

# Dyadic Data Analysis

David A. Kenny  
Deborah A. Kashy  
William L. Cook

FOREWORD by Jeffry A. Simpson



THE GUILFORD PRESS  
New York    London

## Basic Definitions and Overview

The dyad is arguably the fundamental unit of interpersonal interaction and interpersonal relations. Although we commonly think of dating and marital partners when we consider dyadic relationships, friendships are also often experienced as dyadic phenomena, even though they may be nested within larger friendship groups. Even family relations have a strong dyadic component in that we have different relationships with our mothers, fathers, and each of our siblings. Beyond the domain of close relationships, everyday interactions with acquaintances and strangers often occur in pairs (Bakeman & Beck, 1974; DePaulo & Kashy, 1998; James, 1953; Kashy, 1992). This book describes the methodological and data-analytic approaches useful in the study of dyads. The methods that we present in this book can be applied in a variety of contexts that involve two individuals, from relationships between a doctor and a patient to interactions between two people waiting for an experiment to begin to dating couples, pen pals, best friends, siblings, and coworkers. Our focus is on quantitative, as opposed to qualitative, methods.

Many of the phenomena studied by social and behavioral scientists are interpersonal by definition, and as a result, observations do not refer to a single person but rather to multiple persons embedded within a social context. For instance, Harry's response when he is asked how much he likes Sally does not simply reflect something about Harry. Yet because the check mark on the questionnaire is made by Harry, researchers all too often make the fundamental attribution error (Ross, 1977) and treat the measurement as if only Harry caused it. The error of thinking that a dyadic

measure refers to only one of the interaction partners has been called *pseudo-unilaterality* (Duncan, Kanki, Mokros, & Fiske, 1984). Almost certainly, the liking that Harry feels for Sally is driven in part by characteristics of Sally herself, such as how friendly or agreeable she is, as well as by the unique relationship that Harry and Sally have established. The measurement reflects both Harry and Sally and, therefore, is fundamentally dyadic. In general, a dyadic measurement reflects the contribution of two persons, although the function of those contributions can be quite different (Bond & Kenny, 2002).

The intrinsically dyadic nature of many of the measurements in social and behavioral science research means that they are often linked to other measurements in the study, and the strength of these links may be one of the most important research questions to be examined. Consider the following examples:

- Both members in a romantic relationship evaluate whether they are satisfied with the relationship (Feeney, 1994).
- The amount of self-disclosure made by two people interacting is measured to ascertain whether there is reciprocity (Reno & Kenny, 1992).
- Two persons are asked to describe a common target person to determine whether there is agreement in person perception (Park & Judd, 1989).
- Members of a family describe their attachment relationships with one another (Cook, 2000).

In each of these cases, the issues of stability, consistency, and correlation between related measurements are interesting phenomena worth studying in their own right. However, none of them can be addressed easily by standard methods developed for the study of individuals.

Why has social science research tended to focus on individuals? Although there are many reasons for this focus, we think that three are key. First, no doubt much of the attention given to the individual is cultural. The United States is the most individualistic country in the world (Smith & Bond, 1994), and because the United States has dominated social and behavioral research, the prevalence of research concerning individuals is hardly surprising.

A second factor that has contributed to this individualistic orientation is the reliance on standard statistical methods such as analysis of variance

(ANOVA) and multiple regression. Although these two data-analytic approaches are very useful and, as will be shown, form the basis for many of the techniques described in this book, in their standard forms they both make what is known as the *independence assumption*. The independence assumption requires that, after controlling for variation due to the independent variable, the data from each individual in a study be unrelated to the data from every other individual in the study. As discussed later in this chapter, dyadic data typically violate this assumption.

The third reason is that psychologists have dominated research in the social and behavioral sciences. The discipline of psychology emphasizes the individual before higher levels of analysis (Bond & Kenny, 2002). So it is hardly surprising that most methods of analysis focus on the individual.

In spite of the individualistic focus of social and behavioral science research, many theoretical concepts intrinsically involve two persons (e.g., love, conflict, person perception, helping, aggression, attachment, relational competence, communication, influence). The need for a book detailing dyadic data analysis is highlighted by the fact that most of these interpersonal concepts have been studied by examining individuals in isolation. Before we can have a genuinely interpersonal social science, our theories, research methods, and data analyses must take into account the truly interpersonal nature of the phenomena under study. One of the major goals of this book is to provide social scientists with methods that focus on relationships and not individuals.

In this chapter, we define the fundamental concepts for dyadic data analysis. We begin by defining the most essential concept in relationship research: nonindependence. A series of other basic concepts are also defined, including distinguishability, types of dyadic variables (between dyads, within dyads, and mixed), and levels of measurement for dyadic variables. In addition, a typology of dyadic designs is provided. We also offer advice concerning the organization of dyadic data files. We then describe a database that includes 75 studies of relationships from five major journals. This database is used throughout the book, and a catalog of the types of relationships examined in these studies is provided. Finally, we give the reader an overview of the remainder of the book.

Although much of what we discuss in this first chapter is rather elementary, it is essential that the reader thoroughly understand the terminology presented in this chapter, because those terms are referred to repeatedly throughout the book. Thus we encourage all to read the remainder of this chapter, even those who are quite statistically sophisticated.

## NONINDEPENDENCE

Perhaps the most fundamental concept in dyadic data analysis is that of nonindependence. Two members of a dyad are not simply two independent individuals. Rather, they share something in common that we refer to as *nonindependence*. The focus of this entire book is, in essence, the study of nonindependence.

Although we postpone our statistical definition of the concept of nonindependence until the next chapter, it is useful to develop a conceptual definition here. A formal conceptual definition of dyadic nonindependence is: If the two scores from the two members of the dyad are nonindependent, then those two scores are more similar to (or different from) one another than are two scores from two people who are not members of the same dyad. The heightened similarity (or dissimilarity) of scores from dyads is the critical issue that is central to this book. Our discussion tends to focus on nonindependence that results from close interpersonal relationships such as friendships, married or dating couples, and roommates. However, similar issues may arise when the two individuals are initially strangers who have just met in the laboratory or on the Internet. Nonindependence can even occur when two people never actually interact but share a common experience; for example, two patients of the same physician.

The preceding definition presumes that the data are structured in what we define as the *standard dyadic design*: Each person is linked to one, and only one, other person. In Chapters 8, 9, 10, and 11 we investigate other, more complex patterns of nonindependence. In addition, nonindependence can occur as a result of factors other than relationships. For instance, measurements from the same person may be nonindependent, something that we discuss in Chapters 12, 13, and 14.

Nonindependence, or linked scores, can occur in several ways, and it is helpful to distinguish among voluntary linkage, kinship linkage, experimental linkage, and yoked linkage. *Voluntary linkage* is the link between friends or between members of dating couples. We normally think of these persons as having some sort of bond that develops over time. *Kinship linkage* is a linkage that occurs between family members, such as siblings, cousins, or parents and children. *Experimental linkage* is a relationship that is created in a laboratory, as when two persons are asked to get to know each other. Finally, in *yoked linkage*, the two individuals never interact at

all and are not even aware of each other, but they are both exposed to the same environmental stimuli. Very often linkages are combinations of two or more types of linkages: Married couples are linked both voluntarily and by kinship.

Kenny (1996b) and Kenny and Judd (1986) describe four sources that may generate nonindependence in dyads. The first source is simply a *compositional effect*: The two dyad members may have already been similar even before they were paired together. Compositional effects are likely to occur any time dyad members are paired together in a nonrandom way. For example, compositional effects are to be expected with dating and married couples because, even before they meet, members of such couples typically are similar to one another on a wide range of variables, including education level, age, socioeconomic status, religion, and so on (Epstein & Guttman, 1984). This similarity of married couples is sometimes referred to as *assortative mating*. Nonrandom pairing is typically an issue in naturally occurring dyads. For example, married couples likely have similar political attitudes because, in part, they have similar educational backgrounds. Moreover, similarities in political attitudes may have been a factor that created the dyad.

Once dyad members have been paired together, even if the pairing is random so that compositional effects are unlikely, there are three processes that may produce nonindependence between the two individuals. A *partner effect* occurs when a characteristic or behavior of one person affects his or her partner's outcomes. The amount of housework that one roommate does may affect the other roommate's level of satisfaction with his or her living arrangements. Similarly, how much a woman trusts her dating partner may affect the partner's level of commitment to the relationship. *Mutual influence* occurs when both persons' outcomes directly affect one another. Thus mutual influence involves a process of feedback. In a study of initial interactions between strangers, mutual influence might occur for a variable such as liking, so that the more a person likes his or her interaction partner, the more the partner likes that person in return. The third process that may produce nonindependence is *common fate*. Common fate effects occur when both dyad members are exposed to the same causal factors. Consider again the example of roommates living in an apartment complex. If the complex was poorly maintained and the environment unpleasant, then the two roommates' satisfaction might be similar because the unpleasant environment affects both of them.



## BASIC DEFINITIONS

### Distinguishability<sup>1</sup>

One important question in dyadic research and data analysis is whether or not the two dyad members can be distinguished from one another by some variable. Table 1.1 presents examples of dyads in which members are distinguishable and indistinguishable. In heterosexual dating relationships, dyad members are distinguishable because of their gender: Each couple has one man and one woman. In sibling dyads, the two siblings can be distinguished by birth order. In both of these examples, a systematic ordering of the scores from the two dyad members can be developed based on the variable that distinguishes them. However, there are many instances in which there is no such natural distinction. Same-sex friendship pairs, homosexual romantic partners, and identical twins are all examples of dyads in which the members are typically indistinguishable. If dyad members are indistinguishable, then there is no systematic or meaningful way to order the two scores. Thus, by distinguishability, we mean the following: Dyad members are considered distinguishable if there is a meaningful factor that can be used to order the two persons.

Distinguishability is critical to a discussion of quantitative methods for relationship data because the data-analytic techniques appropriate for distinguishable dyads may not be appropriate for indistinguishable dyads. We shall see that the statistical analysis of data from dyad members who are distinguishable is relatively easy. For this reason, researchers sometimes create a variable that can be used to distinguish dyad members. If

such a variable is theoretically and empirically meaningful, this approach is not problematic. However, if the distinguishing variable is not meaningful (e.g., the person who is in the front of the data storage folder is assigned to be "X" and the person who is in the back of the folder is "Y"), this practice engenders an arbitrary component in the data, and it should be avoided.

Technically, the decision of whether or not the dyad members are distinguishable is both empirical and theoretical. Notice that the definition refers to a "meaningful factor" distinguishing the two persons. Sometimes a factor is designated as theoretically "meaningful" (e.g., parent and child). Other times distinguishability is an empirical issue, and the defining question is whether there are differences in the data (e.g., if there are no mean or variance differences between the two members) for the two "types" of partners (Gonzalez & Griffin, 1999). We discuss empirical tests of distinguishability in detail in Chapters 6 and 7.

### Between-Dyads, Within-Dyads, and Mixed Variables

An important distinction that is often made in research is to refer to variables as either *independent* or *dependent* variables. An independent variable is usually assumed to cause a dependent variable. In this book, we use the terms *independent variable* and *outcome variable*. However, we do not necessarily assume that the independent variable is a variable that is manipulated by an experimenter. In some circumstances, the relationship between the independent variable and the outcome variable may be predictive rather than causal.

The nature of the independent variable plays an important role in determining the appropriate data-analytic approach for dyadic data. In this section, we introduce the concept of between-dyads, within-dyads, and mixed variables (Kashy & Kenny, 2000; Kenny, 1988a). Although this distinction can be made for any variable, including outcome variables, it is most important for independent variables.

### Between-Dyads Variables

Scores on a between-dyads variable differ from dyad to dyad, but not within a dyad, and thus both members have identical scores on the variable. For example, in a study of the effects of stress on romantic relationship satisfaction, couples might be randomly assigned to a high-stress condition in which they are asked to discuss a difficult problem in their relationship, or they could be assigned to a low-stress condition in which they are asked to

TABLE 1.1. Illustrations of Distinguishable and Indistinguishable Members

Dyads with distinguishable members	Dyads with indistinguishable members
Husband and wife	Gay couple
Boss and employee	Coworkers
Older and younger siblings	Twins
Person and his or her best friend	Best friends (mutually chosen)
Winner and loser	Opponents
Parent and child	Roommates
Waiter and customer	Pen pals
Teacher and student	Study partners
Sadist and masochist	Business partners
First and second author	Colleagues
Pet owner and pet	Acquaintances

discuss a current event. For this example, the level of stress would be a between-dyads variable, because both dyad members are at the same level of induced stress: Some dyads would be in the high-stress condition, and others would be in the low-stress condition. Other examples of such variables are:

- Gender, in a study of same-sex roommates that includes both men and women.
- Length of a couple's marriage.
- Opinion, when the members of the dyad are asked to come to consensus on an issue.

Some variables that we might think are between-dyads variables may not necessarily be so. For instance, in a study of married couples, the question of whether a couple engaged in premarital sexual intercourse may seem to be a between-dyads variable. However, Liu and Detels (1999) found in one survey that 5% of couples disagreed. In some cases, the scores on an independent variable from the two members can be combined to create a single between-dyads score. For example, if members of dating couples disagree on when they first met, an average, or the earlier of their two responses, could be used. However, we discourage the routine averaging of the scores of dyad members (see Chapter 7 for empirical criteria for averaging).

### *Within-Dyads Variables*

The two scores of a within-dyads variable differ between the two members within a dyad, but when averaged across the two dyad members, each dyad has an identical average score. Gender is a prototypical within-dyads variable in heterosexual couples, in that every couple is composed of both a man and a woman. A less obvious example of a within-dyads variable is the actual proportion of housework done by two roommates. With this variable, the average of the two proportions always equals .50, yet within each dyad the amount of housework varies across the two partners. Examples of other within-dyads variables are:

- Family role in a study of fathers and sons.
- Role when one person is asked to persuade another person.
- Reward allocation when each member of a dyad is rewarded separately, with the constraint that every dyad is assigned the same total amount.

Earlier we discussed the notion of distinguishability of the dyad members. Dyad members are distinguished by a within-dyads variable.

### *Mixed Variables*

The third type of variable in dyadic research is a mixed independent variable in which variation exists both within the dyads and between dyads. A mixed predictor variable is probably a new concept to most researchers. Kenny and Cook (1999) and Kenny, Kashy, and Bolger (1998) present extended discussions of mixed variables. Age is an example of a mixed independent variable in marital research because the two spouses' ages may differ from one another, and, in addition, some couples on average may be older than others. Many variables in dyadic research are mixed in nature in that the two partners' scores differ and some dyads have higher average scores than others. Additional examples of mixed variables include satisfaction and individual productivity. Most outcome variables in dyadic research are mixed. Chapter 7 presents an extended discussion of the analysis of mixed independent variables.

A variable can be a within-dyads, a between-dyads, or a mixed variable, depending on the design of the study. Consider a study of friendship: If only same-sex friends were studied, sex would be a between-dyads variable; if only opposite-sex friendships were studied, sex would be a within-dyads variable, and if both types were studied, then sex would be a mixed variable.

### **Level of Measurement**

Much of our discussion involves variables measured at either the nominal or interval levels of measurement. S. S. Stevens (1946) invented the concept of "level of measurement." The interval level of measurement is defined as measurement in which the interval between the numbers is constant, so that the difference between a score of 4 and a score of 6 is equivalent to the difference between a score of 12 and a score of 14. Generally speaking, the interval level of measurement does not assume an absolute zero (i.e., where a score of 0 implies a total absence of that variable), and so it is not possible to say that a score of 8 is twice as large as a score of 4. Most scales developed and used in social-science research are assumed to be measured on an interval scale. For instance, relationship satisfaction is usually treated as if it were an interval measurement.

Stevens (1946) defined the nominal level of measurement as measurement in which the numbers refer to discrete categories and are meant only to differentiate those categories. When there are just two categories (e.g., experimental and control or male and female), the variable is called a dichotomy. As an example, Hazan and Shaver (1987) defined adult attachment as a nominal variable: One was either secure, avoidant, or anxious-ambivalent. Griffin and Bartholomew (1994) expanded on this definition to create four categories: secure, preoccupied, dismissing, and fearful. They also presented two variables (model of self and model of other) that are measured at the interval level of measurement.<sup>2</sup>

Throughout this book, we generally assume that outcome variables are measured at the interval level. However, in several chapters (see especially Chapters 11 and 14) we focus on methods that can be used when the outcome variable is measured at the nominal level. Chapters 2, 3, and 6 also discuss the analysis of nominal variables. We usually consider analyses appropriate for both nominal and interval independent variables.

### Idiographic and Nomothetic Analyses

Two key issues in dyadic analyses are the unit of the analysis and the unit of generalization. Most often, the dyad is (or should be!) the unit of analysis, and the analysis is called *nomothetic*. In nomothetic analyses, research is conducted across many dyads, and the focus is on establishing general laws of behavior that apply to all dyads of a similar nature. Questions such as whether mothers are more responsive to their children than fathers can be approached from a nomothetic perspective by measuring mother and father responsiveness for many families and then testing for mean differences between mothers and fathers.

*Idiographic* approaches are encountered less frequently: An idiographic analysis is conducted on each dyad separately, and differences between dyads are examined. Thus, in idiographic analyses, an analysis is conducted for each dyad, and the unit of analysis might be time points in a longitudinal study or measures in a study of personality similarity between dyad members (see Chapters 12, 13, and 14). Using the parental responsiveness example, an idiographic approach to this question might involve measuring mother and father responsiveness every time they interact with their children over some period of time and computing mean differences in responsiveness for each family.

### Dyadic Designs

In this book, we detail the statistical analysis of three different types of designs described by Kenny and Winquist (2001). They are the *standard dyadic design*, the *Social Relations Model (SRM) design*, and the *one-with-many design*. The basic structure of these designs is illustrated in Table 1.2. In the table, persons are designated by uppercase letters, such as A, B, and D. It might help to think of person A as Alice, B as Bob, C as Cindy, and D as David. Note that *actor* refers to the person who generated the data point, and *partner* refers to the other member of the dyad. Thus Alice and Bob might be a dyad, and when Alice rates or interacts with Bob, an *x* is placed in the A-row, B-column. On the other hand, when Bob rates or interacts with Alice, an *x*' is placed in the B-row, A-column. So in this table the *x* score refers to the outcome score for one member of a dyad, and the *x*' score refers to the outcome score for the other member of the dyad. In some dyadic designs, only one member of the dyad is measured, and the design is said to be *one-sided*. A one-sided design would occur if only the *x* scores (or only the *x*' scores) are collected. When both members are measured, both *x* and *x*' are gathered, and the design is said to be *two-sided*. We also refer to designs in which both members are measured as *reciprocal*.

The standard design is one in which each person is a member of one and only one dyad. In Table 1.2, for the standard design (the first panel of the table), A and B are members of one dyad, C and D are members of a second dyad, E and F are members of a third dyad, and G and H are the final dyad. In this design, there are *n* dyads and *2n* individuals. When the design is reciprocal, there are *2n* observations per variable (both the *x* and *x*' observations in Table 1.2 are obtained), and when only one of the two persons is measured, there are only *n* observations (either the *x* or the *x*' observations are obtained). Generally, in this book we assume that the standard design is reciprocal. As an example of the standard design, Acitelli (1997) measured 148 married and 90 heterosexual dating couples on satisfaction. The study consisted of 238 men and 238 women. The *x* scores might then represent how satisfied the 238 men were, and the *x*' scores, how satisfied the 238 women were. Based on our survey of dyadic studies (see later in this chapter), the standard design is used in about 75% of dyadic studies. Note that in the standard design, both persons are measured, and, at least for some of the variables, both are measured on the same variables. If father and child were measured, but only the father's child-rearing philosophy and the child's respect for the father were measured, the design would not be reciprocal.

**TABLE 1.2.** Three Major Types of Designs Used to Study Dyads

Standard design

Actor	Partner							
	A	B	C	D	E	F	G	H
A		x						
B	x'							
C				x				
D			x'					
E						x		
F					x'			
G							x'	
H								x

SRM designs

Round robin

Actor	Partner			
	A	B	C	D
A		x'	x'	x'
B	x		x'	x'
C	x	x		x'
D	x	x	x	

Block

Actor	Partner							
	A	B	C	D	E	F	G	H
A					x	x	x	x
B					x	x	x	x
C					x	x	x	x
D					x	x	x	x
E	x'	x'	x'	x'				
F	x'	x'	x'	x'				
G	x'	x'	x'	x'				
H	x'	x'	x'	x'				

One-with-many design

Actor	Partner							
	A	B	C	D	E	F	G	H
A								
B	x'							
C	x'							
D	x'							
E		x	x	x				
F								
G								
H								

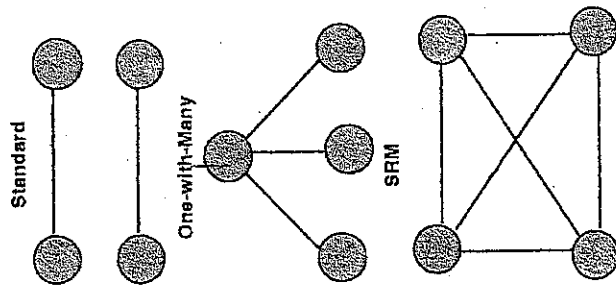
Note. Designs with both x and x' measurements are reciprocal designs, and designs with just an x or an x' measurement are nonreciprocal designs.

In an SRM design, each person is paired with multiple others, and each of these others is also paired with multiple others. As shown in Table 1.2, the prototypical SRM design is a round-robin design in which a group of persons rate or interact with each other. In the table, A and B are one dyad; A and C are also a dyad, as are A and D. Similarly, B and A are a dyad; B and C are also a dyad, as are B and D. For example, Alice may interact once with Bob, again with Cindy, and a third time with David. Bob also interacts with Alice, Cindy, and David, and so on. The round-robin design is inherently a reciprocal design, and all the observations, both x and x', are gathered. In other words, in the round-robin design, each person serves as both the actor and the partner. As an example of a round-robin SRM design, Miller and Kenny (1986) asked members of a sorority to state how willing they were to disclose information to each of the other members of their sorority.

The other major SRM design is the block design, which is also illustrated in Table 1.2. In this design, a group of persons is divided into two subgroups, and members of each subgroup rate or interact with members of the other subgroup. In Table 1.2, persons A through D form one subgroup and E through H form the other subgroup. The block design is reciprocal if both blocks (the x and the x'scores) are gathered. As an example of a block SRM design, DePaulo, Kenny, Hoover, Webb, and Oliver (1987) had one group of persons try to guess how another group of persons perceived them. In Chapters 8 and 9, other variants of SRM designs are presented.

The final design presented in Table 1.2 is the one-with-many design. In this design each person is paired with multiple others, but these others are not paired with any other persons. For example, Alice is paired with Bob, Cindy, and David. However, Bob, Cindy, and David are never paired with each other or anyone else. Like the other designs, this design can either be reciprocal (both x and x' are gathered) or not (only x or x' is gathered). However, with this design the data are typically not reciprocal. As an example of the one-with-many design, Kashy (1992) asked people to rate the physical attractiveness of each person that they had interacted with over a period of 2 weeks. A second example of the one-with-many design would be having patients rate their satisfaction with their physician (so that there are multiple patients each rating the same physician).

We illustrate the differences between the three designs in Figure 1.1. Each circle represents a person, and the line connecting two circles represents a dyadic linkage. We see for the standard design that each circle is



**FIGURE 1.1.** Diagrammatic illustrations of the three design types.

linked to just one other. In the one-with-many design, three circles are linked to one, and in the SRM design, all possible links are formed.

Although the standard design, the one-with-many design, and the SRM design account for the majority of designs used in dyadic research, other designs are possible. For instance, as discussed in Chapter 11, in studies of social networks, the pairing of persons in a group can be relatively haphazard. Other designs are also considered in the final chapter of the book.

## DATA ORGANIZATION

### Standard Design

It is very important to consider the different ways that dyadic data sets can be structured. If the data have the wrong structure, then they cannot be analyzed by the appropriate statistical technique. There are three fundamentally different ways that dyadic data from the standard design (the design in which each person is a member of one and only one dyad) can be

organized. We refer to these as *individual*, *dyad*, and *pairwise* structures. We show that the individual structure is not advisable. The other two structures have their own particular advantages and disadvantages. Because some statistical methods require a particular type of data organization and other methods require another organizational scheme, researchers should be aware that it may be necessary to create multiple data sets, each appropriate for a different statistical method.

In describing the three ways dyadic data sets can be structured, it helps to think of two types of variables. A dyad-level variable is one for which both dyad members have the same score. That is, a dyad-level variable is equivalent to what we previously termed a between-dyads variable. Marital status in a study of dating and married couples would be a dyad-level variable. An individual-level variable is one for which the dyad members each may have different scores (for some dyads, members may have the same score, but this would not be true for all dyads). Both mixed and within-dyads variables are individual-level variables. The highest educational degree obtained would be an example of an individual-level variable. Table 1.3 illustrates the three different data structures using a simple data set with three dyads (six persons) and three variables measured for each person. Variables  $X$  and  $Y$  are individual-level variables (both are mixed), and variable  $Z$  is a dyad-level variable (i.e., between dyads).

Before beginning our description of the three types of structures, we want to urge the reader to document the data carefully. Even for individual data, data management is a difficult problem, and dyadic data are much more complicated. It is essential to document the decisions that are made during the data management process. For example, researchers need to keep a careful record of how categorical variables are coded. Particular care should be given to the decisions concerning what units are excluded, how missing data are coded, and how variables are transformed.

### Individual Structure

In this case, each member of the dyad is treated as a single unit. If there were  $n$  dyads, there would be  $2n$  units in the individual file. In Table 1.3, we see that for the individual structure, there are six records of data, each one corresponding to one of the six persons in the data set. It is imperative that researchers include an identification variable (denoted *Dyad* in Table 1.3) that codes for dyad membership so that linked scores can be identified. Note that in this individual structure the dyad-level variables would have to be entered twice, once for each individual. For instance, a variable

**TABLE 1.3.** Illustration of Data Structures for a Data Set with Three Dyads, Six Persons, and Three Variables ( $X$ ,  $Y$ , and  $Z$ )

Individual		Person			Person		
Dyad	Person	$X$	$Y$	$Z$	Dyad	Person	$Z$
1	1	5	9	3	1	1	3
1	2	2	8	3	1	2	3
2	1	6	3	7	2	1	7
2	2	4	6	7	2	2	7
3	1	3	6	5	3	1	5
3	2	9	7	5	3	2	5

Dyad		Person			Person		
Dyad	Person	$X_1$	$Y_1$	$Z_1$	$X_2$	$Y_2$	$Z_2$ <sup>a</sup>
1	5	9	3	3	2	8	3
2	6	3	7	7	4	6	7
3	3	6	5	5	9	7	5

Pairwise		Person			Person		
Dyad	Person	$X_1$	$Y_1$	$Z_1$	$X_2$	$Y_2$	$Z_2$ <sup>a</sup>
1	1	5	9	3	2	8	3
1	2	2	8	3	5	9	3
2	1	6	3	7	4	6	7
2	2	4	6	7	6	3	7
3	1	3	6	5	9	7	5
3	2	9	7	5	3	6	5

<sup>a</sup>This variable is redundant with  $Z_1$  and need not be included.

$Z$  is entered for both persons 1 and 2. As we later explain, an individual structure is not useful for many dyadic analyses. Nonetheless, it is the typical way that dyadic data are entered, and thus we later discuss how this structure can be transformed into the other structures.

The variable *Person* in the data set designates which member of the data the person is. One person is denoted as 1 and the other as 2. Having such a variable in the data file can be very helpful for some analyses—especially in studies that contain a categorical within-dyads variable (e.g., in married couples, husbands and wives; in sibling dyads, older and younger).

It is advisable to arrange the individual data set so that the data from each member of the dyad are adjacent: Units 1 and 2 represent data from

the two members of dyad 1, units 3 and 4 are from dyad 2, and so on. Having the data ordered in this way facilitates certain analyses and is required for others. In addition, if dyad members are distinguishable, ordering the two members systematically using that distinguishing variable can also be useful. For example, if the dyads are married couples, the husband data would be entered consistently before (or after) the wife data, resulting in a data file in which odd-numbered units would be the husband (wife) data and even-numbered units would be the wife (husband) data. Such an ordering may not be required, but it is still advisable.

Using the individual structure has major disadvantages. The first is that the structure encourages researchers to analyze the data with person as the unit of analysis. Such a data analysis strategy ignores nonindependence, and so is ill advised. The second is that it fails to allow for the influence that partner characteristics can have on the person. The next two structures do allow for that possibility.

### Dyad Structure

In this case there is a single unit for each dyad. If there were  $n$  dyads and  $2n$  individuals, there would be  $n$  records in the dyad file. So the example in Table 1.3 shows three records, one for each dyad. Each unit would have only a single score for dyad-level variables (e.g.,  $Z$ , which might be length of relationship), but there would be two variables,  $X_1$  and  $X_2$ , for each individual-level variable. The variable  $X_1$  refers to person 1's score on  $X$  (e.g., attachment avoidance), and  $X_2$  refers to person 2's score on  $X$ . For example, in a study of roommates, each unit would have one score measuring the total cost of renting an apartment, two scores measuring percentage of housework done by each individual, and two scores measuring general satisfaction with the living arrangements.

Note that an individual file can be read as a dyad structure as long as it is arranged so that dyad members are adjacent. That is, the individual file would be sorted by dyad, and so persons 1 and 2 would be members of the same dyad, persons 3 and 4 the same, 5 and 6, and so on. If this were done, it would be unnecessary to read the dyad-level variables twice, and they would need to be read only on either the odd or even records.

We can create a dyad-structure data set from an individual-structure data set by merging records. The following SPSS syntax would be used to convert the structure of the individual data file depicted in the top panel of Table 1.3 to the dyadic data structure depicted in the middle panel of this table. To use this transformation procedure, it is crucial that there be a

variable that identifies the dyad, such as *Dyad* in the example data set. The SPSS syntax for creating a dyad structure from an individual structure is

```
DELETE VARIABLES Person.
CASESTOVARS
  /ID = Dyad
  /GROUPBY = INDEX
```

We first delete the variable *Person*. Then the following variables are created: *Dyad*, *Z*, *X.1*, *Y.1*, *X.2*, and *Y.2*. Note that SPSS creates a single variable *Z* because it is a between-dyads variable.

### Pairwise Structure

The pairwise structure is a combination of the individual and dyad structures in the sense that there is one record for each individual but both partners' scores occur on each record as well. More specifically, in this file structure (sometimes called a *double-entry structure*), each record includes the person's scores on each of the variables, as well as the person's partner's scores on each of the individual-level variables. So, in a study of married partners, on the wife's data record the wife's scores would be entered as one set of variables, and the husband's scores would be entered as "partner" variables. For example, there might be two variables: *SATISFACTION* (which on the wife's record would be the wife's score on satisfaction) and a variable *PARTNER SATISFACTION* (which on the wife's record would be the husband's score on satisfaction).

As seen in Table 1.3, the pairwise structure is similar to the dyad structure in that it has two sets of *X*, *Y*, and *Z* variables. There are two key differences. First, the pairwise structure has a variable that designates the *Person*. Second, the meanings of the two variables, for example, *X<sub>1</sub>* and *X<sub>2</sub>*, are very different for the two structures. For a pairwise structure, *X<sub>1</sub>* refers to the person whose record it is, and *X<sub>2</sub>* refers to that person's partner. For a dyad structure, *X<sub>1</sub>* refers to person 1, and *X<sub>2</sub>* refers to person 2.

One can create a pairwise structure by cutting and pasting the data from either an individual or a dyad structure. One would first sort the dyad structure by member such that the first *n* records would be for person 1 of the dyad and the last *n* for person 2. One would then copy the data for person 1 and paste it for person 2, and vice versa. For this strategy to work, there can be no missing records, and if a person is missing, a dummy record has to be created.

### Other Designs

Although the standard design is by far the most common design in dyadic research, one-with-many and SRM designs also occur (about 28% of the time; see the next section). Recall that in the one-with-many design each person is linked to multiple others, but these others are linked only to that one person. As an example, consider a design in which each research participant interacts with a confederate who either acts interested or uninterested in getting to know the participant. To strengthen the generalizability of the research, 8 confederates participate in the study, and 10 participants interact dyadically with each confederate, making the total sample size 80. The data can be organized either by the persons (i.e., participants) or by the focal person (i.e., confederate) who has links to the other persons. If the data were organized by person, each record should include an identification variable for the focal person who has links to the other persons (i.e., the confederate in the example). In this way the data can be sorted so that they are linked together. This strategy is particularly useful when there are an unequal number of persons paired to each focal person.

Alternatively, the unit can be the focal person (i.e., the confederate in the previous example) and all of the data about partners (i.e., participants) can be on a single record. A common research design that might call for such an organization scheme occurs when participants are asked to report on their own dyadic relationships with multiple partners. For instance, if a person rates his or her closeness to each member of his or her social network, those ratings might all be placed in a single record. If there are multiple variables, as there usually are, researchers must decide whether partner or variable is "fastest moving." If there are three variables and five partners, the data would have "variable fastest moving" if the three variables for the first partner come before the three variables for the second partner, and so on. "Partner fastest moving" would occur if the five partners' scores on the first variable come before the five partners' scores on the second variable, and so on. Many computer programs require the user to specify whether partner or variable is faster moving.

SRM designs can be viewed as an extension of the one-with-many design with one major difference: Not only does each person interact with or rate multiple partners, but each partner also interacts with or is rated by multiple persons. For example, as in the one-with-many design, in a round-robin SRM design, person A is paired with persons B, C, and D. In an SRM design, however, person B is also paired with A, C, and D, C is paired with A, B, and D, and D is paired with A, B, and C. Thus, in this



design, each individual serves as both an actor (rater) and a partner (target). Basically, there are two ways to order the data from an SRM design. For "dyad input," each unit or record refers to a particular dyadic combination. Thus, in the example, there would be separate records for A's outcomes with B, A's outcomes with C, A's outcomes with D; B's outcomes with A, B's outcomes with C, and so on. If this data structure is used, it is helpful to include identification codes on each record indicating who the actor and partner are for that record.

An alternative strategy with SRM data is to use "person or actor input." In this format each unit contains all of the data from one actor. Thus person A would be treated as a unit that would contain all of A's outcomes with B, C, and D. The issue of whether variable or partner is fastest moving (discussed previously) must be considered with this type of data structure.

### A DATABASE OF DYADIC STUDIES

We conducted a survey of five major journals that often publish research involving dyads (*Child Development*, *Journal of Marriage and the Family*, *Journal of Personality and Social Psychology*, *Journal of Social and Personal Relationships*, and *Personal Relationships*). For each of the five journals, we started with the last paper in the last issue of the year 2000 and worked backward in time until we found 15 dyadic studies per journal—yielding a set of 75 studies in all. We eliminated the following studies:

- Studies that used the same data as another study previously included in the database.
- Meta-analyses.
- Methodological studies.
- Qualitative studies.
- Simulation studies.
- Studies with artificial data.
- Studies that focused on groups and not dyads.
- Studies that used confederates or phantom others.
- Studies that had people rate persons in general, not particular others.

When the article included two or more studies or samples, we chose the first study or the study that was the primary study in the paper. In this

book, we sometimes refer to specific studies from the database, and at other times we characterize the general tenor of these studies.

The database is quite varied. The number of dyads included in the studies ranges from 16 to 4,066. The investigators are various types of psychologists (e.g., social, personality, developmental, and clinical), sociologists, family scientists, and communication scientists. Although most of the researchers were from North America, investigators also came from the Netherlands, Germany, Israel, Korea, and England.

We categorized the 75 studies into different design types. The results of the survey and in parentheses the results for the average study are as follows:

- Standard dyadic design: 54 studies
  - Reciprocal: 25 studies (101 dyads, 202 persons, both measured)
  - Nonreciprocal: 29 studies (200 dyads, 200 persons, one person measured)
- One-with-many design: 11 studies
  - Reciprocal: 1 study (121 persons paired with 2 partners)
  - Nonreciprocal: 10 studies (200 persons paired with 4 partners)
- SRM design: 10 studies
  - Reciprocal: 5 studies (254 persons with 4 partners)
  - Nonreciprocal: 5 studies (68 persons with 2 partners)

### OVERVIEW OF THE BOOK

When is a book that discusses data-analytic issues for dyads not needed? Some dyadic data can easily be handled by methods developed for individual data. There are three cases in which "dyadic" data can be treated as individual data.

First, 39% of "dyadic" studies (the 29 nonreciprocal standard design studies) were really just studies of single persons. An example of these nonreciprocal studies might be one in which women who are involved in heterosexual dating relationships rate their commitment to the relationship.

Second, consider a study of father-child relationships in which both child and father are measured, but in which one set of variables is measured for the father and another set is measured for the child. For example, fathers might be asked about child-rearing style, and the child's self-esteem



might be measured. Although such data are clearly dyadic, the dyad can be treated as an "individual."

Third, a dyad might be studied, but the outcome variable might be a between-dyads variable. For example, dyads are asked to solve anagrams, and the outcome is number of anagrams solved. For such a study, we can use individual methods of data analysis, but we treat the dyad as an "individual."

For those who need to read the book, we urge everyone to read Chapter 2. It defines the concept of nonindependence and considers its measurement. Because nonindependence is the fundamental concept of dyadic analysis, it needs to be thoroughly understood. We refer to nonindependence in every other chapter of the book.

Both multilevel modeling (MLM) and structural equation modeling (SEM) are valuable tools in the estimation of dyadic models. In Chapter 4, we show how MLM can be used to estimate models for dyadic data, and Chapter 5 presents models that are estimated by using SEM.

Much of what we discuss depends on the design of the research. Most readers will be interested in the standard design in which each person is linked or tied to just one other person in the study. If this is the case, then Chapters 3, 6, and 7 should be read. If means or regression coefficients are of primary interest, then Chapters 3 and 7 should be read, though reading Chapters 4 and 5 may be necessary before reading Chapter 7. Chapters 12, 13, and 14 may be relevant if the study has multiple outcome variables and the researcher is interested in conducting an analysis on each dyad, an idiographic analysis. In Chapter 12, we discuss dyadic indices of agreement and similarity, and in Chapters 13 and 14, we discuss over-time methods, and all of these chapters are useful for the analysis of data from the standard design.

Although most of the book presumes a standard design in which each person is a member of one and only one dyad, sometimes a person is a member of more than one dyad. Earlier we made a distinction between SRM- and one-with-many designs. If the researcher's design is an SRM design, then Chapters 8 and 9 are important. If the one-with-many design is used, then it is still advisable to read about the SRM design in Chapters 8 and 9, because some concepts discussed in these chapters are presumed in the discussion of the one-with-many design in Chapter 10. Chapter 11 considers an SRM design, but the level of measurement is categorical.

The reader might be tempted to read selectively; however, we urge the reading of the entire book. Although books necessarily have a sequential

or linear structure of one topic following another, dyadic data analysis is a complex topic that is not necessarily well characterized by a linear progression. Many topics could have been placed in several different chapters. Moreover, we were very surprised to discover that many topics that appeared to be fairly simple were much more complex than we thought. Thus a chapter that might have no intrinsic interest to the reader may provide a useful tool in another context. For instance, in Chapter 12, we discuss the use of pseudo-couple analysis in the study of profile similarity. This strategy represents the random pairing of couples to create a baseline measure. The strategy of pseudo-couple analysis can be very useful for other topics besides profile similarity.

The book emphasizes computer applications. We even sometimes give specific syntax for SPSS and SAS. That syntax is sure to change, and the reader is urged to consult the website <http://davidakenny.net/kkc.htm> for updates, changes, and elaborations. We also invite the reader to send us suggestions and corrections for the software updates.

## SUMMARY AND CONCLUSIONS

In this chapter we provided definitions that are crucial in dyadic research. We defined nonindependence and discussed processes that can generate it, including compositional effects, partner effects, mutual-influence effects, and common-fate effects. We also defined distinguishability of dyad members and types of independent variables that are used in dyadic research. We noted that between-dyads variables vary from dyad to dyad, but within a dyad both individuals have the same score on that variable. Within-dyads variables, on the other hand, vary across the two dyad members but do not vary on average from dyad to dyad. Variables that vary both between and within dyads were defined as mixed variables.

We also introduced three basic dyadic designs: the standard design, the SRM design, and the one-with-many design. The analysis of data that arise from these designs is the central topic of this book. Perhaps one of the most important and pragmatic sections of this chapter was our discussion of data organization. Finally, we presented a database of dyadic studies that informs our discussion of dyadic data analysis in the remainder of this book.

As a final note of introduction, in this book we presume that the members of the dyad are two people. This need not be the case. The dyad

might be two ears or two eyes from the same person or even two personalities. Alternatively, the members of the dyad might be groups of people or countries. The key idea is that the pair of scores are nonindependent—the topic of the next chapter.

## NOTES

1. In some presentations, *indistinguishable* is called *exchangeable*.
2. Fraley and Waller (1998) have empirically determined that adult attachment is an interval, not a nominal, measurement.

# 2

## The Measurement of Nonindependence

In this chapter, we consider the topic that is at the core of dyadic data analysis: the nonindependence of observations from the two dyad members. This is the most important chapter in this book because dyadic data analysis begins with an analysis of the degree of nonindependence.

In the previous chapter, we gave the following definition for nonindependence: If the two scores from the two members of the dyad are nonindependent, then those two scores are more similar to (or different from) one another than are two scores from two people who are not members of the same dyad. We presume that the nonindependence is linear. Although nonlinear nonindependence is possible,<sup>1</sup> it is rarely investigated.

The standard statistical methods that most behavioral and social scientists currently use, analysis of variance (ANOVA) and multiple regression, were developed for agricultural researchers. Those researchers knew that two ears of corn from the same stalk were similar to one another and even that two ears from the same row were similar due to cross-fertilization. They created separate plots of land and computed the crop yields in the different plots. These plots served as “participants.” It could then be safely assumed that the yields from different plots were independent of one another. The key assumption of ANOVA and multiple regression is that once variation due to the independent variables is controlled, the scores of different units are independent.

Why is the independence assumption so important for valid statistical inference? The answer concerns the fundamental logic of statistical analysis. Inference in data analysis is a form of induction and is based on repeatedly replicating a phenomenon. The exact number of replications of a phenomenon must be known in order to conduct a statistical analysis.

Consider a study of the effect of two different methods of conflict resolution on dyad members' attitudes toward one another after a discussion. If a sample of 80 individuals was divided into 40 dyads, and 20 dyads were then randomly assigned to each of the two conflict-resolution conditions, there would be 20 dyads (a total of 40 individuals) within each condition. One way to approach such a study would be to treat individual as the unit of analysis and analyze the data as if there were simply 80 individuals, 40 within each level of conflict style. In this approach there would be a total of 80 data points contributing to estimates of differences in attitudes between the two conflict-resolution methods. However, because the dyad members interact, they likely influence one another, and within a dyad the two members may share the same attitude. In this case, there are really only 20 unique or independent pieces of data within each condition, one from each dyad. Thus dyad, not individual, should be the unit of analysis. If the nonindependence is ignored and individual is treated as the unit, the conclusion from the study will be based upon 40 more pieces of evidence than it should be. The key idea in statistical analysis is independent replication, and nonindependence undermines independent replication.

In this chapter, we focus on measuring nonindependence, that is, developing a quantitative index of the degree of nonindependence. We consider variables measured at either the nominal or the interval level of measurement (see Chapter 1 for definitions). So, if we asked members of couples "Are you satisfied, yes or no?" the level of measurement would be nominal. Alternatively, if we asked them to complete a scale of marital satisfaction such as the Locke-Wallace (1959) marital satisfaction scale, the level of measurement would be interval.

We also assume that the design of the study is a reciprocal standard dyadic design (see Chapter 1). We found in our survey that 33% of the studies involving dyads used this design. Measuring nonindependence in the SRM and one-with-many designs is more complicated and is discussed in Chapters 8, 9, and 10. Chapter 15 also discusses the measurement and interpretation of nonindependence.

In this chapter, we describe how to measure nonindependence, taking into consideration both the interval and nominal levels of measurement, as well as the cases of distinguishable and indistinguishable dyad members.

We then discuss the consequences of ignoring nonindependence and then consider strategies for handling nonindependence that we see as less than optimal. Finally, we consider the power of the test of nonindependence. There are many formulas in this chapter, and the reader might be advised to ignore them when initially reading the chapter. On a subsequent reading, the reader could concentrate on the necessary formulas. One thing to remember is that, throughout the chapter,  $n$  stands for the number of dyads, not the number of persons.

## INTERVAL LEVEL OF MEASUREMENT

### Distinguishable Members

Recall from the previous chapter that distinguishable members of a dyad are those for whom there is some variable that can be used to differentiate between the two persons. For example, parent-child dyad members would be distinguished by their roles, heterosexual romantic couples by gender, and siblings by their relative age.

#### *Bivariate*

Measuring nonindependence with interval-level scores and distinguishable dyad members is straightforward: We correlate the dyad members' scores using a Pearson product-moment correlation coefficient. The distinguishing variable supplies a systematic way of assigning scores as either the  $x$  score or the  $x'$  score, so that the Pearson correlation can be computed. For example, with a measure of marital satisfaction gathered from both husbands and wives, the men's satisfaction scores could be the  $x$  scores, and the women's could be the  $x'$  scores. The unit of analysis in computing the correlation is the dyad, not the individual, and the data file would have to be organized with dyad as the unit of analysis (see Chapter 1). A Pearson correlation coefficient can vary from  $-1$  to  $+1$ , and a value of 0 indicates no linear relationship between the two variables (i.e., independence).

It is important to realize that the correlation between dyad members' scores might be negative, as well as positive. For instance, if the measure is how much housework each member does, it might well be the case that the correlation is negative: The more work one person does, the less work the other does. Thus the test of the correlation coefficient usually should be two-tailed, not one-tailed. If it made no sense at all for the correlation to be negative, the test would be one-tailed.

The statistical test of the null hypothesis that the correlation coefficient equals 0 is

$$t(n-2) = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

( $r$  being the Pearson correlation and  $n$  being the number of dyads), which under the null hypothesis that the population correlation is zero has a  $t$  distribution with  $n-2$  degrees of freedom. This formula is the standard method for testing the statistical significance of Pearson correlation coefficients. For example, if  $n$  is 45 and  $r$  is .34, the result is  $t(43) = 2.37$ , which has a  $p = .022$ .

The usual assumptions for statistical significance testing apply to the test of the correlation coefficient, including the assumption of a bivariate normal distribution and the assumption of independence of the observations. In this case, however, independence refers to independence from dyad to dyad. If there are concerns about violation of the bivariate normality assumption, then other measures of association (e.g., Spearman's rho) might be computed, although inferential tests of correlation coefficients are relatively robust against violations of the normality assumption (Havlicek & Peterson, 1977).

Note that the correlation between the two members of a dyad can be very large, yet the two persons might have very different scores. For instance, consider the correlation between parents' and children's ratings of how bossy the parent is. That correlation might be large, and parents and children appear to agree, but it might be that children on average think that parents are much bossier than parents think they are on average. That is, once we control for the mean difference, children and parents agree in their assessments of parental bossiness. A correlation coefficient measures the correspondence between relative rank orderings, not absolute scores. When measuring nonindependence, the mean difference between the two dyad members should be controlled, as it is when the correlation coefficient is used.

Just how much nonindependence might we expect? Cohen (1988) defines .5 as a large correlation, .3 as medium, and .1 as small. We examined the 25 studies from our survey of 75 studies that used a reciprocal standard design. Of the 25 studies, only 2 reported the degree of nonindependence. We were both surprised and disappointed by the rareness of reporting this statistic. The degree of nonindependence is a fundamental datum of a dyadic study and should be reported. The average degree of nonindependence from these two studies was .4, but with so few studies this value is not very informative. Presumably the size of the correlation

largely depends on the type of dyad studied (correlations being larger for intact vs. experimental pairs) and the variable being measured (correlations being larger for more relational variables).

As we have noted, the measure of nonindependence can be positive such that the two members are relatively similar to one another, or it can be negative such that the two members are relatively different from one another. Negative nonindependence can occur when there is:

- *Compensation*: If one person has a large score, the other person lowers his or her score. For example, if one person acts very friendly, the partner may distance him or herself (Argyle & Dean, 1965).
- *Social comparison*: The members of the dyad use the relative difference on some measure to determine some other variable. For instance, satisfaction after a tennis match is determined by who won or lost that match.
- *Zero sum*: The sum of two scores is the same for each dyad. For instance, the two members divide a reward that is the same for all dyads. Similarly, the percentage of time that the two persons talk in an interaction is measured such that the total time sums to 100.
- *Division of labor*: Dyad members assign one member to do one task and the other member to do another. For instance, the amount of housework done in the household may be negatively correlated.

Although nonindependence in dyads is usually positive, it can be negative, and researchers should not presume that it is positive.

### Confidence Interval for $r$

Besides knowing whether a correlation is statistically significant, it is also useful to know its confidence interval. The confidence interval provides a way of determining the stability of a correlation; that is, the range of possibilities for the correlation if it were reestimated in a follow-up study. Additionally, by seeing whether 0 is in the confidence interval, we can perform a significance test. The formula for the confidence interval for the correlation requires first transforming  $r$  to Fisher's  $z$  or  $z_r$ , which uses the following formula:

$$z_r = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right),$$

where "ln" is the natural logarithm. The 95% confidence interval for this Fisher's  $z$  value is  $z_r \pm 1.96/\sqrt{n-3}$ . If a different width were desired, a value

different from 1.96 would be substituted. Typically, we would report the confidence interval for  $r$  by transforming the Fisher's  $z$  value confidence interval endpoints back to  $r$ 's by the following formula:

$$r = \frac{e^{2z_r} - 1}{e^{2z_r} + 1},$$

where  $e$  is the transcendental number that approximately equals 2.718. For example, with a correlation of .44 and  $n$  of 51, the Fisher's  $z$  value is  $z_r = .472$ , and the 95% confidence interval in terms of  $z_r$  is  $.472 \pm .283$ , or 0.189 to 0.755. Transforming the endpoints of the interval back to  $r$  yields a 95% confidence interval for  $r$  of .187 to .638. Because 0 is not in the confidence interval, the correlation is statistically significant. Note that because the upper limit for a correlation coefficient is 1, the confidence interval for  $r$  (but not  $z_r$ ) is asymmetric.

### Multiple Distinguishing Variables

Sometimes there is more than one distinguishing variable. For instance, Badr (2003) studied heterosexual married couples in which one of the two members had a serious illness. There are then two factors that can be used to distinguish dyad members: gender (male or female) and disease status (present or absent). In this situation it is not obvious which distinguishing variable should be treated as the critical variable in the analysis of nonindependence, and as far as we know, there is no simple way to answer this question. We propose the following strategy. First, determine which variable has larger differences in both means and variances on the key outcome variables (e.g., marital satisfaction). Note that differences in variances are especially critical. Then use the variable with the larger differences as the distinguishing variable. In this situation, the other distinguishing variable should be used as a control variable in the analysis of nonindependence, a topic that we now discuss.

### Controlling for Independent Variables

If there are independent variables (or additional distinguishing variables) in the study, we should control for these variables when measuring nonindependence. Failing to control for these variables may lead us to mistakenly conclude that there is nonindependence when, in fact, there is none. Less likely, though still possible, we might fail to find nonindependence when there is some. Either of these errors could occur if the independent variables have a large effect on the variable of interest.

We can control for these independent variables by computing a partial correlation between the two dyad members' scores, partialing out the effects of the independent variables. For instance, in correlating the degree of marital satisfaction between husband and wife, we may need to control for the length of marriage (or disease status, as in Badr, 2003).

The formula for a partial correlation, correlating variables 1 and 2 and partialing out variable 3, is

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}.$$

The correlation is tested like any correlation coefficient (see the preceding formula), but the degrees of freedom are  $n - 3$ , not  $n - 2$ . (Recall that  $n$  is the number of dyads in the study, not the number of persons.)

For example, if the correlation of husband's and wife's marital satisfaction is  $r_{12} = .50$ ,  $n = 100$ , and husband's satisfaction correlates with length of marriage,  $r_{13} = -.31$ , and wife's satisfaction correlates with length of marriage,  $r_{23} = -.33$ , then the partial correlation of marital satisfaction between husbands and wives, controlling for length of marriage, is .443. The test of statistical significance is  $t(97) = 4.87$ ,  $p < .001$ . Computing a partial correlation often, but not always, lowers the correlation.

Although it is not obvious, if the variable to be controlled is a second distinguishing variable, the same computations would be used, but the second distinguishing variable would need to be recoded. Consider the variable of disease status in the Badr (2003) study. A new variable, such as "disease status of the husband," could be created such that a 1 would indicate that the husband was the ill partner and a 0 would indicate that the husband was not the ill partner (rather, the wife was ill). The correlation  $r_{13}$  would then be the correlation between husband's satisfaction and husband's disease status, and, similarly, the correlation  $r_{23}$  would be the correlation between the wife's satisfaction and the husband's disease status.

The partial correlation can be viewed as a correlation between scores that have been "residualized." By residualized, we mean that a separate regression analysis for each type of dyad member (recall that we are discussing distinguishable dyads) has been run in which the independent variables (i.e., the variables to be controlled) are the predictors and in which the errors, or residuals, for each equation have been computed. The partial correlation can be viewed as the Pearson correlation between the errors across the dyad members. Note that one additional degree of freedom is lost for every variable that is controlled. Note also that if the value on the independent variable is not the same for the two dyad members

(i.e., if it is mixed), then there are two variables that must be controlled, and so two degrees of freedom are lost.

Multivariate Tests

In this case, we have a series of measures, and we wish to evaluate whether the set of variables is independent across the dyad members. We can use a canonical correlation (Tabachnick & Fidell, 2001) to perform such a test. We treat one member's scores as the predictor variables and the other member's scores as the criterion variables. For example, say that we have a study in which 10 measures are assessed for husbands and 10 measures are assessed for wives. We can compute the canonical correlation by having the husbands' data predict the wives' data, or we can equivalently have the wives' data predict the husbands' data. It makes no difference which person is used to predict the other, as the results would be identical. The test that the canonical correlation is 0 evaluates whether there is a correlation in observations (i.e., nonindependence).

It is possible, although rather unlikely, to obtain a statistically significant canonical correlation even when none of the bivariate correlations between the two dyad members is statistically significant. There are two ways that this could happen. First, it might be that all the correlations are small but that they are consistently positive or negative. Second, several of the correlations between different variables (e.g., between husband's satisfaction and wife's commitment) might be large, but the correlations of the same two variables might be small.

Indistinguishable Members

As is shown throughout this book, dyadic data from distinguishable dyads can often be analyzed using conventional methods. However, when the dyad members are indistinguishable (e.g., coworkers), new methods need to be developed. As is shown, we cannot use the ordinary correlation coefficient to measure nonindependence; rather, we use the intraclass correlation.

Consider as an example the data in Table 2.1 from a fictitious study of liking between same-gender roommates. In this dyad-structured data set there are 10 pairs of roommates, and individuals are asked to rate how much they like their roommates on a 9-point scale. The following demonstration illustrates why a Pearson correlation would be the wrong approach for estimating the correspondence between liking scores when

TABLE 2.1. Intraclass Correlation Example Using Data from a Fictitious Study of Roommates

Dyad	Scores				
	$x_i$	$x'_i$	$m_i$	$d_i$	
1	8	6	7.0	2	
2	5	3	4.0	2	
3	7	2	4.5	5	
4	8	5	6.5	3	
5	8	7	7.5	1	
6	5	6	5.5	-1	
7	3	4	3.5	-1	
8	8	9	8.5	-1	
9	6	7	6.5	-1	
10	2	3	2.5	-1	

$M = 5.60$ ;  $MS_B = 7.42$ ;  $MS_W = 2.40$ ;  $r_I = .51$ ;  
 $F(9,10) = 7.42/2.40 = 3.09$ ,  $p = .093$

dyad members are indistinguishable. Because there is no meaningful way to assign one set of scores as the  $x$  scores and the other set as the  $x'$  scores, we arbitrarily decide that the 10 scores in the left column (those entered on the data records first) are to be treated as the  $x$  scores and the 10 scores in the right column are to be treated as the  $x'$  scores. The Pearson correlation between  $x$  and  $x'$  is  $r(8) = .53$ ,  $p = .12$ . Now, because the assignment to  $x$  and  $x'$  was totally arbitrary, we might reverse the scores for the last five roommate pairs (e.g., now the  $x$  score for the last dyad is 3, and the  $x'$  score is 2). The Pearson correlation now is  $r(8) = .79$ ,  $p = .006$ . Clearly, the fact that the size and the statistical significance of the Pearson correlation can change substantially depending upon such an arbitrary rearrangement of the data implies that this particular statistic is not well suited for assessing the relatedness of scores from dyads with indistinguishable members.

The *intraclass correlation*, symbolized in this book as  $r_I$ , provides a unique estimate of the relationship between scores from indistinguishable dyad members. Intraclass correlations for dyads are interpreted in the same fashion as Pearson correlations. Thus, if a dyad member has a high score on a measure and the intraclass correlation is positive, then the other dyad member also has a relatively high score; if the intraclass correlation is negative, then the other dyad member has a relatively low score. A common alternative interpretation of a positive intraclass correlation is the proportion of variation in the outcome measure that is accounted for by

dyad. That is, if the intraclass correlation equals .40, then 40% of the variation in the scores is accounted for by the particular dyad to which individuals belong. The shared-variance explanation of the nonindependence becomes problematic when the correlation is negative.

### ANOVA Methods: Univariate

The intraclass correlation can be computed via several approaches, the first of which is based on ANOVA techniques. We denote the scores from two members of dyad  $i$  as  $X_{1i}$  and  $X_{2i}$ . There are a total of  $n$  dyads, and the average of all  $2n$  scores is denoted as  $M$ . Let

$$d_i = X_{1i} - X_{2i}$$

and

$$m_i = \frac{X_{1i} + X_{2i}}{2}$$

Thus  $d$  represents the difference between the dyad members' scores, and  $m$  represents the average of the scores. We define the mean square between dyads as

$$MS_B = \frac{2 \sum (m_i - M)^2}{n - 1},$$

and we define the mean square within dyads as

$$MS_W = \frac{\sum d_i^2}{2n}.$$

The intraclass correlation for dyads, or  $r_i$  (see Chapter 10 for a more general formula), is then defined as

$$r_i = \frac{MS_B - MS_W}{MS_B + MS_W}. \quad (2.1)$$

Tests of the statistical significance of the intraclass correlation depend on the sign of  $r_i$ . If the correlation is positive, the expression  $MS_B/MS_W$  (or, equivalently,  $[1 + r_i]/[1 - r_i]$ ) has an  $F$  distribution with  $n - 1$  degrees of freedom on the numerator and  $n$  on the denominator. If the correlation is negative (i.e., dyad members are less similar than two persons in different dyads are), the term  $MS_W/MS_B$  (or, equivalently,  $[1 - r_i]/[1 + r_i]$ ) has an  $F$  distribution with  $n$  degrees of freedom on the numerator and  $n - 1$  on the

denominator. Because the test is two-tailed (unlike the usual  $F$ -test), the tabulated  $p$  value must be doubled. For the example data in Table 2.1,  $MS_B = 7.42$ ,  $MS_W = 2.40$ , and the intraclass correlation equals .51. Thus, in this example, the test of the intraclass correlation is  $F(9, 10) = 3.09$ ,  $p = .093$ , two-tailed.

The values of  $MS_B$  and  $MS_W$  can also be computed directly from an ANOVA. The data would need to be organized as either a pairwise or an individual structure (see Chapter 1). The independent variable is dyad, which has  $n$  levels. The  $MS_B$  measures the variance in the dyad means times 2, and equals 0 when all the dyad means are equal. The  $MS_W$  measures the variance in the two scores in the dyad divided by 2, and it is 0 when the two members of each dyad both have the same score. The intraclass correlation itself equals 0 not when  $MS_W$  is 0, but rather when  $MS_B = MS_W$ . In such a case, the variability in the dyad means can be explained by differences within the dyads: Two members of the same dyad are as similar as two members of different dyads.

One interesting historical fact is that the famous statistician R. A. Fisher invented the intraclass correlation. Not so well known is that Fisher developed the ANOVA from the intraclass correlation. The intraclass correlation is a measure of effect size in a one-way ANOVA.

The ANOVA intraclass correlation is a biased measure. As seen in Table 2.2, when the true correlation is zero, the estimated intraclass correlation tends to be negative. This bias declines as the number of dyads,  $n$ , increases and is trivial as long as  $n$  is greater than 30. For example, when the true correlation is .500 and the sample size is 50, the average estimate of the correlation

TABLE 2.2. Bias in the Estimation of the ANOVA Intraclass Correlation ( $r_i$ )

$n^a$	True intraclass correlation				
	-.5	-.3	-.1	0	.1
5	-.434	-.261	-.102	-.025	.053
10	-.465	-.276	-.095	-.006	.084
15	-.476	-.283	-.096	-.002	.091
20	-.482	-.287	-.096	-.001	.094
25	-.485	-.290	-.097	-.001	.095
30	-.488	-.291	-.097	-.001	.096
40	-.491	-.293	-.098	-.000	.097
50	-.492	-.295	-.098	-.000	.098
60	-.494	-.296	-.098	-.000	.098

<sup>a</sup>Number of dyads.



is .492, a bias of only .008. We also note that the Pearson product-moment correlation is biased when nonzero, but this bias is usually ignored. The bias is mainly an attenuation bias in that the estimated correlation is somewhat smaller in absolute value than the true correlation.

The formula for the confidence interval for the intraclass correlation (Shrout & Fleiss, 1979) is relatively complicated. First, determine the upper and lower critical values for the  $F$ -test of the intraclass and denote them as  $F_U$  and  $F_L$ , and denote  $F$  as the computed value of the  $F$ -test of the actual intraclass correlation. Then the upper limit of the confidence interval is given by

$$\frac{(F_U)(F) - 1}{(F_U)(F) + 1}$$

and the lower limit by

$$\frac{(F_L)(F) - 1}{(F_L)(F) + 1}.$$

For example, consider an intraclass correlation of .44 with 51 dyads. The test of statistical significance is  $F(50, 51) = 2.571$ . The value for  $F_U$  (the value that  $F$  would have to be significant for a positive  $r_i$  with 50 and 51 degrees of freedom, respectively) is 1.746 (note that the area under the curve for this  $F$  distribution from 0 to 1.746 is .975); and the value for  $F_L$  (the value that  $F$  would have to be significant for a negative  $r_i$  with 50 and 51 degrees of freedom, respectively) is 0.572 (note that the area under the curve for this  $F$  distribution from 0 to 0.572 is .025). The 95% confidence interval extends from .190  $[(0.572)(2.571) - 1]/[(0.572)(2.571) + 1]$  to .636  $[(1.746)(2.571) - 1]/[(1.746)(2.571) + 1]$ . Because 0 is not in the confidence interval, the correlation would be statistically significant. It is interesting that this confidence interval is virtually identical in size to the confidence interval for the Pearson  $r$ . It seems reasonable to use the confidence interval for the Pearson correlation as an approximation of that for  $r_i$ .

The reader might wonder what would happen if an intraclass correlation were computed with distinguishable members of the dyad: How similar would the intraclass correlation be to the ordinary correlation? Typically, the two values are quite similar. They are nearly identical when the means and variances for the two types of members (e.g., husbands and wives) are similar.<sup>2</sup> However, if the means and variances are quite different, the intraclass correlation is usually smaller than the ordinary correlation coefficient. Gonzalez and Griffin (1999) recommend testing whether

the means and variances are equal across the distinguishing variable for distinguishable dyad members. If they are not statistically different, the dyad can be treated as if the members were indistinguishable. In the next chapter, we review how to test the equality of means when dyad members are distinguishable, and in Chapter 6 we present methods for testing equality of variance.

### ANOVA Methods: Partial Correlations

As seen earlier, the intraclass correlation can be defined in terms of mean squares,  $MS_B$  and  $MS_W$ . Most dyadic studies contain causal variables, experimental variables, couple-level variables (such as length of the relationship), and individual differences (such as personality variables). If there are variables to control for, we need to adjust these two mean squares and their degrees of freedom. One strategy for making the necessary adjustments is to compute a set of two regressions using a dyad-structured data set and combine their results. In the first regression, the sum of the two persons' outcome scores are regressed on any between-dyad variables, as well as the sum of any mixed variables. The sum of squared residuals from this regression equation equals  $SS_B$ . Because the degrees of freedom equal  $n - k - 1$ , where  $k$  is the number of predictor variables in the regression equation, the  $MS_B$  equals  $SS_B/(n - k - 1)$ . In the second regression equation, the dependent variable is the difference in the dyad outcome scores, and predictor variables include the dyad difference scores for any mixed variables and any within-dyad independent variables. Note that the intercept should not be included in this second equation. The intercept is not included because the direction of the difference is arbitrary. (We discuss this extensively in Chapter 3.) The sum of squared residuals from this second regression equals  $SS_W$ , with degrees of freedom equal to  $n - k$ , where  $k$  is the number of predictors in the regression equation. The ratio of this sum of squares to its degrees of freedom yields the  $MS_W$ . The intraclass correlation would then be estimated and tested as before with the new degrees of freedom.

### Pairwise Correlational Methods

There is an alternative way of computing the intraclass correlation—the double-entry method. This method, which is as old as the ANOVA method, has been introduced by Griffin and Gonzalez (1995; Gonzalez & Griffin, 1999) into the psychological literature. They refer to the correla-



tion as the *pairwise correlation*, and we symbolize this correlation as  $r_p$ . In this method, one of the persons is designated as  $X$ , and the other  $Y$ . Then the data are doubled, but the two persons' scores are flip-flopped, making each  $X$  a  $Y$  and each  $Y$  an  $X$ . Thus the sample size is  $2n$ , not  $n$ . The correlation is computed for the "double-entered" data set using the standard Pearson correlation formula. An equivalent formula for the pairwise correlation is

$$r_p = \frac{SS_B - SS_W}{SS_B + SS_W},$$

where  $SS_B = df_B MS_B$  and  $SS_W = df_W MS_W$  (Gonzalez & Griffin, 2001a). As with the ANOVA intraclass correlation, the formula changes if groups, not dyads, are studied.<sup>3</sup>

A pairwise correlation cannot be tested in the usual way. Griffin and Gonzalez (1995) recommend using  $1/\sqrt{n}$  as the standard error for the test of  $r_p$ . The resulting test statistic is treated as a  $Z$  statistic. Such a test is asymptotic, and under the null hypothesis it approximates the  $Z$  distribution as the sample size increases. With a small number of dyads ( $n < 30$ ), the test is somewhat liberal. The 95% confidence interval for  $r_p$  would be  $r_p \pm 1.96/\sqrt{n}$ .

We have reorganized the data in Table 2.1 to compute  $r_p$  and present it in Table 2.3. The resulting correlation is .47, a value slightly lower than the ANOVA intraclass correlation value. The  $Z$  test of the null hypothesis is 1.49,  $p = .137$ . The 95% confidence interval for  $r_p$  is  $.47 \pm .620$ , or  $-.150$  to  $1.090$ . However, because the largest value that  $r_p$  can have is 1.00, the interval extends from  $-.150$  to 1.000.

The difference between the two measures of the intraclass correlation, ANOVA minus pairwise, is

$$r_i - r_p = \frac{2SS_B SS_W (q-1)}{(SS_B + SS_W)(qSS_B + SS_W)}, \quad (2.2)$$

where  $q = df_W/df_B$ . Typically,  $r_p$  is a bit smaller than  $r_i$ . Both measures are negatively biased, but the pairwise measure is more biased. The amount of bias lessens as the sample size increases, becoming essentially trivial when  $n$  is 50 or more. Note, too, that the standard error is larger and the confidence interval is wider for  $r_p$  than for  $r_i$ . The reason is that the pairwise standard error is asymptotic. The pairwise correlation can be shown to be a maximum likelihood estimate of the correlation between the two persons' scores, and the ANOVA approach is the restricted maximum likelihood estimate (Gonzalez & Griffin, 2001a).

**TABLE 2.3.** Pairwise Correlation Example Using Data from a Fictitious Study of Roommates (see Table 2.1)

Dyad	Scores	
	$x_i$	$x_j$
1	8	6
2	5	3
3	7	2
4	8	5
5	8	7
6	5	6
7	3	4
8	8	9
9	6	7
10	2	3
1	6	8
2	3	5
3	2	7
4	5	8
5	7	8
6	6	5
7	4	3
8	9	8
9	7	6
10	3	2

$$r_p = .47, Z = 1.49, p = .136$$

### Multilevel Modeling

The intraclass correlation can also be computed using a statistical technique called *multilevel modeling* (MLM). We describe this method in detail for the standard dyadic design in Chapter 4, but for our purposes here we assume that the reader is familiar with this technique. The intraclass correlation would be equal to the dyad variance (the variance of the dyad intercepts) divided by the sum of the residual variance plus the dyad variance. A major problem with using the multilevel approach occurs if the intraclass correlation is negative, because the dyad variance is estimated as zero in this case, and therefore the intraclass correlation is estimated to be zero. We discuss the use of MLM to estimate the intraclass correlation in more detail in Chapter 4.

### Using the Intraclass When Members Are Distinguishable

In some instances it is appropriate to compute the intraclass correlation even when members are distinguishable. Consider the study by Sneeuw, Alberisen, and Aaronson (2001), who studied 72 patients with prostate cancer and their spouses. Patients and spouses independently completed several questionnaires. Because the purpose of the study was to investigate using spouse's reports as a proxy for patient reports, the intraclass correlation is the appropriate measure of nonindependence in that it measures exact agreement in responding.

### CATEGORICAL MEASURES

We have thus far assumed that the level of measurement is interval, but sometimes the level of measurement may be nominal, and therefore dyad members' responses are categorical. In this section we consider the measurement of nonindependence for such variables.

#### Distinguishable Members

With a nominal variable, each member of the  $n$  dyads responds to a question that has two or more categories; they may be asked their ethnicity or which of several political candidates they prefer. As an example, consider the case in which we have both husbands and wives each state their religion. The religions might be Catholic, Protestant, Jewish, Muslim, Other, and None. We would then have a resulting 6-by-6 table with husband's choice on one side and wife's choice on the other. We can use Cohen's kappa (Cohen, 1960), or  $\kappa$  to determine whether the scores are independent or not. If  $\kappa$  were different from zero, then there would be evidence of nonindependence. In this case nonindependence would indicate that husbands and wives are more likely to be members of the same religion than members of two different religions.

Table 2.4 presents hypothetical data from 200 dyads in which each member of a distinguishable dyad makes a categorical response of either A, B, or C. To estimate kappa, we compute  $n_o$ , the number of times the two members of the dyad have the same response, that is, the number of agreements. In Table 2.4,  $n_o = 106 + 28 + 6 = 140$ . We also compute  $n_e$ , the expected or chance number of agreements. For any given category the expected number of agreements is the product of the number of times each judge chose the category divided by the total sample size. That is, the

expected number of agreements is the row marginal frequency times the column marginal frequency divided by the number of dyads. (This is the same way expected frequencies are obtained in a standard chi-square test of independence.) The term  $n_e$  is the sum across categories of the expected number of agreements. In the example,  $n_e = (130)(120)/200 + (50)(60)/200 + (20)(20)/200 = 95$ . Cohen's kappa equals

$$\kappa = \frac{n_o - n_e}{n - n_e},$$

where  $n$  is the number of dyads. For the example,  $\kappa = (140 - 95)/(200 - 95) = .429$ . If we define  $p_o$  as  $n_o/n$  and  $p_e$  as  $n_e/n$ , kappa can be alternatively written as

$$\kappa = \frac{p_o - p_e}{1 - p_e}.$$

As another example, Christensen, Sullaway, and King (1983) asked 50 heterosexual couples to report on their interactions over the preceding 24 hours. Each member decided, for instance, whether he or she had had an argument in the past 24 hours. The average value of kappa across the behaviors in the study was about .50.

A kappa of zero implies that  $p_o = p_e$  and that agreement is at chance levels. A kappa of 1 implies that  $p_o = 1$  and there is complete agreement. Kappa is affected by margin differences. If, for instance, the wives are less likely to designate their religion as "None" than their husbands are, this would lower the value of kappa. Kappa can potentially equal 1 only when the margins are exactly equal. The upper limit of  $p_o$  is  $\Sigma \text{Min}(p_i, p_j)$  where "Min" means the smaller of the two values. The term  $p_i$  is the number of times the first judge chose category  $i$  divided by  $n$ , and  $p_j$  is the number of times the second judge chose category  $j$  divided by  $n$ . (Note for this example that  $p_i$  is  $120/200 = .60$  and  $p_j$  is  $130/200 = .65$ .) For the example in

**TABLE 2.4.** Kappa Example with Two Members and Three Categories (A, B, and C)

	Category	Member 2			Total
		A	B	C	
Member 1	A	106	10	4	120
	B	22	28	10	60
	C	2	12	6	20
	Total	130	50	20	200

Table 2.4, given the margins, the maximum value of kappa is not 1 but .905. If kappa is negative, then agreement is less than expected by chance. Note that the lower limit of kappa can be less than -1.

The standard error of kappa, assuming that kappa is zero, which we designate as  $s_{\kappa}$  (the standard error for  $\kappa$  given the null hypothesis that kappa is zero) equals

$$s_{\kappa} = \sqrt{\frac{p_c + p_e - \sum [p_i p_i (p_i + p_i)]}{n(1 - p_e)^2}},$$

where the summation is across categories (Fleiss, 1981). To test whether kappa is statistically greater than zero, it is divided by its standard error. The resulting value is a one-tailed Z test.

The computation of a confidence interval for kappa is fairly complex. We cannot use the previous standard error, or  $s_{\kappa}$ , as it presumes, under the null hypothesis, that kappa is zero. The proper standard error for a confidence interval, which we denote as  $s_{\kappa}$ , is

$$s_{\kappa} = \sqrt{\frac{\sum p_i [1 - (p_i + p_i)(1 - \kappa)]^2 + (1 - \kappa)^2 \sum p_i (p_i + p_i)^2 - [\kappa - p_i(1 - \kappa)]^2}{n[1 - p_e]^2}}.$$

The 95% confidence interval for kappa is  $\kappa \pm 1.96s_{\kappa}$ .

For the example data in Table 2.4, the estimated standard error of kappa is  $s_{\kappa}$ , which is 0.0555, and the test of the null hypothesis yields  $Z = 7.72$ , which is statistically significant,  $p < .001$ . The other standard error, or  $s_{\kappa}$ , is 0.0537, and so its confidence interval is  $.429 \pm .105$ , or between .323 and .534.

Kappa can be modified to weight various types of discrepancies more than others, and the resulting measure is called *weighted kappa* (Cohen, 1968). For the religion example, it might be that a Catholic-Protestant (a Catholic married to a Protestant) marriage is less discrepant than a Non-Muslim marriage. As discussed by Krippendorff (1970), if the weights are chosen in a certain fashion (categories are given a numeric score and the weights are the squared difference between scores), the value of kappa is identical to an ANOVA intraclass correlation coefficient. Thus the intraclass correlation can be viewed as a special case of kappa.

### Indistinguishable Members

We know of no previous work that presents methods for assessing and testing nonindependence on a categorical variable when dyad members are indistinguishable. Such a situation might occur in a study investigating

similarity in marital status (e.g., single-never married, married, divorced, and so on) for same-sex best friends.

We suggest the following strategy. Arbitrarily designate the two dyad members as persons 1 and 2 and create a two-way table of association. Next, combine corresponding off-diagonal cells and average them. Considering the data in Table 2.4, the average of the two AB cells is 16, the average of the AC cells is 3, and the average of the BC cells is 11. This average may not be an integer (e.g., 12.5), but this is not problematic. We place this average in each of the two corresponding off-diagonal cells. Thus the AB (i.e., the cell in which member 1 says "A" and member 2 says "B") and BA cells (i.e., the cell in which member 1 says "B" and member 2 says "A") are both 16. We then recompute the row and column margins, which now should be equal. The A margin is now 125, the B margin is 55, and the C margin is 20. Using frequencies in the resulting new symmetric table, we compute Cohen's kappa. For the data in Table 2.4, we obtain a value of .427.

This measure of kappa can be interpreted in the same fashion as the standard measure of kappa. One thing to note is that one source of disagreement, marginal disagreement, is impossible. Additionally, the standard error presented previously is likely biased, and so using it in a test of statistical significance would be problematic. However, we know of no alternative.

### CONSEQUENCES OF IGNORING NONINDEPENDENCE

In this section we consider the effect of treating person as the unit of analysis and ignoring the nonindependence in dyad members' scores on standard significance testing. As we show, sometimes tests of statistical significance are too liberal, and at other times the tests are too conservative. If a test is too liberal, the inferential test statistic ( $t$  or  $F$  statistic) is too large in absolute value, and the  $p$  value is smaller (i.e., more "significant") than its nominal value (e.g., .05). A liberal test results in too many Type I errors, concluding that the null hypothesis is false when it is not. If a test is too conservative, the inferential test statistic ( $t$  and  $F$  statistic) is too small in absolute value, and the  $p$  value is larger (i.e., less "significant") than it truly is. With a conservative test, the likelihood of rejecting a true null hypothesis (i.e., a Type I error) is lower than the nominal value of .05; therefore, such a test has less power than it should and results in an increase in the probability of a Type II error.

We should note that nonindependence does not bias the effect estimates themselves. So *unstandardized* regression coefficients and mean differences are not affected by nonindependence. What are biased are the variances, and the biased variances then likely affect the standard errors of test statistics (e.g.,  $t$  and  $F$ ), making tests of statistical significance, and their associated  $p$  values, biased. Because variances are biased, standardized measures (e.g.,  $r$ ,  $\beta$ , and  $d$ ) are also biased.

In addition to the bias in the variance, there is also a bias in the degrees of freedom in tests of significance. Nonindependence always results in fewer degrees of freedom than there would be if the data were independent. This loss in degrees of freedom is easy to see when there is perfect nonindependence. With perfect nonindependence, the information supplied by one dyad member is completely redundant with that supplied by the other person, and although there are  $2n$  data points, there are only  $n$  unique pieces of information. When nonindependence is weaker, fewer degrees of freedom are lost. The exact quantitative formulas for degrees of freedom loss are presented in Kenny, Kashy, and Bolger (1998). However, if the nonindependence is less than .5, the loss in degrees of freedom has relatively little effect on significance testing results. For instance, if the number of dyads is 50 and the intraclass correlation is .5, the effective degrees of freedom are 78.08, not 98. However, the critical  $t$  (alpha of .05) is increased to only 1.9845 from 1.9908, a trivial increase.

As we stated earlier, the typical effect of nonindependence is to bias variances. If the correlation between dyad members' scores were positive, then the variance of the observations would be smaller than it should be; and if the correlation were negative, then the variance would be larger than it should be. As we have mentioned, the bias in  $p$  values can either be too conservative or too liberal. A key factor in determining the direction of the  $p$  value bias is the nature of the independent variable—that is, whether it is a within- or between-dyads independent variable. This bias also depends on the direction of the correlation between the two members on the outcome variable. So we have two factors—the type of independent variable and the direction of the correlation on the outcome variable. Together they form four possible combinations.

As shown in Table 2.5, the test is too liberal when the independent variable is between dyads and the correlation is positive. This is perhaps the most prototypical case for the study of nonindependence. In this situation, the inferential statistic is too large, the  $p$  value too small, and too many Type I errors are made. The test is also too liberal when the independent variable is within dyads and the correlation is negative. There are two

cases in which the test of statistical significance is too conservative: when the independent variable is within dyads and the correlation is positive, and when the independent variable is between dyads and the correlation is negative.

Consider an example of heterosexual couples, half of whom are distressed and half of whom are not distressed. Two outcome variables are measured: One variable, marital satisfaction, likely has a positive correlation, and the other, how much housework each person does, likely has a negative correlation. If we test whether distress (a between-dyads variable) has an impact on these two outcome variables and treat individual as the unit of analysis, the test of whether distress affects marital satisfaction is likely to be overly liberal, and the test of whether distress affects housework is likely to be overly conservative.

Table 2.5 is very useful when the independent variable is between or within dyads. However, if it is mixed, the direction of bias is less certain. To determine the direction of bias, we need to compute the intraclass correlation for the independent variable. Consider gender as the independent variable. If gender were a between-dyads independent variable, in some of the dyads both members would be men, and in others both would be women. The intraclass correlation for a dummy-coded gender variable would be 1. If gender were a within-dyads independent variable, each of the dyads would consist of one man and one woman, and the intraclass correlation for gender would be  $-1$ . If gender were a mixed independent variable, some of the dyads would be same sex and others would be opposite sex. If most of the dyads were same sex, gender would be very much like a between-dyads variable and would have a positive intraclass correla-

**TABLE 2.5.** Effects of Nonindependence on Significance Testing

Independent variable	Correlation for the outcome variable	
	Positive	Negative
Between dyad	Too high <sup>a</sup>	Too low <sup>b</sup>
Within dyad	Too low <sup>c</sup>	Too high <sup>d</sup>

<sup>a</sup>Married couples assigned to conditions, with satisfaction the outcome variable.

<sup>b</sup>Married couples assigned to conditions, with amount of housework the outcome variable.

<sup>c</sup>Gender of person in married couples is the independent variable, and satisfaction is the outcome variable.

<sup>d</sup>Gender of person in married couples is the independent variable, and amount of housework is the outcome variable.

tion. If most of the dyads were opposite sex, then the intraclass correlation would be negative. However, if there were approximately the same number of male-male, female-female, and male-female dyads, the intraclass correlation would be near zero. When the intraclass correlation for the independent variable is near zero and the individual is the unit of analysis, tests of significance are generally neither very conservative nor very liberal.

Kenny and colleagues (1998) present the formulas for computing the degree of bias in  $p$  values created by nonindependence. For example, in dyadic research with a between-dyads predictor variable, if there were 50 dyads, or 100 individuals, in two treatment conditions and the intraclass correlation for the dependent variable were .45, then test statistics that normally would be associated with an alpha of .05 would actually indicate an alpha of .10 (Kenny et al., 1998, p. 238). Thus, in this case, mistakenly treating individual as the unit of analysis when there is nonindependence can result in substantial increases in Type I errors.

When the intraclass correlation is negative and the independent variable is between dyads, inferential tests are overly conservative. For example, again consider dyadic research with 100 individuals and two treatment conditions; given an intraclass correlation equal to  $-.50$ , the actual alpha would be only .006, not .05. There would also be a corresponding increase in the probability of Type II errors and a drop in power.

## WHAT NOT TO DO

If the data are determined to be nonindependent, what should a researcher do? We want to emphasize that the determination of whether there is nonindependence is in part an empirical question and in part a theoretical one. The theoretical aspects of the decision refer to the type of dyad being studied, the type of variable being studied, the research context, and whether previous researchers doing similar studies have found nonindependence. For example, if unacquainted people are randomly assigned to dyads and are asked to discuss an issue, their prior attitudes are likely independent, but their postdiscussion attitudes are probably nonindependent. An empirical analysis, as described in this chapter, would reveal whether there was any dependence.

The rest of this book considers how to handle nonindependence. However, we discuss in this section some strategies of data analysis and design that we see as less than optimal for dyadic data but that are nonetheless commonly used.

The first flawed strategy is just to ignore the nonindependence. As we discussed in the previous section, this strategy often results in biased significance tests. Sometimes the test is too liberal, and other times, too conservative. As long as the nonindependence is nontrivial, it is generally a mistake to ignore it.<sup>4</sup>

The second less-than-optimal strategy is to discard the data from one dyad member and analyze only one member's data. If there are no dyads and only individuals, then there is no nonindependence, and this is a "solution." Obviously, this strategy results in a loss of precision. Additionally, one would obtain different results if different data were discarded. Moreover, dyadic effects could not be measured.

The third less-than-optimal strategy is very much like the previous one. However, instead of discarding one person's data, data are collected from only one member of the dyad when they could have easily been collected from both. This strategy has the same drawbacks as the previous strategy.

The fourth flawed strategy is to treat the data as if they were two samples. This approach is often used when dyad members are distinguishable. For example, in a study of heterosexual couples, the data from men and women would be separately analyzed.<sup>5</sup> Although there are some advantages to this approach, it also has several key disadvantages. Most important, it presumes differences between genders (or whatever the distinguishing variable is) when in fact there may be no such differences. Additionally, power is likely lost by not combining results.

A fifth suboptimal strategy is not statistical but experimental. It is possible to prevent the nonindependence from occurring by ensuring that there is no social interaction between participants. There are two different ways in which this can be achieved. Participants interact not with another person but with a computer program that mimics a person, or they interact with an accomplice of the researcher. In some sense, this strategy takes the "social" out of social science. Duncan and Fiske (1977) discuss the difficulties of using accomplices in dyadic research.

Sometimes an alternative to using accomplices is to ask the participants themselves to alter their behavior. One participant may be asked to smile frequently during the interaction to determine the effect of smiling on how much a partner likes them. Alternatively, a variable can be manipulated (e.g., tell one participant that the other participant does not like him or her) in order to avoid the use of an experimental accomplice.

In essence, each of these strategies treats dyadic data as if they were individual data, not dyadic. Researchers need to confront the reality of

dyadic data and reject these individualistic strategies. Although dyadic data create complications in the analysis, such data also create opportunities. Moreover, in many cases, if there were independence of observations, we would be distressed because, as researchers in the area of relationships, we expect and even require nonindependence; when we do not find it, we are very disappointed. Rather than avoiding dyadic data, researchers need to embrace it and apply the methods that are presented in this book.

## POWER CONSIDERATIONS

Power, an important consideration in data analysis, refers to the probability of rejecting the null hypothesis when, in fact, it is false. Because the null hypothesis is virtually always false, power should be an important concern in the analysis. In dyadic analysis, there are two fundamentally different questions about power. The first question concerns the power of the test of nonindependence: If the data are nonindependent, will we be able to detect this nonindependence? The second question concerns the reduction or increase in power of the test of the independent variable that occurs when we change the unit of analysis from individual to dyad. That is, how much power in our key tests of independent variables is lost or gained when we use dyad as the unit of analysis rather than individual? In this chapter we consider the first topic. We consider the second topic in Chapters 3 and 7.

### Power of the Test of $r$

Discussing the interval level of measurement first, in Table 2.6, we present the power for the test of a Pearson correlation coefficient for various correlations and for different sample sizes. (The power of the tests of the intraclass correlation, either ANOVA or pairwise, is slightly less than that of the Pearson correlation.) We used Cohen (1988) to determine the power. We might note that Cohen considered .1 a small correlation, .3 a medium one, and .5 a large one. For instance, if we conduct a study with 80 pairs of siblings, and we think the population correlation is .3, then the power of such a test would be .78. Importantly, there is relatively little power in the tests of nonindependence when that nonindependence is small in size, even when we have as many as 200 dyads.

How many dyads are there in the typical study? From our survey of 75 studies, we examined the subset of those 25 that measured both mem-

**TABLE 2.6.** Power of the Test of Nonindependence (Pearson  $r$ ) given Alpha of .05 for a Two-Tailed Test

$n^a$	Absolute value of the population correlation				
	.1	.2	.3	.4	.5
10	.06	.08	.13	.21	.33
20	.07	.14	.25	.43	.64
40	.09	.24	.48	.74	.92
80	.14	.43	.78	.96	*
120	.19	.59	.92	*	*
160	.24	.72	.97	*	*
200	.29	.81	.99	*	*
300	.46	.97	*	*	*
500	.61	*	*	*	*
800	.81	*	*	*	*

\*Power greater than .995.

<sup>a</sup>Number of dyads.

bers of the dyad and used the standard design. For studies of this type, the number of dyads ranged from 25 to 411, with the median number across the 25 studies being 101. Because this value was obtained from publications in major journals, it is likely that the number of dyads in a typical research study would be somewhat smaller. Therefore, we would estimate the typical sample size as being 80 dyads. Using this number, we determined the power of the test for a small, medium, and large correlation. Cohen recommended setting power at .80, or an 80% chance of ruling out the null hypothesis. As shown in Table 2.6, there is typically not very much power in tests of correlations when the effect size is small and  $n$  equals 80. When the correlation is large, there does appear to be sufficient power. The minimal  $n$  for sufficient (80%) power to detect a small effect size is 783 dyads. We need to realize that enrolling large numbers of dyads into a study can be difficult; therefore, other strategies for increasing power besides increasing the sample size should be explored (McClelland, 2000).

Kenny and colleagues (1998) have defined *consequential nonindependence* as the level of nonindependence that would have to be present to result in the probability of committing a Type I error (i.e., alpha) being .10 when it should equal .05. The value for consequential nonindependence for dyads and a between-dyads independent variable is about .45. A corre-



lation of about .45 between dyad members' scores would result in the probability of making a Type I error of .10, if individual were the unit of analysis. The key question, then, is how many dyads are needed to have enough power (i.e., power of at least .80) to detect consequential nonindependence? If we assume a two-tailed test of nonindependence,<sup>6</sup> an alpha of .05, and a between-dyads independent variable, we would need to have at least 44 dyads. However, it is conventionally recommended (e.g., Myers, 1979) that the test for nonindependence be quite liberal and that a two-tailed alpha be .20. Using this value, the number of dyads needed to be able to test for consequential nonindependence is 28. We recommend having a minimum of 25 dyads before testing for nonindependence and adopting a liberal test value for alpha (.20) in the test of nonindependence. If there are fewer than 25 dyads (fortunately, a rare occurrence according to our survey of the literature), it becomes necessary to assume that there is nonindependence, even if the test of statistical significance of nonindependence is not statistically significant.

### Power of Kappa

Turning now to the nominal level of measurement, we suggest the following strategy for determining the power of the test that kappa is greater than zero. We return to the example in Table 2.4, but imagine that the values in that table are theoretical, not empirical. That is, a researcher plans to have an  $n$  of 200 and has specified the expected numbers for each of the nine cells. From this table, the hypothetical value of kappa and the two standard errors of  $s_{\kappa}$  and  $s_{\kappa}$  can be determined. The test statistic is  $\kappa\sqrt{n/s_{\kappa}}$  or  $Z_{\kappa}$ . Power is the probability that  $Z_{\kappa}$  is greater than 1.645 (assuming the conventional .05 level of statistical significance and one tail). We treat  $Z_{\kappa}$  as a normally distributed variable with a mean of  $\kappa\sqrt{n/s_{\kappa}}$  and a variance of  $s_{\kappa}^2/s_{\kappa}$ .<sup>2</sup> With this information we can use the standard normal distribution or the  $Z$  distribution to determine power.

Again treating the values in Table 2.4 as theoretical, not empirical, the value of  $\kappa$  is .4286,  $n = 200$ ,  $s_{\kappa} = 0.0555$ , and  $s_{\kappa} = 0.0537$ . Then the mean of that  $Z_{\kappa}$  is 48.82, and its standard deviation is 0.967. The probability that that  $Z_{\kappa}$  would be greater than 1.645 is 1. That is, for a sample size of 200 and a kappa of .4286, it is a virtual certainty that kappa will be statistically significant. However, if  $n$  were 40, then  $s_{\kappa} = 0.7851$ , and  $s_{\kappa} = 0.7596$ . The mean of that  $Z_{\kappa}$  is 3.452, and its standard deviation is 0.967. The probability that that  $Z_{\kappa}$  would be greater than 1.645 is .969.

### SUMMARY AND CONCLUSIONS

The primary focus of this chapter has been to provide methods for estimating and testing the degree to which two dyad members' scores are nonindependent. For distinguishable dyads, we saw that nonindependence for variables measured at the interval level can be estimated using standard methods such as the Pearson correlation coefficient. We discussed the intraclass correlation as the measure of nonindependence for interval level outcomes when the dyad members are indistinguishable. We considered several data-analytic approaches to estimating the intraclass correlation, including ANOVA, pairwise, and MLM solutions. With categorical outcomes, we suggested kappa as the measure of nonindependence for distinguishable dyads, and we proposed a variant of kappa for indistinguishable dyads.

Our discussion turned to several important issues involved in measuring and testing the degree of nonindependence. We defined negative nonindependence and discussed processes that can generate it. We also described methods for estimating nonindependence that control for the effects of independent variables.

The latter part of the chapter considered the consequences of ignoring nonindependence when conducting tests of the effects of independent variables. We showed that ignoring nonindependence can bias hypothesis-testing results in either the overly liberal or overly conservative direction, depending on the direction of nonindependence and the type of independent variable (e.g., between dyads or within dyads). We also discussed several less-than-optimal strategies for "controlling" nonindependence. Finally, we described a range of issues concerning the power of the test of nonindependence.

All too often nonindependence is viewed as a problem in the analysis of dyadic data. However, very often the nonindependence is what is most interesting about dyadic data. Consider two examples. Niederhoffer and Pennebaker (2002) were interested in the degree to which interaction partners matched each other's linguistic style, for example, using big words and prepositions and speaking tentatively. They found a matching of linguistic styles on many variables. However, O'Rourke and Cappellez (2003) examined whether marital aggrandizement, or an exceedingly positive portrayal, was similar in romantic couples. Contrary to what they predicted, they found no correlation of the two marital-aggrandizement measures once satisfaction was controlled.

In the next chapter, we consider hypotheses concerning means and regression coefficients when the independent variable is either between or within dyads. Later, in Chapter 7, we discuss mixed independent variables. We also return to the issue of measuring the similarity of responses of the two members of the dyad in Chapter 12. There, we develop a measure for each dyad, and we revisit the Pearson and intraclass correlation coefficients as measures of nonindependence. Finally, in Chapter 15, we review the conceptual meaning of nonindependence in dyadic research.

## NOTES

1. The testing of nonlinear nonindependence is straightforward when dyad members are distinguishable. Nonlinear terms for both persons (e.g., squared terms) would be created and then correlated. We know of no way to measure nonlinear nonindependence when dyad members are indistinguishable, because nonlinearity implies an asymmetry.
2. Assuming equal variances, if the means of the two members are exactly equal, the intraclass and the ordinary correlation coefficients have the same value only when the correlation is exactly +1 or -1.
3. Some have suggested using eta squared, or  $\eta^2$ , as a measure of a dyadic or group effect:  $SS_B/(SS_B + SS_W)$ . We do not think that this measure should be used, because it is very positively biased. For instance, if the true correlation is zero, eta squared equals about .5.
4. When the independent variable is mixed and its intraclass correlation is near zero, there is relatively little bias in the  $p$  value. However, as will be seen in Chapter 7, there are still advantages in considering the dyadic nature of the data.
5. Some might wonder about adopting the following strategy: Analyze the data of men and women separately, and then combine the results using meta-analytic techniques. Although there is some merit in such a strategy, the two sets of results are themselves nonindependent, and traditional meta-analytic methods would improperly treat them as if they were independent.
6. The careful reader may note that the minimum number of dyads needed differs from the number given in Table 6 in Kenny and colleagues (1998). The difference is that they assume a one-tailed test of the intraclass correlation, whereas we assume a two-tailed test.

## 3

# Analyzing Between- and Within-Dyads Independent Variables

In the previous chapter, we discussed how nonindependence biases significance test results (i.e.,  $p$  values) when individual is treated as the unit of analysis. In this chapter, we turn our attention to methods of analyzing the effects of either between- or within-dyads independent variables for the standard dyadic design, treating dyad as the unit of analysis (Kashy & Kenny, 2000). That is, we consider the cases in which the scores on the independent variable are the same for both dyad members (i.e., a between-dyads variable) or in which the scores sum to the same value for every dyad (i.e., a within-dyads variable). We also provide methods for assessing the degree of nonindependence in the outcome scores after partialing out the effects of the independent variables, as well as advice concerning the appropriate data-analytic strategy if the data appear to be independent. If there are missing data, then the methods presented in Chapter 7 should be used instead of the methods presented in this chapter.

In the first two sections of this chapter, we consider outcome or dependent variables that are measured at the interval level of measurement. The first section considers categorical independent or predictor variables, and the second section discusses predictor variables measured at the interval level of measurement. In the third section of this chapter, we consider outcome variables that are dichotomous. We assume throughout that there are a total of  $n$  dyads and  $2n$  persons measured, and we denote the intraclass correlation as  $r_i$ .



application to the dyadic context: A dyad can be viewed as a group with two members.

Turning to the specifics of the multilevel data structure, a basic multilevel model contains two levels. First, there is the lower level, or level 1; for the case of persons nested within groups, the lower-level unit is person. Second, there is the upper level, or level 2, and for persons in groups, the upper-level unit is group. (In a repeated-measures context, the lower level is observation and the upper level is person.)

Variables can also be characterized by level. A level-1 variable is a variable for which a score is obtained for each lower-level unit; some examples are student achievement scores in a classroom study and momentary stress in a repeated-measures experience sampling study. The outcome measure is always a level-1 variable in MLM. Variables can also be measured for the upper-level units. In a classroom study, a measure of the teacher's experience or ability is an example of a level-2 variable. Another level-2 variable would be class size. In a repeated-measures study, a level-2 variable might be a score on a personality scale or the person's ethnicity. Note that the score on the level-2 variable is the same for all of the level-1 units nested within the level-2 unit.

In illustrating the basic model, it is helpful to consider a hypothetical study of the effectiveness of two teaching methods. This study involves 60 classrooms, each with 10 children, 5 boys and 5 girls. Thirty of the classrooms have teachers who use a new experimental teaching method; teachers in the other 30 classrooms use the standard teaching method. At the conclusion of the academic marking period, the amount learned is measured for each student. Thus the level-1 unit is student, and the level-2 unit is classroom. Two level-1 variables are measured: student learning, which is the outcome, and gender of the student. The level-2 variable is teaching method (experimental vs. standard). We denote the outcome (learning) as  $Y$ , the level-1 predictor variable (gender) as  $X$ , and the level-2 predictor variable (teaching method) as  $Z$ .

We first consider a conventional ANOVA of the classroom data set. We then explain how that same data set can be analyzed using MLM. Finally, we adapt MLM to the analysis of dyadic data.

## MIXED-MODEL ANOVA

Using conventional ANOVA to analyze the data from the classroom study would result in a source table, as presented in Table 4.1. In the table, par-

## Using Multilevel Modeling to Study Dyads

Multilevel modeling (MLM), also commonly referred to as hierarchical linear modeling, is a relatively new statistical technique that is particularly useful for the analysis of dyadic data. In fact, many of the analyses discussed in this book can be accomplished only by using MLM, and almost every chapter in the remainder of this book includes sections detailing such analyses. This chapter is intended to provide an introduction to multilevel models in general and then to describe specifically how they can be applied to dyadic data. There are many issues involved in MLM that we do not venture into in this brief chapter, and we urge the reader to consult more detailed treatments of this topic (Bryk & Raudenbush, 1992; Hox, 2002; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Some of those more advanced issues are discussed in Chapters 10 and 13.

As the name suggests, in a multilevel data structure, there are multiple levels within the data. That is, there is a hierarchy of units, with one set of units nested within another. There are two classic cases that generate multilevel data. In one case, persons are nested in groups; examples might be students nested within classrooms or workers nested within work groups. In the other case, known as a *repeated-measures study*, each person is observed several times, and observations are nested within persons. Experience sampling or daily diary studies generate this second type of multilevel data. In this chapter, we focus our attention on the multilevel case in which persons are nested within groups, because it has a direct

**TABLE 4.1.** ANOVA Source Table and Corresponding Multilevel Parameters for the Group Data Example: The Effects of Gender and Teaching Method on Learning in Classrooms

Source	df	Error term	Multilevel parameter
Between groups	60		
Mean	1	G/Z	$a_0$
Condition (Z)	1	G/Z	$a_1$
Classroom (G/Z)	58	S/G/Z	$\sigma_d^2$
Children within groups	540		
Gender (X)	1	XG/Z	$c_0$
Gender by condition (XZ)	1	XG/Z	$c_1$
Gender by classroom (XG/Z)	58	S/XG/Z	$\sigma_f^2$
Error (S/XG/Z)	480		$\sigma_e^2$
Total	600		

Note. There are 600 children (300 boys and 300 girls), who are members of 60 classrooms.

participant gender is symbolized as X and experimental condition as Z; G represents classrooms or groups, and S represents participants. Presented in the table are the sources of variance, their degrees of freedom, and the error terms for the F-tests (the denominator of the F-ratio) that evaluate whether each effect differs significantly from zero. The MLM terms that correspond to each effect are presented in the last column of the table; these terms are introduced later in the chapter.

It is helpful to have an understanding of the different sources of variance. The between-group variation in Table 4.1 refers to the variation in the 60 classroom means, derived by averaging the students' learning scores over the 10 students within each classroom. This between-group variation can be partitioned into three sources: the grand mean, condition (Z), and classroom within condition (G/Z). The mean term represents how different the grand mean is from zero, which is a question not typically estimated or tested in traditional ANOVA. The condition variation measures the effect of the intervention—that is, the condition effect estimates the degree to which average learning scores are higher in classes that use the experimental teaching method relative to those that use the standard method. The third source of variation results from differences between classrooms that were in the same level of the treatment condition: Within the classrooms in which the teaching method was the same (either experimental or standard), were average learning scores higher in some classrooms than in others?

The children-within-groups variation refers to differences among children in each class: Do children within the same classroom differ in how much they learn? The gender effect (X) refers to whether boys or girls learn more on average, controlling for condition and classroom. The gender-by-condition interaction (XZ) refers to whether the effectiveness of the two teaching methods differs for boys or girls. The gender-by-classroom interaction (XG/Z) estimates the degree to which gender differences vary from classroom to classroom when controlling for condition. That is, the gender-by-classroom interaction assesses the degree to which the gender difference (i.e., the difference between the mean of the girls' scores minus the mean of the boys' scores) varies from group to group within the same treatment condition. Finally, there is variation due to child (S/XG/Z), and this source of variance measures variation in learning from child to child after controlling for gender, classroom, and condition.

Within this ANOVA model, there are three random effects: classroom (G/Z), gender  $\times$  classroom (XG/Z), and error (S/XG/Z). The classroom variance, symbolized as  $\sigma_d^2$ , for reasons that will become clear, measures the variation in average learning from classroom to classroom after controlling for condition. The gender  $\times$  classroom variance, symbolized as  $\sigma_f^2$ , measures the degree to which the gender difference in learning varies from classroom to classroom after controlling for condition. Denoting  $a$  as the number of levels of X ( $a = 2$  in this example) and  $b$  as the number of students within one level of X ( $b = 5$  in this example), the standard ANOVA estimates of these variances are given by

$$\text{classroom: } \sigma_d^2 = \frac{MS_{G/Z} - MS_{S/XG/Z}}{ab}, \quad (4.1)$$

$$\text{gender} \times \text{classroom: } \sigma_f^2 = \frac{MS_{XG/Z} - MS_{S/XG/Z}}{b}. \quad (4.2)$$

An exact estimate of the student variance cannot be obtained because it is confounded with error variance, and thus we represent the combination of student variance and error variance as  $\sigma_e^2$ . Finally, although it is not usually estimated, we can compute the covariance between classroom and gender  $\times$  classroom by computing the covariance between the mean amount learned by the group minus the amount learned for condition and the mean difference in the amount learned by boys and girls in each classroom. Such a covariance would represent the tendency for groups that learn more to have larger (or smaller) gender differences. Although this covariance is rarely estimated within ANOVA, the method does permit such a covariance.

Table 4.1 also presents the usual mixed-model error terms for each of the sources of variance. Tests of the fixed between-groups sources of variance, both the grand mean and the condition main effects, are conducted treating the  $MS_{GZ}$  as the error term. To test whether there are classroom differences,  $MS_{GZ}$  is divided by  $MS_{SGZ}$ . The error term for the fixed within-classroom effects, including the gender main effect and the gender-by-condition interaction, is  $MS_{XGZ}$ . Finally, the error term for  $MS_{XGZ}$  is  $MS_{SGZ}$ , which itself cannot be tested.

### MULTILEVEL-MODEL EQUATIONS

Estimation in multilevel models can be thought of as a two-step procedure. In the first step, a separate regression equation, in which  $Y$  is treated as the criterion variable that is predicted by the set of  $X$  variables, is estimated for each upper-level unit. In our classroom example, this involves computing a separate regression equation for each classroom. In the formulas that follow, the term  $i$  represents the upper-level unit (classroom or group), and  $j$  represents the lower-level unit (student). In the example, the first-step regression equation for student  $j$  in classroom  $i$  is

$$Y_{ij} = b_{0i} + b_{1i}X_{ij} + e_{ij}, \quad (4.3)$$

where  $b_{0i}$  represents the average learning for classroom  $i$ , and  $b_{1i}$  represents the coefficient for the relationship between learning and gender in group  $i$ . Assuming that we have used effect coding for gender (i.e., one gender is coded as 1 and the other as -1), the slope and the intercept are interpreted as follows:

- $b_{0i}$ : The average learning in classroom  $i$  across boys and girls.
- $b_{1i}$ : The difference between girls' and boys' mean learning, divided by 2, for classroom  $i$ .

The second step of the analysis involves treating the slopes and intercepts from the first-step analyses as outcome variables in two regressions. For these second-step analyses, the regression coefficients from the first step (see equation 4.3) are assumed to be a function of a group-level predictor variable  $Z$ :

$$b_{0i} = a_0 + a_1Z_i + d_i, \quad (4.4)$$

$$b_{1i} = c_0 + c_1Z_i + f_i. \quad (4.5)$$

There are two second-step regression equations, the first of which treats the first-step intercepts as a function of the  $Z$  variable and the second of which treats the first-step regression coefficients as a function of  $Z$ . So, in our example, the criterion variable in equation 4.4 is the average classroom learning, and it is modeled to be a function of the teaching method condition. The outcome variable in equation 4.5 is the gender difference in learning, and it is also modeled to be a function of the teaching method condition. In general, if there are  $p$  level-1 predictor variables (i.e.,  $X$  variables) and  $q$  level-2 predictor variables (i.e.,  $Z$  variables), and all possible fixed effects were estimated, there would be  $p + 1$  second-step regressions, each with  $q$  predictors and an intercept. There are then a maximum total of  $(p + 1)(q + 1)$  fixed-effect parameters in the set of second-step regressions.

Assuming that we have used effect coding for the teaching method condition variable (i.e., 1 = treated, -1 = control) and gender (1 = girl, -1 = boy), the parameters in equations 4.4 and 4.5 estimate the following effects:

- $a_0$ : Grand mean for learning across all students and classrooms.
- $a_1$ : Degree to which students in the experimental classrooms learned more than those in the control classrooms.
- $c_0$ : Degree to which girls learned more than boys.
- $c_1$ : Degree to which the effect of teaching method differs for boys and girls.

Table 4.2 presents a more general interpretation of these four parameters. For the intercepts ( $b_{0i}$ ,  $a_0$ , and  $c_0$ ) to be interpretable, both  $X$  and  $Z$  must be scaled so that either zero is meaningful or the grand mean of the variable is subtracted from each score (i.e., the  $X$  and  $Z$  variables are centered). As mentioned, in the example used here,  $X$  and  $Z$  (gender and experimental condition, respectively) are both effect-coded (-1, 1) categorical variables, and therefore zero can be thought of as an "average" across boys and girls and teaching method conditions.

In introducing MLM, we have thus far taken the sequential, two-step approach using two sets of equations, one set for level 1 and the other for level 2. An alternative approach is to combine all of the equations into a single equation. That combined equation can be obtained by substituting equations 4.4 and 4.5 directly into 4.3 to obtain

**TABLE 4.2.** Definition of Fixed Effects and Variance Components in the Multilevel Context<sup>a</sup>

Effect estimate	Parameter	Definition of effect
<b>Fixed effect</b>		
Constant	$a_0$	Average response on Y for persons scoring 0 on both X and Z
Level-2 predictor (Z)	$a_1$	Effect of Z on the average response on Y
Level-1 predictor (X)	$c_0$	Effect of X on Y for groups scoring 0 on Z
Interaction (X by Z)	$c_1$	Effect of Z on the effect of X on Y
<b>Variance</b>		
Group	$\sigma_d^2$	Group variation in average Y, controlling for X and Z
X by Group	$\sigma_f^2$	Group variation in the effects of X on Y, controlling for Z
Error	$\sigma_e^2$	Within-group variation, controlling for X (includes error variance)

$$Y_{ij} = a_0 + a_1 Z_i + d_i + (c_0 + c_1 Z_i + f_i) X_{ij} + e_{ij}, \quad (4.6)$$

and rearranging terms yields

$$Y_{ij} = (a_0 + d_i) + a_1 Z_i + (c_0 + f_i) X_{ij} + c_1 Z_i X_{ij} + e_{ij}. \quad (4.6)$$

The combined equation is not as complex as it might first appear to be. The intercept in this equation involves two components— $a_0$ , which is the fixed-effect piece that estimates the grand mean, and  $d_i$ , which is the random-effects piece that indicates that the intercept also varies from classroom to classroom. We then have the main effect of Z, teaching method, which is straightforward. Next we have the main effect of X, gender, which has an overall fixed-effect component,  $c_0$ , as well as a random effect,  $f_i$ , that varies from classroom to classroom. The last fixed effect is the gender-by-teaching-method interaction, or XZ. Finally, the last term in the model is error.

As was the case in the ANOVA discussion for balanced data, there are three random effects in multilevel models. First, there is the error component,  $e_{ij}$ , in the lower-level or first-step regressions (see equation 4.3 or 4.6). This error component represents variation in responses across the lower-level units after controlling for the effects of both the lower-level and upper-level predictor variables; its variance can be represented as  $\sigma_e^2$ .

In the example, this component represents variation in learning from student to student within classroom, controlling for both teaching method and gender.

There are also random effects in each of the two second-step regression equations. In equations 4.4 and 4.6, the random effect  $d_i$  represents variation in the intercepts that is not explained by Z. For the example,  $d_i$  represents variation in the classroom mean learning scores that is not explained by the teaching method variable. Note that  $d_i$ , in this context, is parallel to  $MS_{GZ}$  within the balanced repeated-measures ANOVA context, as shown in equation 4.1. The variance in  $d_i$  is a combination of  $\sigma_d^2$ , which was previously referred to as classroom variance, and  $\sigma_e^2$ . Finally, in equations 4.5 and 4.6, the random effect  $f_i$  represents the degree to which the size of the gender difference in learning varies from classroom to classroom. Note that  $f_i$  here is parallel to  $MS_{XGZ}$  within the ANOVA context, as shown in equation 4.2. The variance in  $f_i$  is a combination of  $\sigma_f^2$ , which was previously referred to as the gender-by-classroom variance, and  $\sigma_e^2$ . A more general description of these variances is given in Table 4.2.

The multilevel model, with its multistep regression approach, seems radically different from the ANOVA model. However, as we have pointed out, the seven parameters of this multilevel model correspond directly to the seven mean squares of the ANOVA model for balanced data. The multilevel approach has numerous advantages over ANOVA, including its ability to handle continuous, as well as categorical, predictor variables and its ability to handle unbalanced designs. Thus the multilevel model provides a more general and more flexible approach to analyzing multilevel data than that given by ANOVA.

## MULTILEVEL MODELING WITH MAXIMUM LIKELIHOOD

The two-step regression approach to MLM involves the repeated estimation of regression equations for each upper-level unit. That is, for the example, we would compute 60 regression equations, one for each classroom. The estimates derived from these regression analyses are then pooled at the second stage of estimation via a regression analysis across the level-2 units. In the example, this involves aggregating the results across the classrooms by predicting the step-1 results as a function of the teaching method condition variable.

Although it is possible to do a simple (and perhaps simplistic) analysis in which the first-step coefficients ( $b_{0i}$  and  $b_{1i}$ ) are literally the outcomes

in an ordinary regression at level 2—a solution known as ordinary least squares, or OLS—there are reasons why this may not be the optimal solution. An OLS approach treats all of the first-step regression coefficients as if they were of equal quality without consideration of such important factors as the sample size for each group and the amount of variation in  $X$  for each group. The issue is that the regression coefficients for larger groups or groups for which there is greater variance in  $X$  are estimated with greater precision than are those estimated for smaller groups or groups with lower variance in  $X$ .

Our example was rather basic in the sense that neither of these factors varied across classrooms: Every classroom had 10 children, and the distribution of  $X$ , or gender, was identical for each group (i.e., there were 5 boys and 5 girls in each classroom). In this special case of a balanced design, there is no variation in the precision of the first-step regression coefficients from group to group. If the first-step coefficients do vary in precision, however, factoring the quality or precision of those estimates into the second-step regressions (i.e., equations 4.4 and 4.5) is an important element of MLM (Kenny et al., 1998).

One common estimation strategy used for MLM is maximum likelihood estimation. With maximum likelihood, the two stages of estimation are done simultaneously in one step, and the solution is weighted so that the precision of the coefficients (e.g.,  $b_{0i}$ ,  $b_{1i}$ ) is taken into account. The maximum likelihood weights are a function of the standard errors and the variance of the term being estimated. For example, the weight given to a particular  $b_{0i}$  is a function of its standard error and the variance of  $d_i$ . Specialized stand-alone computer programs, such as HLM6 (Raudenbush, Bryk, Cheong, & Congdon, 2004), have been written that use maximum likelihood to derive estimates for multilevel data. Within major statistical packages, SAS's PROC MIXED and SPSS's Mixed Model can be used to estimate multilevel models.

It turns out that maximum likelihood estimation provides biased estimates of variances. For instance, the maximum likelihood estimate of an ordinary variance divides the sums of squares by  $N$ . We are more familiar with the unbiased estimator of the variance that uses  $N - 1$  in the denominator; such an estimate is called a restricted maximum likelihood estimate, or REML. Generally REML estimation is preferred over maximum likelihood, and REML is the default estimation method in most computer programs and packages.

With both maximum likelihood and REML, the log likelihood, denoted as  $L$ , is minimized. For a given model we can compute the model's deviance, or  $-2L$ . Although the deviance of a particular model cannot be

readily interpreted, the deviances of two nested models can be compared. This comparison is different depending on whether maximum likelihood or REML is used. Consider, first, maximum likelihood and the example that we have been discussing, with eight parameters, four fixed effects, three variances, and one covariance (between slope and intercept). If we removed one of the parameters from the model (i.e., set it to zero), we can difference the deviances of the two models to obtain a value that is distributed as a  $\chi^2$  with 1 degree of freedom. For REML, the  $\chi^2$  difference can be used only if the fixed parameters in the two models are the same. For example, the deviance of a model that has two fixed parameters (e.g., intercept and gender of student) and a random component (e.g., the effect of gender is random and varies by classroom) can be compared with the deviance of a model without the random component only if the same fixed effects (intercept and gender) are in the reduced model. Thus, with REML, the  $\chi^2$  difference test can be used to compare changes in only the random portion of the model, whereas maximum likelihood can be used for changes in the fixed, random, or combinations of both.

Additionally, the computer program MLwiN uses generalized least squares (GLS) estimation and not maximum likelihood. However, the results using GLS and restricted GLS, or RGLS, closely parallel the results using maximum likelihood and REML. Also, if RGLS is chosen for estimation, the GLS value is given for the deviance value. So, for the program MLwiN, the deviances can be subtracted regardless of which estimation method is chosen.

For both maximum likelihood and REML,  $t$ -tests are given for the fixed effects. To obtain the proper degrees of freedom, the Satterthwaite approximation should be used (Kenny, Bolger, & Kashy, 2002). We explain this complex approximation in more detail in Chapter 7. We note here that both SAS and SPSS use the approximation; the computer program HLM gives approximate degrees of freedom. Finally, the program MLwiN does not provide any degrees of freedom, and tests of the coefficients must be computed by dividing each estimate by its standard error and using a  $Z$  table to obtain an approximate  $p$  value.

## ADAPTATION OF MULTILEVEL MODELS TO DYADIC DATA

The adaptation from the person-within-group perspective to dyadic data is straightforward: Statistically, dyads can be viewed as groups composed of two persons. As with group data, with dyadic data the outcome variable,  $Y$ , is measured once for each person within the dyad, and so each dyad has

two scores on  $Y$ . Similarly, there can be level-1 predictor variables, the  $X$  variables, as well as level-2 predictor variables, the  $Z$  variables.

As an example, consider the data from a hypothetical study of 10 cohabiting heterosexual couples in Table 4.3. We realize that the example does not include enough dyads for meaningful estimates and significance tests, but by using a small data set the reader can easily replicate our results. In this study, researchers wanted to examine whether the contribution a person made to the household (a composite score based on financial contributions, as well as household labor) was associated with that person's predictions for the relationship's future—specifically, the likelihood of marriage within the next 5 years, measured as a percentage. The researchers were also interested in the degree to which there is cultural variation in this association, and so some of the cohabiting couples were Asian and others were American. Note that both the sex and culture vari-

**TABLE 4.3.** Hypothetical Dyadic Data Set for a Study Predicting Likelihood of Marriage

Dyad	Person	Future	Gender	Contribution	Culture
1	1	75	-1	-10	1
1	2	90	1	-5	1
2	1	55	-1	0	1
2	2	75	1	10	1
3	1	45	-1	-10	1
3	2	33	1	-15	1
4	1	70	-1	5	1
4	2	75	1	15	1
5	1	50	-1	0	1
5	2	40	1	-5	1
6	1	85	-1	-10	-1
6	2	90	1	20	-1
7	1	75	-1	-5	-1
7	2	80	1	0	-1
8	1	90	-1	5	-1
8	2	68	1	0	-1
9	1	65	-1	0	-1
9	2	78	1	15	-1
10	1	88	-1	-15	-1
10	2	95	1	5	-1

*Note.* For gender, women = -1 and men = 1; and for culture, American = 1 and Asian = -1. Contribution is a composite score and has been centered about the sample mean. The variable Future is a prediction of the likelihood of marriage in the next 5 years.

ables have been effect coded (1, -1) and that contribution has already been centered around its grand mean value for the entire sample so that interactions among the predictors can be readily interpreted.

With these data, we can specify a model that estimates several effects. We can estimate the grand mean for future predictions ( $a_0$ ). If we restrict ourselves to a single  $X$  variable—contribution—then we can examine whether individuals who make greater contributions to their relationships have more positive predictions for the future ( $c_0$ ). If we also include a  $Z$  variable—culture—we can test a main effect of whether predictions are more optimistic in one culture or the other ( $a_1$ ). Finally, we can estimate the interaction between  $X$  and  $Z$  to test whether the relationship between household contribution and predictions for the future differs across cultures ( $c_1$ ). Thus we can estimate the same four fixed effects as in the more general multilevel model with groups.

Application of MLM to dyadic data requires one major restriction that need not be made for groups with more than two members. In the most general multilevel model, the coefficients from the first-stage analyses (intercepts and slopes) are allowed to vary from group to group, but with dyadic data the slopes (i.e., the effect of  $X$  on  $Y$  for each dyad) must be constrained to be equal across all dyads. In other words, we must constrain the model to include only a fixed effect with respect to the effect of  $X$  on  $Y$ —the random component,  $f_i$ , is omitted from the model. The reason is that dyads do not have enough lower-level units (i.e., dyad members) to allow the slopes to vary from dyad to dyad. However, the intercepts for the dyads can vary, and it is through the variation of intercepts that the nonindependence in the members' scores within dyads is modeled. Thus a dyadic model with one  $X$  variable and one  $Z$  variable has four fixed effects and two random effects (variation in the intercepts and error variance).

One might assume that being unable to estimate a model with different slopes for each dyad biases the estimates of the multilevel model for dyads. It does not. Rather, what it does is confound the variance of the slopes with error variance. Tests of the null hypothesis are not biased by having differential slopes.

As mentioned, PROC MIXED in SAS can be used to estimate multilevel models. Singer (1998) gives an excellent overview of how various multilevel models can be specified using this procedure. In this chapter, we focus solely on the problem posed by dyadic data. We initially provide the statements used by SAS. Later, we discuss the use of SPSS, HLM, and MLwiN. All multilevel programs presume that data set is an individual or



pairwise data set. The SAS code to estimate the model in which CONTRIBUTION and CULTURE predict likelihood of marriage or FUTURE is

```
PROC MIXED COVTEST;
CLASS DYAD;
MODEL FUTURE = CONTRIB CULTURE CONTRIB*CULTURE
/SOLUTION DDFM=SATTERTH;
RANDOM INTERCEPT / SUBJECT = DYAD;
```

(Note that SAS requires a semicolon at the end of each statement and that the variable DYAD is a dyad identification number for which both partners have the same value.)

The default estimation method used by PROC MIXED is REML. If maximum likelihood estimation is desired, the optional statement METHOD = ML is added to the PROC MIXED line. The COVTEST option requests that Z tests of the random effects (e.g., the variance of the intercepts) be included in the output. The CLASS statement specifies the variable that identifies dyad membership (i.e., DYAD). FUTURE is the individual's outcome, or Y score; CONTRIB is the person's contribution score, or X; CULTURE is the person's Z score. Note that we have included the X-by-Z interaction in this model. The SOLUTION option in the MODEL statement requests that SAS print the estimates for the fixed effects:  $a_0$ , the intercept or grand mean;  $a_1$ , the effect of culture on future predictions;  $c_0$ , the effect of contribution on future predictions; and  $c_1$ , the interaction between culture and contribution. The DDFM = SATTERTH option requests the Satterthwaite (1946) approximation to determine the degrees of freedom for the intercept and slopes.

The final line of SAS code can take one of two forms. The first, as presented previously, is the RANDOM statement, which identifies the random-effect components of the model. In the dyadic model, only the INTERCEPT is random. The SUBJECT option identifies the level at which there is independence. Thus having SUBJECT = DYAD instructs the program that there is independence from dyad to dyad. This method of specifying the random effects has a significant disadvantage in dyadic research because it assumes that the nonindependence in the outcome scores is positive. That is, nonindependence is specified as a variance rather than as a correlation. Fortunately, there is an alternative:

```
PROC MIXED COVTEST;
CLASS DYAD;
MODEL FUTURE = CONTRIB CULTURE CONTRIB*CULTURE
/SOLUTION DDFM=SATTERTH;
REPEATED / TYPE=CS SUBJECT=DYAD;
```

The REPEATED statement treats the individual scores as repeated measures in the dyad, and CS implies what is called *compound symmetry*, which forces the degree of unexplained variance for the dyad members to be equal. Nonindependence is estimated as a correlation (more technically a covariance) and not as a variance.

Because the nonindependence in Y scores from Table 4.3 is positive, both methods of specifying the model yield the same results. The intercept (analogous to the grand mean in this example) for future predictions is  $a_0 = 71.83$ , indicating that across the 20 participants, the average estimate of likelihood of future marriage was slightly under 72%. The main effect of contribution was statistically significant, with  $c_0 = 0.84$ ,  $t(13.9) = 3.30$ ,  $p < .01$ , indicating that individuals who made greater financial/household-labor contributions to their relationships had more positive predictions for the future. More exactly, this coefficient estimates that with each 1-unit increase in contribution, the likelihood of future marriage increased by .84. The effect of culture, as well as the interaction between culture and contribution, were both marginally significant, with  $a_1 = -9.03$ ,  $t(7.66) = 2.02$ ,  $p < .08$ , and  $c_1 = 0.49$ ,  $t(13.9) = 1.91$ ,  $p < .08$ . Given the coding scheme used, the culture main effect indicates that Asian participants were generally more positive in their forecasts of the future than American participants. The interaction suggests that the relationship between contribution and future predictions is especially pronounced for American couples. The random-effect parameter for the model is the variance in the intercepts; it is estimated as 176.42 and is statistically significant,  $\chi^2(1) = 8.44$ ,  $p < .001$ . The deviance of the model is 142.93.

An alternative estimation strategy is to use SPSS. One needs version 11 or later of SPSS; we recommend using at least version 12.0. The SPSS syntax statements are as follows:

```
MIXED
  FUTURE WITH CONTRIB CULTURE
  /FIXED = CONTRIB CULTURE CONTRIB*CULTURE
  /PRINT = SOLUTION TESTCOV
  /RANDOM INTERCEPT | SUBJECT(DYAD) COVTYPE(VC).
```

(Note that SPSS requires a period after the last statement and "VC" stands for variance component.) The SPSS output is identical to the SAS output, with one exception. The  $p$  values for variances are two-tailed when they should be one-tailed. After dividing the  $p$  values in SPSS by 2, we obtain the same value as SAS. To allow for compound symmetry in SPSS, we remove the RANDOM statement and add the following statement:

```
/REPEATED = SEX | SUBJECT(DYAD) COVTYPE(CS).
```

Note that SPSS requires a repeated-measures variable (such as SEX in the example) in the REPEATED statement even if that variable is never used in analyses.

The code for HLM equations is entered as follows:

```
LEVEL 1 MODEL
FUTURE =  $\beta_0$  +  $\beta_1$ (CONTRIB) +  $\epsilon$ 
LEVEL 2 MODEL
 $\beta_0 = \gamma_{00} + \gamma_{01}(\text{CULTURE}) + u_0$ 
 $\beta_1 = \gamma_{10} + \gamma_{11}(\text{CULTURE}) + u_1$ 
```

Observe that the "+  $u_1$ " in the last line is lighter than the rest of the text to signify that the effect of CONTRIB is fixed; that is, it does not vary across dyads. In the output, the same model is expressed as

```
Level-1 Model
Y = B0 + B1*(CONTRIB) + R
Level-2 Model
B0 = G00 + G01*(CULTURE) + U0
B1 = G10 + G11*(CULTURE)
```

Finally, the estimates given by HLM6 are as follows:

Fixed Effect	Coefficient	Standard Error	T-ratio	d.f.	Approx. p-value
For INTRCPT1, B0					
INTRCPT2, G00	71.830862	4.469434	16.072	8	0.000
CULTURE, G01	-9.032863	4.469434	-2.021	8	0.077
For CONTRIB slope, B1					
INTRCPT2, G10	0.844758	0.255669	3.304	16	0.005
CULTURE, G11	0.487241	0.255669	1.906	16	0.074

If we compare the HLM estimates to the SAS and SPSS estimates, they are identical within the limits of rounding error. The estimates of the two variances and the deviance are also identical. The only difference is the degrees of freedom and their corresponding  $p$  values. As we discuss in Chapter 7, the degrees of freedom and  $p$  values given in SAS and SPSS are more appropriate.

Finally, we used MLwiN to estimate the parameters of the model. The equations for the model, including the estimated parameter values, are as follows:

```
FUTUREij ~ N(XB,  $\Omega$ )
FUTUREij =  $\beta_{0ij}$ INTERCPT + 0.845(0.256) CONTRIBij +
-9.033(4.470)CULTURj + 0.487(0.256)CONTRIBXCULij
 $\beta_{0ij} = 71.831(4.470) + u_{0j} + e_{0ij}$ 
 $[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [176.430(89.223)]$ 
 $[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [43.753(19.567)]$ 
-2*loglikelihood(IGLS) = 150.373(20 of 20 cases
in use)
```

The first line indicates that the outcome has a normal distribution. In parentheses next to each estimate is the standard error of the parameter. Notice that no degrees of freedom are given for the estimates. Notice also that the deviance, called IGLS by MLwiN, of -2\*loglikelihood is comparable to the maximum likelihood deviance, not the REML deviance. The estimates are RIGLS.

### Computing the Intraclass Correlation and Variance Explained with Multilevel Models

We begin by assuming the simplest possible multilevel model—one with no  $X$  or  $Z$  variables. The only fixed factor in this model is the intercept,  $a_0$ , which is equal to the grand mean of the observations. There are two random factors in this model, the first of which is the dyad covariance  $s_{dd}$ . Note that the dyad covariance is the same thing as the variance of the intercepts if the nonindependence in the outcome is positive. The second random factor is the error variance, or  $s_e^2$ . The estimate of the intraclass correlation can be computed by  $s_{dd}/(s_{dd} + s_e^2)$ . The SAS code to estimate this model is

```
PROC MIXED COVTEST;
CLASS DYAD;
MODEL FUTURE = /SOLUTION DDFM=SATTERTH;
REPEATED / TYPE=CS SUBJECT=DYAD;
```

Using the data in Table 4.3, the results of this analysis are  $s_{dd} = 255.45$  and  $s_e^2 = 82.30$ . The intraclass correlation is then  $255.45/(255.45 + 82.30)$ , which equals .76. Note that, with no missing data, this REML estimate is identical to the ANOVA intraclass correlation described in Chapter 2, and the maximum likelihood estimate is identical to the pairwise estimate of the intraclass correlation (Gonzalez & Griffin, 2001a). A key advantage that the MLM estimate has over the ANOVA approach is that ANOVA requires complete data (i.e., both dyad



members must supply scores to be included in the analysis), whereas with MLM we can use all the obtained data, even if scores from some dyad members are missing at random.

The comparable code for SPSS is

```
MIXED
  FUTURE          |
  /FIXED =        |
  /PRINT = SOLUTION TESTCOV
  /REPEATED = SEX | SUBJECT(DYAD) COVTYPE(CS).
```

(The statement "/FIXED =" looks wrong but it is not.)

More typically, there are predictor variables at both the upper and lower levels, as in our initial example with the data from Table 4.3. One can again compute the ratio of the dyad covariance (i.e., the variance in the intercepts) to the sum of the dyad covariance and the error variance. This ratio represents the proportion of variance due to dyads and is a partial intraclass correlation from which the effects of the predictor variables are partialled. For our model, we find  $s_{dd} = 176.42$  and  $s_e^2 = 43.76$ . The partial intraclass correlation is .80.

The units of measurement of the predictor variables are always an important consideration, but this is especially true in MLM. Normally, it is advisable to center all X and Z variables; that is, subtract the grand means from the variables.

Two types of centering are not advisable. The first type is called *group centering*: Here, a mean is computed for the dyad, and then it is subtracted from each member's score. Such a strategy removes all the variance due to dyad. The second type of inadvisable centering occurs with distinguishable dyads when the mean for each type of dyad member is subtracted from that type of individuals' scores. For instance, if the dyads were heterosexual couples, researchers might subtract the women's mean from each of the women's scores, and likewise for men. The problem with this type of centering is that it prevents researchers from examining the effects of the distinguishing variable on the outcome. Although grand mean centering affects the values of the coefficients, for dyads it has no effect<sup>3</sup> on  $s_{dd}$  and  $s_e^2$ .

One method of determining variance explained in a multilevel model is to compute  $s_{dd}$  and  $s_e^2$  for the model with the fixed variables included and then again with them excluded. The latter is sometimes called the unrestricted model. The degree of variance explained, sometimes called a *pseudo*  $R^2$ , is defined as

$$R^2 = 1 - \frac{s_{dd} + s_e^2}{s_{dd}' + s_e'^2},$$

where the prime signifies the estimates of variance and covariance from the unrestricted model (i.e., the model without any predictors). There is a chance that this  $R^2$  value might be estimated as a negative value. This would occur if the predictor variables were essentially unrelated to the outcome, so the values of  $s_{dd}$  and/or  $s_e^2$  might actually be larger than  $s_{dd}'$  and  $s_e'^2$ . If the  $R^2$  value were estimated to be negative, it would be reported as zero. For the example, the proportion of variance explained by the two predictor variables and their interaction is  $1 - (176.42 + 43.76)/(255.45 + 82.30)$ , or .348.

### Random Effects with Dyads

Typically any level-1 or X variable can be treated as a random variable in a multilevel model. That is, the effect of that variable can vary across level 2. Returning to the earlier classroom example, we can allow the gender effect to vary across classrooms such that in some classrooms girls perform better than boys, whereas in other classrooms boys perform better than girls. Normally, we can have as many random variables as we want, with the restriction that there must be more level-1 units within each level-2 unit than there are random variables. With dyadic data, there are only two level-1 units within each level-2 unit (i.e., two people in each dyad). Thus we can allow for only one random variable. Almost without exception, that one random effect is the intercept or constant. We cannot allow for any other variation. However, this does not bias our results, because that variation is properly included in the error variance. Although only one random effect can be estimated, models with dyads can still include the fixed effects from multiple X variables and multiple Z variables.

### Distinguishable Dyads

The preceding method of using multilevel models is particularly useful when members of the dyads are indistinguishable. It is not as useful when dyad members are distinguishable. There are several strategies for handling dyadic data within MLM when members are distinguishable. We discuss three different strategies, starting with the simplest. We want to mention that structural equation modeling (SEM), discussed in the next

chapter, tends to handle dyadic data with distinguishable dyads more simply.

The first strategy is identical to the one presented previously for handling indistinguishable dyads, but a dummy variable is added to code for the distinguishing variable. For the example data set, the dyad members are distinguishable by their gender. When we include gender in the model, we find that its effect is not statistically significant,  $b = -1.26$ ,  $t(7.77) = -0.70$ ,  $p = .50$ .

Additionally, we might wish to allow for interactions between the other variables in the model and the distinguishing variable. Ordinarily, we can accomplish this by multiplying the dummy variable for the distinguishing variable by each other level-1 variable in the model. (Remember that we cannot multiply the distinguishing variable by a level-2 variable because that variable is the same for both members of the dyad.) For the example, CONTRIBUTE is a level-1 variable, and we can examine its interaction with gender. We find that there is no effect of this interaction,  $b = -0.087$ ,  $t(7.56) = -0.33$ ,  $p = .75$ . This first strategy presumes that the variances are the same for both types of members, that is, homogeneity of variance.

The second strategy is an extension of the previous strategy but allows for heterogeneity of variance across levels of the distinguishing variable. For the example data set, that would mean allowing for different variances for men and women. If we were to use SAS, the data within dyad must be sorted by the distinguishable variable. Thus, for the example, the dyads would either all be sorted as "man then woman" or sorted as "woman then man." If there are missing cases, these cases would be added into the data set with the outcome variable coded as missing. Additionally, within SAS we would change the "REPEATED" statement to

```
REPEATED / TYPE=CSH SUBJECT=DYAD;
```

For SPSS, the comparable statement is

```
/REPEATED = GENDER | SUBJECT(DYAD) COVTYPE(CSH) .
```

The term CSH refers to *heterogeneous compound symmetry*. Note that "gender" is added to the statement in SPSS even though it need not be a variable in the model. It is included just to code for a repeated-measures variable, that is, to designate the two members as "1" and "2." Both HLM and MLwiN have a repeated-measures option that we do not illustrate.

We estimated such a model for the example data set using both SPSS and SAS, and we find that the error variance for women is 180.00 and for men is 253.77. Because the fixed variables are the same, we can use the difference between the REML deviance values from models with and without heterogeneous compound symmetry to evaluate whether in fact the variances are different. We find that they are not statistically different,  $\chi^2(1) = 0.68$ ,  $p = .59$ .

The third strategy for handling distinguishable dyad members, originally suggested by Raudenbush, Brennan, and Barnett (1995), is the most complicated. We call this model the *two-intercept model*. We estimate the following equation, for member  $j$  of dyad  $i$ :

$$Y_{ij} = a_i X_{1i} + b_i X_{2i},$$

where  $X_1$  is 1 for member 1 and 0 for member 2, whereas  $X_2$  is 0 for member 1 and 1 for member 2. (The correlation between  $X_1$  and  $X_2$  is  $-1$ .) Within the model, the effect of  $X_1$  and  $X_2$  are random variables (both  $a$  and  $b$  have an  $i$  subscript). Be aware also that in this model there is no intercept, at least not in the usual sense. In addition, there is no error term in the model, making this a very unusual model. The model does have a variance-covariance matrix of  $a_i$  and  $b_i$  with three elements: the variance of  $a_i$  or  $s_a^2$  (the error variance for the 1s), the variance of  $b_i$  or  $s_b^2$  (the error variance for the 2s), and the covariance between the two, or  $s_{ab}$  (the degree of nonindependence). We can test whether the two variances are equal and whether the covariance is statistically different from zero.

If there are any  $X$  or  $Z$  variables, they are added to the model, but any  $X$  variables need to be multiplied by each of the two  $X$  dummies. In that way, we can test whether the effect of the  $X$  variable is the same for the two types of members.

Because there is currently no way to force the error variance to zero within SAS or SPSS, the two-intercept model cannot be estimated using these programs. (We shall see in Chapter 7 that for both SAS and SPSS, there is a way to mimic the two-intercept model.) These models can, however, be estimated using MLwiN. Also, HLM6 can be used by setting the error variance to a very small value (e.g., 0.00001) or by using the multivariate option.

The results using MLwiN with RIGLS to estimate the two-intercept model for the example data are as follows:

$$\begin{aligned}
\text{FUTURE}_{ij} &\sim N(XB, \Omega) \\
\text{FUTURE}_{ij} &= \beta_{0ij}X1 + \beta_{1ij}X2 + 0.885(0.311) \text{CONTRIBUTE}_{ij} \\
&\quad + -9.667(4.418)\text{CULTUR}_{ij} + 0.459(0.275)\text{CONTRIB}\times\text{CUL}_{ij} \\
\beta_{0ij} &= 72.882(4.498) + u_{0j} \\
\beta_{1ij} &= 70.695(5.248) + u_{1j} \\
\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 189.886(84.916) \\ 180.181(90.805) \quad 263.265(117.735) \end{bmatrix} \\
-2 \times \log \text{likelihood(IGLS)} &= 149.638 \text{ (20 of 20 cases in use)}
\end{aligned}$$

These results are identical to the results obtained using the second strategy for handling distinguishing variables by allowing for heterogeneous variances.

## SUMMARY AND CONCLUSIONS

This chapter provides a brief introduction to MLM in general and MLM specifically in the dyadic context. We saw that this data-analytic approach can be used when data are hierarchically nested, with observations at a lower level (e.g., persons) being nested within an upper level (e.g., groups). We noted that this method is used when the outcome variable is measured for each lower-level unit but that independent or predictor variables can occur at either the lower level or the upper level.

Our introduction to MLM began with the balanced case, and we considered conventional mixed-model ANOVA as a data-analytic method. We then described the multilevel equations using the two-step procedure as a means for outlining the MLM technique. We defined the fixed and random effects that are generally estimated in MLM, and we discussed the restrictions required for applications of this method to dyadic data. Two estimation methods, maximum likelihood and REML, were briefly described, and implementation of MLM analyses using the data-analytic programs SAS, SPSS, HLM, and MLwiN was provided. Finally, we discussed the case of distinguishable dyads and described three estimation models that can accommodate such data. Thus, although MLM is particularly useful for studying dyad members that are indistinguishable, it can also be used when members are distinguishable.

Given the brevity of our discussion, we urge the interested reader to consult the books that we have mentioned. We also urge the reader to use

the sample data to reproduce the parameter estimates we reported in this chapter.

Many of the following chapters employ MLM. Chapter 7 is especially important because we recommend MLM as the best method for estimating the Actor-Partner Interdependence Model (APIM) when dyad members are indistinguishable. MLM is also useful for testing hypotheses about variances and covariances, as is discussed in Chapter 6. Chapter 10 uses MLM extensively in the analysis of the one-on-many design when partners (the many) are indistinguishable. For Social Relations Model (SRM) designs, MLM has been proposed as an estimation technique (Snijders & Kenny, 1999), but at the moment, this method is far too complicated to recommend for general use. As we stated in the introduction to this chapter, there are two types of multilevel data: persons within groups and repeated measures. In this chapter we focused on the within-groups (i.e., dyads) application. It is possible to combine the two types when dyads are measured over time, and this complex three-level model (dyads, persons, and times) is discussed in Chapters 13 and 14.

MLM is a very important tool for the estimation of dyadic models and will likely become increasingly popular. If one is serious about analyzing dyadic data, one should learn how to use it.

## NOTES

1. Sometimes models are called HLM for hierarchical linear modeling. Because HLM is also a computer program for multilevel modeling, the term "HLM" for multilevel modeling is not desirable.
2. Recent treatment of multilevel models (e.g., Raudenbush & Bryk, 2002) allows for crossing, as well as nesting, of units.
3. As seen in Chapters 10, 13, and 14, for models in which  $X$  variables are random, the units of measurement affect estimates of the variances and covariances. However, dyadic models discussed in Chapter 7 almost never contain random  $X$  variables.

# Analyzing Mixed Independent Variables

## The Actor-Partner Interdependence Model

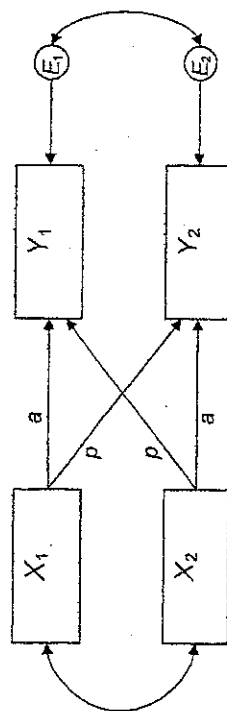
dyad mean,  $X_{ij} - M_j$ . The analysis of mixed independent variables allows researchers to investigate issues of mutual influence and is, therefore, a very important part of dyadic data analysis.

This chapter first introduces the model that we use to estimate the effects of a mixed independent variable. We focus on the two key components of that model, actor and partner, and consider their conceptual interpretation. Next, we present the statistical estimation of the model, first describing the analysis for indistinguishable dyads and second describing the analysis for distinguishable dyads. We consider three different estimation techniques: pooled regression, MLM, and SEM. For MLM we discuss four computer programs: SAS, SPSS, HLM, and MLwin. We also present several extensive examples. Then we discuss the measurement of effect sizes for mixed independent variables. Finally, we discuss specification error in the model.

### THE MODEL

As an example, consider the effects of depression on marital satisfaction. It may be that a wife's depression influences both her own and her husband's marital satisfaction. The effect of a wife's depression on her own marital satisfaction is called an *actor effect*, and the effect of her depression on her husband's satisfaction is called a *partner effect* (Kenny, 1996b). That is, an actor effect occurs when a person's score on a predictor variable affects that same person's score on an outcome variable; a partner effect occurs when a person's score on a predictor variable affects his or her partner's score on an outcome variable.

Figure 7.1 depicts the model that we call the *Actor-Partner Interdependence Model* (APIM). There are two dyad members and two variables,  $X$  and  $Y$ , for each. We denote the two  $X$  scores as  $X_1$  and  $X_2$  and the two  $Y$  scores as  $Y_1$  and  $Y_2$ . We assume that  $X$  causes, or is antecedent to,  $Y$ . A per-



**FIGURE 7.1.** APIM model where  $a$  is the actor effect and  $p$  is the partner effect.

People involved in dyadic relationships (or even brief dyadic interactions) can, and often do, influence each other's thoughts, emotions, and behaviors. This certainly occurs in romantic relationships, where the potential for mutual influence may be the quintessential feature of closeness in relationships (Kelley et al., 1983). Indeed, virtually all major theories of romantic relationships, including theories of equity (Messick & Crook, 1983; Walster, Walster, & Berscheid, 1978), commitment (Rusbult, 1980), trust (Rempel, Holmes, & Zanna, 1985), interdependence (Kelley & Thibaut, 1978; Thibaut & Kelley, 1959), and attachment (Bowlby, 1969, 1973, 1980) acknowledge the idea that one partner's attributes and behaviors can affect the other partner's outcomes. Mutual influence is also present in other types of dyadic relationships, such as friendships and work relationships.

In Chapter 3, we described analyses appropriate for independent or predictor variables that vary either only between dyads or only within dyads. This chapter focuses on the analysis of independent variables that are mixed in nature (see Chapter 1). Mixed independent variables vary both between and within dyads; they can vary on average from dyad to dyad, and they can vary from person to person within each dyad. Thus two sources of variation contribute to variation in mixed variables: The between-dyad source is characterized by variation in the dyad means,  $M_j$ , and the within-dyads source is characterized by variation in the deviations of each individual's score from the

son's own  $X$  affects his or her own  $Y$ , and the effects of  $X_1$  on  $Y_1$  and of  $X_2$  on  $Y_2$  are called *actor effects*. Actor effects are symbolized by  $a$  in our discussion and in Figure 7.1. In addition, a person's  $X$  affects his or her partner's  $Y$ , and the effects of  $X_1$  on  $Y_2$  and of  $X_2$  on  $Y_1$  are called *partner effects*. Partner effects are symbolized by  $p$ . If dyad members are distinguishable, there are potentially two actor effects—one for person 1, or  $a_1$ , and one for person 2, or  $a_2$ . There are also potentially two partner effects, one from person 1 to person 2, or  $p_{21}$ , and one from person 2 to person 1, or  $p_{12}$ . Because partner effects are interpersonal, they are given double subscripts, and we follow the usual convention for regression coefficients of listing the effect first and then the cause. Note that an actor effect is like an intrapersonal effect (see Chapter 6) in that it refers to one person and a partner effect is like an interpersonal effect in that it refers to two persons.

There are two correlations in the model. The two  $X$ 's might be correlated, represented by the curved line on the left. Such a correlation might be due to a compositional effect (see Chapter 1). The second correlation is the residual nonindependence in the outcome scores ( $Y$ 's), which is represented by the correlation between  $E_1$  and  $E_2$ . This represents the nonindependence not explained by the APIM. If we denote the correlation between  $X_1$  and  $X_2$  as  $r$ , then that part of the correlation between  $Y_1$  and  $Y_2$  that can be explained by actor and partner effects can be shown to equal

$$a_1p_{21} + a_2p_{12} + r(a_1a_2 + p_{12}p_{21}).$$

[In the case in which actor and partner effects are equal, with  $a_1 = a_2 = a$  and  $p_{12} = p_{21} = p$ , the formula becomes  $2ap + r(a^2 + p^2)$ .] Note that there are four sources of correlation. The first two involve the combination of actor and partner effects. The effect of  $X_1$  on  $Y_1$  and  $Y_2$  is represented by  $a_1p_{21}$ , and the effect of  $X_2$  on  $Y_1$  and  $Y_2$  is represented by  $a_2p_{12}$ . These are two sources of nonindependence. The impact of the next two sources depends on the size of the correlation between  $X_1$  and  $X_2$ , or  $r$ .

The analysis implied by the APIM can be used to estimate actor and partner effects for both dyadic and group data (see Kenny, Mannetti, Pietro, Livi, & Kashy, 2002, for a discussion of the application of the APIM to groups) when the independent variable is mixed, and it allows either categorical or continuous independent variables. It can also incorporate independent variables that are not mixed (i.e., those that vary only within dyads or only between dyads); however, separation of actor and partner effects can occur only with mixed predictor variables. Notice that when  $X$  is a between-dyads variable, the correlation between  $X_1$  and  $X_2$  is +1, and

when  $X$  is a within-dyads variable, the correlation between  $X_1$  and  $X_2$  is -1. In either case, there is perfect multicollinearity, and only one effect, not two, can be estimated.

The APIM is increasingly being used in the social sciences. For example, investigators have used it in researching such varied topics as emotion (Butler et al., 2003), health (Butterfield & Lewis, 2002; Franks, Wendorf, Gonzalez, & Ketterer, 2004), leisure activities (Berg, Trost, Schneider, & Allison, 2001), communication competence (Lakey & Canary, 2002), personality (Robins, Caspi, & Moffitt, 2000), commitment (Kurdek, 1997a, 1997b, 2000), interpersonal perception (Kenny & Acitelli, 2001), relationship violence (Landolt & Dutton, 1997; Moffitt, Robins, & Caspi, 2001), social influence (Ortña, Wood, & Simpson, 2002), and attachment style (Campbell, Simpson, Kashy, & Rholes, 2001). The model has also been recommended for use in the study of families (Rayens & Svavardottir, 2003) and small groups (Bonito, 2002).

## CONCEPTUAL INTERPRETATION OF ACTOR AND PARTNER EFFECTS

Many researchers are most comfortable with intraindividual or psychological explanations of behavior. This may partially account for the fact that actor effects are routinely estimated, whereas partner effects are often ignored. By studying only actor effects, researchers focus on the individual level of analysis. However, when partner effects are included, identifying truly relational phenomena becomes a possibility. In fact, the presence of partner effects implies that something relational has occurred, because a person's response depends on some characteristic of his or her partner.

### The Relative Size of Actor and Partner Effects

Although the sizes of actor and partner effects are independent in principle, specific combinations of these effects are particularly relevant to the study of couples. Following Kenny and Cook (1999), four models have important implications in relationships research:

- Actor-oriented:  $a \neq 0, p = 0$
- Partner-oriented:  $a = 0, p \neq 0$
- Couple-oriented:  $a = p$
- Social comparison:  $a + p = 0$

(The reader should consult Kelley & Thibaut, 1978, for a parallel formulation.)

In the actor-oriented model, a person's outcomes are a function of that person's characteristics only; the partner's characteristics have no impact. Most researchers who adopt the actor-oriented perspective assume that there are actor effects but no partner effects. They rarely bother to test that assumption. This approach to dyadic data occurs frequently with studies of heterosexual dating or married couples in which separate analyses are conducted for men and women (Kashy, Campbell, & Harris, 2006). We suggest that even if a researcher believes that a process is individualistic, it is still necessary to estimate partner effects to show that they are zero. For instance, Rusbult's (1983) social exchange model is individualistic, and when Bui, Peplau, and Hill (1996) looked for partner effects, they found very few. Thus the individualistic perspective of the Rusbult model was validated. Moreover, if there were unestimated partner effects, the estimated actor effects would be biased.

In the partner-oriented model, a person is affected by his or her partner's score on  $X$  but is not affected by his or her own  $X$  score. One plausible example for which there could be partner effects and not actor effects is the effects of physical attractiveness on relationship satisfaction. It could very well be the case that a person's own attractiveness has little association with his or her own relationship satisfaction, but the person's partner's attractiveness could be strongly related to the person's satisfaction. Although such a partner-effect-only model is plausible, it is still helpful to include actor effects. The physical attractiveness of two members of a couple tends to be correlated (Feingold, 1988), and the curved line between  $X_1$  and  $X_2$  in Figure 7.1 represents this correlation. If there are actor effects that are ignored, then partner effects would be overestimated. Accordingly, even if partner effects are of primary interest, controlling for actor effects in the analysis is still necessary.

In the couple-oriented model, the actor and partner effects are equal such that the person is affected as much by his or her own  $X$  as by his or her partner's  $X$ . This pattern would occur if the person were as concerned with the partner's outcomes as with his or her own outcomes. For example, playing one's best in a tennis match may make one satisfied with the outcome of the game, and knowing that one's spouse played his or her best could also lead to satisfaction. When one's spouse has a bad game, one's own satisfaction may be less, independent of how one played. Such an orientation would be characteristic of communal relationships (Clark & Mills, 1979).

It is helpful to examine the equation when actor and partner effects are equal. Consider the model:

$$Y_1 = aX_1 + pX_2 + E_1.$$

If actor and partner effects are equal ( $a = p$ ), then we can rewrite the equation as

$$Y_1 = 2a(X_1 + X_2)/2 + E_1.$$

Thus the sum or average of the  $X$  variables—that is,  $(X_1 + X_2)/2$ —can be used to predict  $Y_1$  instead of both  $X_1$  and  $X_2$ . This analysis then implies that researchers who use the dyad sum or average of the independent variable to predict a person's outcomes are implicitly assuming a couple orientation. The APIM permits an empirical test of the validity of using a dyad sum: equal actor and partner effects.

It may happen that one member of the couple is couple oriented and the other is not. For example, in a tennis match between a parent and a child, the parent may be couple oriented, feeling satisfaction with either his or her own game and with the child's game. In other words, both the actor and the partner effect are positive for the parent. However, the child may not be couple oriented, feeling less satisfied the better the parent plays. Thus, for the child, the actor effect (i.e., playing well) would be a positive predictor of satisfaction, and the partner effect (i.e., the parent's playing well) would be a negative predictor of satisfaction. A competitive orientation such as this leads to the next pattern.

In the social comparison model, the actor and partner effects are relatively equal in absolute magnitude and have opposite signs. Usually, the actor effect is positive, and the partner effect is negative. Here the person implicitly or explicitly compares him- or herself with the partner. In contrast to the couple-oriented case, in which the partner's success is valued as much as one's own outcome, the social comparison orientation typically involves dissatisfaction with the partner's success. Both imply couple effects, but their conceptual meanings are totally opposite.

### Partner-Oriented Interaction Effects

If there were partner effects, then there would be evidence that the two persons are part of an interdependent system. Conversely, if there were no interdependence, there would be no partner effect. It follows logically that



the greater the interdependence, the greater the partner effect—or, in other words, that the degree of interdependence moderates the size of partner effects. Therefore, the more important the partner is to someone (consciously or unconsciously), the more he or she would be affected by the partner's characteristics. Hence, researchers can validate a measure of relationship closeness by demonstrating that it interacts with a partner score in producing a particular outcome.

Such an analysis suggests that a relationship scale's "neutral point" or "zero point" can be empirically defined. This neutral point is the predicted score on the relationship scale for individuals who are not interdependent on one another. Imagine a dyadic study of the effects of two partners' happiness on their own and their partner's health. Imagine further that a measure of relationship closeness interacts with the partner effect for feelings of happiness. We can compute at what value of closeness the partner effect for happiness is zero, and that value would be the neutral point of the closeness measure. Those scoring at the neutral point on the measure of relationship closeness would tend to be actor oriented (i.e., the partner's happiness is irrelevant in determining the actor's health); those scoring above the neutral point on the measure of relationship closeness would tend to be couple oriented (i.e., the partner's happiness increases the actor's health); and those scoring below the neutral point would be guided by social comparison (i.e., the partner's happiness decreases the actor's health, or "I do best when you are miserable").

### Actor-Partner Interactions

The major research question in dyadic research often involves actor-partner interactions. Interaction variables are typically operationalized as the product of two independent variables. Thus, for couples, the actor-by-partner interaction term would be the product of  $X_1$  and  $X_2$ . However, multiplying is just one of an infinite number of ways to specify an interaction between two continuous variables. For instance, consider a study of the effect of personality similarity on relationship outcomes. Personality similarity is typically operationalized as the absolute difference between the person's own personality ( $X_1$ ) and the partner's personality ( $X_2$ ). It can be viewed as an interaction of actor and partner effects. If  $X$  is a dichotomy, then the absolute difference and the product are statistically the same (i.e., the absolute difference and the product correlate perfectly with each other).

Still other specifications of actor-partner interactions are possible and often reasonable. For example, a researcher may speculate that relation-

ships need only one member with a certain skill or attribute in order for an outcome to occur; similarly, a deficit (e.g., substance abuse) on the part of only one member can precipitate some negative consequence. In such cases, it makes sense to use only the higher or lower of the two members'  $X$  scores as the interaction term. Consider the effects of interpersonal competence on relationship outcomes. If the more competent member of the couple is able to compensate for the deficits of the less competent member, then relationships may only need one interpersonally competent member in order for both members to be satisfied. In this case, the interaction variable should be operationalized as the score of the member with the higher score. In some couples, this will be the  $X_1$  score, and in other couples, the  $X_2$  score, and so it is not simply a main effect (unless  $X_1$  is always greater or less than  $X_2$ ). If dyad members are distinguishable, then the maximum or minimum measure of interaction should not be used if one member's value of  $X$  is always greater than the other member's (e.g., if the woman's  $X$  is always greater than the man's  $X$ ).

Although there are many ways to operationalize actor-partner interactions, empirically it is very difficult to distinguish among them. For instance, the multiplicative operationalization of interaction and the discrepancy measure are usually highly correlated. Generally, theory, and not statistical analysis, must be used to choose the appropriate operationalization. In addition, one should always interpret the actor-partner interaction using the measure that was operationalized. That is, if a product is used, one should not interpret it as a similarity measure.

Whenever interactive effects are estimated, the main effects should usually be controlled. (For a contrasting view, see Brauer & Judd, 2000.) For instance, if a discrepancy score is computed to form a similarity score, the components that make up that discrepancy should also be included in the analysis. All too often, interaction scores are computed, but the main effects of actor and partner are not controlled. