 and X_2

To illustrate the computation of the partial slopes, we will assess the combined effects of number of children (X_1) and husband's education (X_2) on husband's contribution to housework. All the relevant information can be calculated from Table 14.1 and is reproduced here:

Husband's Housework	Number of Children	Husband's Education
$\bar{Y} = 3.3$	$\bar{X}_1 = 2.7$	$\bar{X}_2 = 13.7$
$s_y = 2.1$	$s_1 = 1.5$	$s_2 = 2.6$
	Zero-Order Correlations	
	$r_{y1} = 0.50$	
	$r_{y2} = -0.30$	
	$r_{12} = -0.47$	

The partial slope for the first independent variable (X_1) is:

$$b_1 = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}\right)$$

$$b_1 = \left(\frac{2.1}{1.5}\right)\left(\frac{0.50 - (-0.30)(-0.47)}{1 - (-0.47)^2}\right)$$

$$b_1 = (1.4)\left(\frac{0.50 - 0.14}{1 - 0.22}\right)$$

$$b_1 = (1.4)\left(\frac{0.36}{0.78}\right)$$

$$b_1 = (1.4)(0.46)$$

$$b_1 = 0.65$$

For the second independent variable (SES or X_2), the partial slope is:

$$b_2 = \left(\frac{s_y}{s_2}\right)\left(\frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2}\right)$$

$$b_2 = \left(\frac{2.1}{2.6}\right)\left(\frac{-0.30 - (+.50)(-0.47)}{1 - (-0.47)^2}\right)$$

$$b_2 = (0.81)\left(\frac{-0.30 - (-0.24)}{1 - 0.22}\right)$$

$$b_2 = (0.81)\left(\frac{-0.30 + 0.24}{0.78}\right)$$

$$b_2 = (0.81)\left(\frac{-0.06}{0.78}\right)$$

$$b_2 = (0.81)(-0.08)$$

$$b_2 = -0.07$$

ONE STEP AT A TIME Computing and Interpreting Partial Slopes

These procedures apply when there are two independent variables and one dependent variable. For more complex situations, use a computerized statistical package, such as SPSS, to do the calculations:

Step ***Operation***

To compute the partial slope associated with the first independent variable by using Formula 14.4:

1. Divide s_y by s_1.
2. Multiply r_{y2} by r_{12}.
3. Subtract the value you computed in step 2 from r_{y1}.
4. Square the value of r_{12}.
5. Subtract the value you computed in step 4 (r_{12}^2) from 1.
6. Divide the value you computed in step 3 by the value you computed in step 5.
7. Multiply the value you computed in step 6 by the value you computed in step 1. *This value is the partial slope associated with the first independent variable.*

ONE STEP AT A TIME ***(continued)***

To compute the partial slope associated with the second independent variable by using Formula 14.5:

1. Divide s_y by s_2.
2. Multiply r_{y1} by r_{12}.
3. Subtract the value you computed in step 2 from r_{y2}.
4. Square the value of r_{12}.
5. Subtract the value you computed in step 4 from 1.
6. Divide the value you computed in step 3 by the value you computed in step 5.
7. Multiply the value you computed in step 6 by the value you computed in step 1. *This value is the partial slope associated with the second independent variable.*

To Interpret Partial Slopes

The value of a partial slope is the increase in the value of *Y* for a unit increase in the value of the associated independent variable while controlling for the effects of the other independent variable.

Finding the *Y* Intercept. Now that partial slopes have been determined for both independent variables, the *Y* intercept (*a*) can be found. Note that *a* is calculated from the mean of the dependent variable (symbolized as $\bar{Y}$) and the means of the two independent variables ($\bar{X}_1$ and $\bar{X}_2$):

FORMULA 14.6

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

Substituting the proper values for the example problem at hand, we would have:

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$
$$a = 3.3 - (0.65)(2.7) - (-0.07)(13.7)$$
$$a = 3.3 - (1.8) - (-1.0)$$
$$a = 3.3 - 1.8 + 1.0$$
$$a = 2.5$$

The Least-Squares Multiple Regression Line and Predicting *Y*'. For our example problem, the full least-squares multiple regression equation would be:

$$Y = a + b_1X_1 + b_2X_2$$
$$Y = 2.5 + (0.65)X_1 + (-0.07)X_2$$

ONE STEP AT A TIME **Computing the *Y* Intercept**

Step ***Operation***

To find the Y intercept by using Formula 14.6:

