and widely used statistical tools for multivariate analysis are introduced in this chapter. We cover techniques that are used to analyze causal relationships and to make predictions—both crucial endeavors in any science.

These techniques are based on Pearson's *r* (see Chapter 13) and are most appropriately used with high-quality, precisely measured interval-ratio variables. As we have noted on many occasions, such data are relatively rare in social science research, and the techniques presented in this chapter are commonly used on variables measured at the ordinal level and with nominal-level variables in the form of dummy variables (see Chapter 13).

We first consider partial correlation analysis—a technique that allows us to examine bivariate relationships while controlling for a third variable. The second technique—multiple regression and correlation—allows researchers to assess the effects—separately and in combination—of more than one independent variable on the dependent variable.

Throughout this chapter, we focus on research situations involving three variables. This is the simplest application, but extensions to situations involving four or more variables are relatively straightforward. However, computations become very complex when dealing with more than two independent variables, and you should use computerized statistical packages (such as SPSS) for these situations.

Partial Correlation

In Chapter 13, we used Pearson's *r* to measure the strength and direction of bivariate relationships. To provide an example, we looked at the relationship between husband's contribution to housework (the dependent, or *Y,* variable) and number of children (the independent, or *X*, variable) for a sample of 12 families. We found a positive relationship of moderate strength $(r = 0.50)$ and concluded that husband's contribution to housework tends to increase as the number of children increases.

You might wonder, as researchers commonly do, if this relationship holds true for *all* types of families. For example, might husbands in strongly religious families respond differently than those in less religious families? Would politically conservative husbands behave differently than husbands who are politically liberal? How about more educated husbands? Would they respond differently than less educated husbands? We can address these kinds of issues by means of a technique called **partial correlation**, in which we observe how the bivariate relationship changes when a third variable, such as religiosity, education, or ethnicity, is introduced. Third variables are often referred to as *Z* variables or **control variables**.

Partial correlation proceeds by first computing Pearson's *r* for the bivariate (or **zero-order**) relationship and then computing the partial (or **first-order**) correlation coefficient. If the partial correlation coefficient differs from the zero-order correlation coefficient, we conclude that the third variable has an effect on the bivariate relationship. For example, if well-educated husbands respond differently to an additional child than less-well-educated husbands, the partial correlation coefficient will differ in strength (and perhaps in direction) from the bivariate correlation coefficient.

Before considering matters of computation, we will consider the possible relationships between the partial and bivariate correlation coefficients and what they might mean. There are three possible patterns, and we will consider each in turn.

Types of Relationships

Direct Relationship. One possible outcome is that the partial correlation coefficient is essentially the same value as the bivariate coefficient. For example, imagine that after we controlled for husband's education, we found a partial correlation coefficient of +0.49, compared to the zero-order Pearson's *r* of +0.50. This would mean that the third variable (husband's education) has no effect on the relationship between number of children and husband's hours of housework. In other words, regardless of their education, husbands respond in a similar way to additional children. This outcome is consistent with the conclusion that there is a direct or causal relationship (see Figure 14.1) between *X* and *Y* and that the third variable (*Z*) is irrelevant and should be discarded from further consideration, although the researcher would probably run additional tests with other likely control variables (e.g., the researcher might control for the religion or ethnicity of the family).

FIGURE 14.1 A Direct Relationship Between *X* and *Y*

Spurious and Intervening Relationships. A second possible outcome occurs when the partial correlation coefficient is much weaker than the bivariate correlation—perhaps even dropping to zero. This outcome is consistent with two different relationships. The first is called a *spurious relationship*: The control variable (*Z*) is a cause of the independent (*X*) and the dependent (*Y*) variable (see Figure 14.2). This outcome would mean that *X* and *Y* are not actually related. They appear to be related only because both are dependent on a common cause (*Z*). Once *Z* is taken into account, the apparent relationship between *X* and *Y* disappears.

What would a spurious relationship look like? Imagine that we controlled for the political ideology of parents and found that the partial correlation coefficient was much weaker than the bivariate Pearson's *r*. This might indicate that the number of children does not actually change the husband's contribution to housework (that is, the relationship between *X* and *Y* is not direct). Rather, political ideology is the mutual cause of both of the other variables: More conservative families are more likely to follow traditional gender role patterns (in which husbands contribute less to housework) *and* have more children.

This pattern (partial correlation much weaker than the bivariate correlation) is also consistent with an **intervening** relationship between the variables (see Figure 14.3). In this situation, *X* and *Y* are not linked directly but are connected through the control variable. Again, once *Z* is controlled, the apparent relationship between *X* and *Y* disappears.

How can we tell the difference between spurious and intervening relationships? This distinction cannot be made on statistical grounds: Spurious and intervening relationships look exactly the same in terms of statistics. The researcher may be able to distinguish between these two relationships in terms of

FIGURE 14.2 A Spurious Relationship Between *X* and *Y*

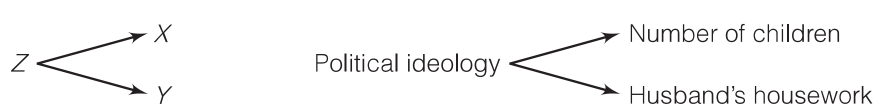


FIGURE 14.3 An Intervening Relationship Between *X* and *Y*

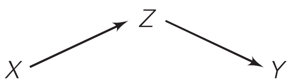
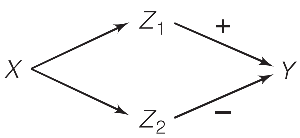


Figure 14.4 An Interactive Relationship Between *X*, *Y*, and *Z*



the time order of the variables (i.e., which came first) or on theoretical grounds, but not on statistical grounds.

Interaction. A third possible relationship between variables should be mentioned, even though it *cannot* be detected by partial correlation analysis. This relationship—called *interaction*—occurs when the relationship between *X* and *Y* changes markedly under the various values of *Z*. For example, if we controlled for social class and found that husbands in middle-class families increased their contribution to housework as the number of children increased while husbands in working-class families did just the reverse, we would conclude that there was interaction between these three variables. In other words, there would be a positive relationship between *X* and *Y* for one category of *Z* and a negative relationship for the other category, as illustrated in Figure 14.4.

Computing and Interpreting the Partial Correlation Coefficient

The logic and computation of the partial correlation coefficient requires some new concepts and terminology. We will introduce the terminology first and then move on to the formula.

Terminology and Formula. Partial correlation requires that we deal with more than one bivariate relationship, and we need to differentiate between them with subscripts. Thus, the symbol r_{yx} will refer to the correlation coefficient between variable *Y* and variable *X*; r_{yz} will refer to the correlation coefficient between *Y* and *Z*; and r_{xz} will refer to the correlation coefficient between *X* and *Z*. Recall that correlation coefficients calculated for bivariate relationships are often referred to as zero-order correlations.

Partial correlation coefficients, or first-order partials, are symbolized as $r_{yx.z}$. The variable to the right of the dot is the control variable. Thus, $r_{yx.z}$ refers to the partial correlation coefficient that measures the relationship between variables *X* and *Y* while controlling for variable *Z*. The formula for the first-order partial is:

FORMULA 14.1

$$r_{yx.z} = \frac{r_{yx} - (r_{yz})(r_{xz})}{\sqrt{1 - r_{yz}^2}\sqrt{1 - r_{xz}^2}}$$

Note that you must first calculate the zero-order coefficients between all possible pairs of variables (variables *X* and *Y*, *X* and *Z*, and *Y* and *X*) before solving this formula.

Computation. To illustrate the computation of a first-order partial, we will return to the relationship between number of children (*X*) and husband's contribution to housework (*Y*) for 12 dual-career families. The zero-order *r* between these two variables ($r_{yx} = 0.50$) indicate a moderate positive relationship (as number of children increased, husbands tended to contribute more to housework). Suppose the researcher wished to investigate the possible effects of husband's education on the bivariate relationship. The original data (from Table 13.1) and the scores of the 12 families on the new variable are presented in Table 14.1.

The zero-order correlations, as presented in the correlation matrix in Table 14.2, indicate that the husband's contribution to housework is positively related to number of children ($r_{yx} = 0.50$), that better-educated husbands tend to do less housework ($r_{yz} = -0.30$), and that families with better-educated husbands have fewer children ($r_{xz} = -0.47$).

TABLE 14.1 Scores on Three Variables for 12 Dual-Wage-Earner Families and Zero-Order Correlations

Family	Husband's Housework (*Y*)	Number of Children (*X*)	Husband's Years of Education (*Z*)
A	1	1	12
B	2	1	14
C	3	1	16
D	5	1	16
E	3	2	18
F	1	2	16
G	5	3	12
H	0	3	12
I	6	4	10
J	3	4	12
K	7	5	10
L	4	5	16

TABLE 14.2 Zero-Order Correlations

	Husband's Housework (*Y*)	Number of Children (*X*)	Husband's Years of Education (*Z*)
Husband's Housework (*Y*)	1.00	0.50	−0.30
Number of Children (*X*)		1.00	−0.47
Husband's Years of Education (*Z*)			1.00

Is the relationship between husband's housework (*X*) and number of children (*Y*) affected by husband's years of education? Substituting the zero-order correlations into Formula 14.1, we would have:

$$r_{yx.z} = \frac{r_{yx} - (r_{yz})(r_{xz})}{\sqrt{1 - r_{yz}^2}\sqrt{1 - r_{xz}^2}}$$

$$r_{yx.z} = \frac{(0.50) - (-0.30)(-0.47)}{\sqrt{1 - (-0.30)^2}\sqrt{1 - (0.47)^2}}$$

$$r_{yx.z} = \frac{(0.50) - (0.14)}{\sqrt{1 - 0.09}\sqrt{1 - 0.22}}$$

$$r_{yx.z} = \frac{0.36}{\sqrt{0.91}\sqrt{0.78}}$$

$$r_{yx.z} = \frac{0.36}{(0.95)(0.88)}$$

$$r_{yx.z} = \frac{0.36}{0.84}$$

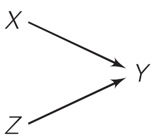
$$r_{yx.z} = 0.43$$

Interpretation. The first-order partial ($r_{yx.z} = 0.43$), which measures the strength of the relationship between husband's housework (*Y*) and number of children (*X*) while controlling for husband's education (*Z*), is lower in value than the zero-order coefficient ($r_{yx} = 0.50$), but the difference in the two values is not great. This result suggests a direct relationship between variables X and Y. That is, when controlling for husband's education, the statistical relationship between husband's housework and number of children is essentially unchanged. Regardless of education, husband's hours of housework increase with the number of children.

Our next step in statistical analysis would probably be to select another control variable. The more the bivariate relationship retains its strength across a series of controls for third variables (Zs), the stronger the evidence for a direct relationship between X and Y.

In closing, I should mention an additional possible outcome, in which the partial correlation coefficient is greater in value than the zero-order coefficient ($r_{yx.z} > r_{yx}$). This outcome would be consistent with a causal model in which the variable taken as independent and the control variable each had a separate effect on the dependent variable and were uncorrelated with each other. This relationship is depicted in Figure 14.5. The absence of an arrow between X and Z indicates that they have no mutual relationship.

Figure 14.5 A Possible Causal relationship Among three Variables



This pattern means that X and Z should be treated as independent variables, and the next step in the statistical analysis would probably involve multiple correlation and regression—techniques that will be presented in the remainder of this chapter. (*For practice in computing and interpreting partial correlation coefficients, see problems 14.1 to 14.3.*)

ONE STEP AT A TIME — Computing and Interpreting the Partial Correlation Coefficient

To begin, compute Pearson's r for all pairs of variables. Be clear about which variable is independent (X), which is dependent (Y), and which is the control (Z).

Step ***Operation***

To find the partial correlation coefficient, solve Formula 14.1:

1. Multiply r_{yz} by r_{xz}.
2. Subtract the value you found in step 1 from r_{yx}. *This value is the numerator of Formula 14.1.*
3. Square the value of r_{YZ}.
4. Subtract the quantity you found in step 3 from 1.
5. Take the square root of the quantity you found in step 4.
6. Square the value of r_{XZ}.
7. Subtract the quantity you found in step 6 from 1.
8. Take the square root of the quantity you found in step 7.
9. Multiply the quantity you found in step 8 by the value you found in step 5. *This value is the denominator of Formula 14.1.*
10. Divide the quantity you found in step 2 by the quantity you found in step 9. *This is the partial correlation coefficient.*

To interpret the partial correlation coefficient, compare its value to the zero-order correlation:

Choose the description below that comes closest to matching the relationship between the two values:

1. The partial correlation coefficient is roughly the same value as the zero-order or bivariate correlation. A good rule of thumb for "roughly the same" is a difference of less than 0.10.

(*continued next page*)