1. Multiply the mean of X_2 by b_2.
2. Multiply the mean of X_1 by b_1.
3. Subtract the quantity you found in step 1 from the quantity you found in step 2.
4. Subtract the quantity you found in step 3 from the mean of *Y*. *The result is the value of a: the Y intercept.*

ONE STEP AT A TIME	Using the Multiple Regression Line to Predict Scores on *Y*
Step	***Operation***
1.	Choose a value for X_1. Multiply this value by the value of b_1.
2.	Choose a value for X_2. Multiply this value by the value of b_2.
3.	Add the values you found in steps 1 and 2 to the value of *a*: the *Y* intercept. *The resulting value is the predicted score on Y.*

As was the case with the bivariate regression line, this formula can be used to predict scores on the dependent variable from scores on the independent variables. For example, what would be our best prediction of husband's housework (Y') for a family of four children ($X_1 = 4$), where the husband had completed 11 years of schooling ($X_2 = 11$)? Substituting these values into the least-squares formula, we would have:

$$Y' = 2.5 + (0.65)(4) + (-.07)(11)$$
$$Y' = 2.5 + 2.6 - 0.8$$
$$Y' = 4.3$$

Our prediction would be that this husband would contribute 4.3 hours per week to housework. This prediction is, of course, a kind of "educated guess," which is unlikely to be perfectly accurate. However, we will make fewer errors of prediction using the least-squares line (and, thus, incorporating information from the independent variables) than we would using any other method of prediction (assuming, of course, that there is a linear association between the independent and the dependent variables). (*For practice in predicting Y scores and in computing slopes and the Y intercept, see problems 14.1 to 14.6.*)

Multiple Regression: Assessing the Effects of the Independent Variables

The least-squares multiple regression equation (Formula 14.3) is used to isolate the separate effects of the independents and to predict scores on the dependent variable. However, in many situations, using this formula to determine the relative importance of the various independent variables will be awkward, especially when the independent variables differ in terms of units of measurement (e.g., number of children vs. years of education). When the units of measurement differ, a comparison of the partial slopes will not necessarily tell us which independent variable has the strongest effect and is thus the most important. Comparing the partial slopes of variables that differ in units of measurement is a little like comparing apples and oranges.

We can make it easier to compare the effects of the independent variables by converting all variables in the equation to a common scale and thereby eliminating variations in the values of the partial slopes that are solely a function of differences in units of measurement. For example, we can standardize all distributions by changing the scores of all variables to Z scores. Each distribution of scores would then have a mean of 0 and a standard deviation of 1 (see Chapter 5), and comparisons between the independent variables would be much more meaningful.

Computing the Standardized Regression Coefficients

To standardize the variables to the normal curve, we could actually convert all scores into the equivalent Z scores and then recompute the slopes and the Y intercept. This would require a good deal of work; fortunately, a shortcut is available for computing the slopes of the standardized scores directly. These **standardized partial slopes**—called **beta-weights**—are symbolized $\boldsymbol{b^*}$.

Beta-Weights. The beta-weights show the amount of change in the standardized scores of Y for a one-unit change in the standardized scores of each independent variable while controlling for the effects of all other independent variables.

Formulas and Computation for Beta-Weights. When we have two independent variables, the beta-weight for each is found by using Formulas 14.7 and 14.8:

FORMULA 14.7

$$b_1^* = b_1\left(\frac{s_1}{s_y}\right)$$

FORMULA 14.8

$$b_2^* = b_2\left(\frac{s_2}{s_y}\right)$$

We can now compute the beta-weights for our sample problem to see which of the two independent variables has the stronger effect on the dependent. For the first independent variable, number of children (X_1):

$$b_1^* = b_1\left(\frac{s_1}{s_y}\right)$$

$$b_1^* = (0.65)\left(\frac{1.5}{2.1}\right)$$

$$b_1^* = (0.65)(0.71)$$

$$b_1^* = 0.46$$

For the second independent variable, SES (X_2):

$$b_2^* = b_2\left(\frac{s_2}{s_y}\right)$$

$$b_2^* = (-0.07)\left(\frac{2.6}{2.1}\right)$$

$$b_2^* = (-0.07)(1.24)$$

$$b_2^* = -0.09$$

ONE STEP AT A TIME **Computing and Interpreting Beta-Weights (*b**)**

These procedures apply when there are two independent variables and one dependent variable. For more complex situations, use a computerized statistical package, such as SPSS, to do the calculations.

Step	Operation
To compute the beta-weight associated with the first independent variable using Formula 14.7:	
1.	Divide s_1 by s_y.
2.	Multiply the value you found in step 1 by the partial slope of the first independent variable (b_1). *This value is the beta-weight associated with the first independent variable.*
To compute the beta-weight associated with the second independent variable using Formula 14.8:	
1.	Divide s_2 by s_y.
2.	Multiply the value you found in step 1 by the partial slope of the second independent variable (b_2). *This value is the beta-weight associated with the second independent variable.*

To Interpret Beta Weights

A beta-weight (or standardized partial slope) shows the increase in the value of Y for a unit increase in the value of the associated independent variable while controlling for the effects of the other independent variable after all variables have been standardized (or transformed to Z scores).

Comparing the value of the beta-weights, we see that number of children has a stronger effect than SES on husband's housework. Furthermore, the net effect (after controlling for the effect of SES) of the first independent variable is positive, while the net effect of the second independent variable is negative.

The Standardized Least-Squares Regression Line. Using standardized scores, the least-squares regression equation can be written as:

FORMULA 14.9

$$Z_y = a_z + b_1^* Z_1 + b_2^* Z_2$$

where: Z indicates that all scores have been standardized to the normal curve

The standardized regression equation can be further simplified by dropping the term for the Y intercept because this term will always be zero when scores have been standardized. This value is the point where the regression line crosses the Y axis and is equal to the mean of Y when all independent variables equal 0. This relationship can be seen by substituting 0 for all independent variables in Formula 14.6:

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$
$$a = \bar{Y} - b_1(0) - b_2(0)$$
$$a = \bar{Y}$$

Because the mean of any standardized distribution of scores is zero, the mean of the standardized Y scores will be zero and the Y intercept will also be zero $(a = \bar{Y} = 0)$. Thus, Formula 14.9 simplifies to:

FORMULA 14.10

$$Z_y = b_1^* Z_1 + b_2^* Z_2$$

The standardized regression equation, with beta-weights noted, would be

$$Z_y = (0.46)Z_1 + (-0.09)Z_2$$

and it is immediately obvious that the first independent variable has a much stronger direct effect on Y than does the second independent variable.

As we have seen, multiple regression analysis permits the researcher to summarize the linear relationship among two or more independents and a dependent variable. The unstandardized regression equation (Formula 14.2) permits values of Y to be predicted from the independent variables in the original units of the variables. The standardized regression equation (Formula 14.10) allows the researcher to easily assess the relative importance of the various independent variables by comparing the beta-weights. (*For practice in computing and interpreting beta-weights, see any of the problems at the end of this chapter. It is probably a good idea to start with problem 14.1 because it has the smallest data set and the least complex computations.*)

STATISTICS
IN EVERYDAY LIFE

The research literature on the crime of rape generally finds that the underlying motivation is about power and dominance. However, an analysis by economist Todd Kendall suggests that sexual desire may be the important force behind these attacks. He finds a correlation between the availability of pornography on the Internet and the rate of rape. In fact, his regression analysis "… implies that an increase in home internet access of 10 percentage points is associated with a 7.3% decline in rape" (p. 23) net of the effect of other variables in the regression equation (poverty, unemployment, alcohol consumption, and so forth). These results suggest that pornography is a substitute for rape—a finding that is in sharp contrast to most previous literature.

Source: Kendall, Todd. 2006. *Pornography, Rape, and the Internet*. Available from http://www.toddkendall.net/internetcrime.pdf.

Applying Statistics 14.1 Multiple Regression and Correlation

Five recently divorced men have been asked to rate subjectively the success of their adjustment to single life on a scale ranging from 5 (very successful adjustment) to 1 (very poor adjustment). Is adjustment related to the length of time married? Is adjustment related to socioeconomic status as measured by yearly income?

Case	Adjustment (Y)	Years Married (X_1)	Income (dollars) (X_2)
A	5	5	30,000
B	4	7	45,000
C	4	10	25,000
D	3	2	27,000
E	1	15	17,000
	$\bar{Y} = 3.4$	$\bar{X}_1 = 7.8$	$\bar{X}_2 = 28,800$
	$s = 1.4$	$s = 4.5$	$s = 9,173.88$

(*continued next page*)

Applying Statistics 14.1 *(continued)*

The zero-order correlations among these three variables are:

	Adjustment (Y)	Years Married (X_1)	Income (X_2)
Adjustment (Y)	1.00	–0.62	0.62
Years married (X_1)		1.00	−0.49
Income (X_2)			1.00

These results suggest a strong but opposite relationship between each independent variable and adjustment. Adjustment decreases as years married increases, and it increases as income increases.

To find the multiple regression equation, we must find the partial slopes.

For years married (X_1):

$$b_1 = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}\right)$$

$$b_1 = \left(\frac{1.4}{4.5}\right)\left(\frac{(-0.62) - (0.62)(-0.49)}{1 - (-0.49)^2}\right)$$

$$b_1 = (0.31)\left(\frac{(-0.62) - (-0.30)}{1 - 0.24}\right)$$

$$b_1 = (0.31)\left(\frac{-0.32}{0.76}\right)$$

$$b_1 = (0.31)(-0.42)$$

$$b_1 = -0.13$$

For income (X_2):

$$b_2 = \left(\frac{s_y}{s_2}\right)\left(\frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2}\right)$$

$$b_2 = \left(\frac{1.4}{9{,}173.88}\right)\left(\frac{(0.62) - (-0.62)(-0.49)}{1 - (-0.49)^2}\right)$$

$$b_2 = (0.00015)\left(\frac{0.62 - 0.30}{1 - 0.24}\right)$$

$$b_2 = (0.00015)\left(\frac{0.32}{0.76}\right)$$

$$b_2 = (0.00015)(0.42)$$

$$b_2 = 0.000063$$

The Y intercept would be:

$$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2$$

$$a = 3.4 - (-0.13)(7.8) - (0.000063)(28{,}800)$$

$$a = 3.4 - (-1.01) - (1.81)$$

$$a = 3.4 + 1.01 - 1.81$$

$$a = 2.60$$

The multiple regression equation is:

$$Y = a + b_1X_1 + b_2X_2$$

$$Y = 2.60 + (-0.13)X_1 + (0.000063)X_2$$

What adjustment score could we predict for a male who had been married 30 years ($X_1 = 30$) and had an income of \$50,000 ($X_2 = 50{,}000$)?

$$Y' = 2.60 + (-0.13)(30) + (0.000063)(50{,}000)$$

$$Y' = 2.60 + (-3.90) + (3.15)$$

$$Y' = 1.85$$

To assess which of the two independents has the stronger effect on adjustment, the standardized partial slopes must be computed.

For years married (X_1):

$$b_1^* = b_1\left(\frac{s_1}{s_y}\right)$$

$$b_1^* = (-0.13)\left(\frac{4.5}{1.4}\right)$$

$$b_1^* = -0.42$$

For income (X_2):

$$b_2^* = b_2\left(\frac{s_2}{s_y}\right)$$

$$b_2^* = (0.000063)\left(\frac{9173.88}{1.4}\right)$$

$$b_2^* = 0.41$$

The standardized regression equation is:

$$Z_y = b_1^*Z_1 + b_2^*Z_2$$

$$Z_y = (-0.42)Z_1 + (0.41)Z_2$$

and the independent variables have nearly equal but opposite effects on adjustment. To assess the combined effects of the two independents on adjustment, the coefficient of multiple determination must be computed:

$$R^2 = r_{y1}^2 + r_{y2.1}^2(1 - r_{y1}^2)$$

$$R^2 = (-0.62)^2 + (0.46)^2(1 - (-0.62)^2)$$

(*continued next page*)

Applying Statistics 14.1 *(continued)*

$$R^2 = 0.38 + (0.21)(1 - 0.38)$$
$$R^2 = 0.38 + (0.21)(0.62)$$
$$R^2 = 0.38 + 0.13$$
$$R^2 = 0.51$$

The first independent variable—years married—explains 38% of the variation in adjustment by itself. To this quantity, income explains an additional 13% of the variation in adjustment. Taken together, the two independent variables explain a total of 51% of the variation in adjustment.

Applying Statistics 14.2 R^2 and Beta-Weights

The following table presents information on three variables for a small sample of 10 nations. The dependent variable is the percentage of respondents who said they are "very happy" on a survey administered to random samples from each nation. The independent variables measure health and physical well-being (life expectancy, or the number of years the average citizen can expect to live at birth) and income inequality (the amount of total income that goes to the richest 20% of the population). Our expectation is that happiness will have a positive correlation with life expectancy (the greater the health, the happier the population) and a negative relationship with inequality (the greater the inequality, the greater the discontent and the lower the level of happiness). In this analysis, we focus on R^2 and the beta-weights only.

Nation	Percent "Very Happy" (Y)	Life Expectancy (X_1)	Income Inequality (X_2)
Brazil	34.1	73	56.99
Canada	46.4	81	32.56
Germany	19.9	80	29.31
Ghana	50.1	59	40.79
India	29.0	64	36.80
Japan	29.2	83	24.85
Mexico	58.5	75	46.05
Russia	21.2	68	39.93
Rwanda	11.9	48	46.79
United States	34.4	78	40.81
Mean =	33.47	70.90	39.48
Standard deviation =	13.82	10.61	8.83

A correlation matrix presents the zero-order correlations:

	Happiness	Life Expectancy	Inequality
Happiness	1.00	0.29	0.13
Life expectancy		1.00	–0.46
Inequality			1.00

Consistent with our expectations, there is a weak-to-moderate positive relationship between life expectancy and happiness. Unexpectedly, however, the relationship between inequality and happiness is positive, although weak in strength. The relationship between the two independent variables is moderate and negative, indicating that nations with more income inequality have lower life expectancy.

(*continued next page*)

Applying Statistics 14.2 *(continued)*

The combined effect of life expectancy and inequality on happiness is found by computing R^2:

$$R^2 = r^2y_1 + r^2y_{2.1}(1 - r^2y_1)$$
$$R^2 = (0.29)^2 + (0.31)^2(1 - 0.29^2)$$
$$R^2 = 0.08 + (0.10)(\ 0.92)$$
$$R^2 = 0.08 + 0.09$$
$$R^2 = 0.17$$

By itself, life expectancy explains 8% of the variance in happiness. To this, income inequality adds another 9%, for a total of 17%. This leaves about 83% of the variance unexplained—a sizable proportion but not unusually large in social science research.

We must calculate the beta-weights to assess the separate effects of the two independent variables. The unstandardized partial slopes needed for this computation are 0.58 for X_1 (life expectancy) and 0.53 for X_2 (income inequality).

For the first independent variable (life expectancy):

$$b_1^* = b1\left(\frac{s_1}{sy}\right)$$
$$b_1^* = (0.58)\left(\frac{10.61}{13.82}\right)$$
$$b_1^* = (0.58)(0.77)$$
$$b_1^* = 0.45$$

For the second independent variable (income inequality):

$$b_2^* = b2\left(\frac{s_2}{sy}\right)$$
$$b_2^* = (0.53)\left(\frac{8.83}{13.82}\right)$$
$$b_2^* = (0.53)(0.64)$$
$$b_2^* = 0.34$$

Recall that the beta-weights show the effect of each independent variable on the dependent variable while controlling for the other independent variables in the equation. In this case, life expectancy has the stronger effect and the relationship is positive. The effect of income inequality is also positive.

In summary, for these nations, level of happiness has a moderate positive relationship with life expectancy and a weaker positive relationship with income inequality. Taken together, the independent variables explain 17% of the variation in happiness.

Multiple Correlation

We use the multiple regression equations to disentangle the separate direct effects of each independent variable on the dependent variable. Using **multiple correlation** techniques, we can also ascertain the combined effects of all independent variables on the dependent variable. We do so by computing the **multiple correlation coefficient (*R*)** and the **coefficient of multiple determination (*R*)**^{2}. The value of the latter statistic represents the proportion of the variance in Y that is explained by all the independent variables combined.

We have seen that number of children (X_1) explains 25% of the variance in Y ($r^2_{y1} = (.50)^2 = .25 \times 100 = 25\%$) by itself and that husband's education explains 9% of the variance in Y ($r^2_{y2} = (-.30)^2 = .09 \times 100 = 9\%$). The two zero-order correlations cannot simply be added together to ascertain their combined effect on Y because the two independent variables are also correlated with each other; therefore, they will "overlap" in their effects on Y and explain some of the same variance. This overlap is eliminated in Formula 14.11:

FORMULA 14.11

$$R^2 = r^2_{y1} + r^2_{y2.1}(1 - r^2_{y1})$$

where: R^2 = the coefficient of multiple determination
r^2_{y1} = the zero-order correlation between Y and X_1, the quantity squared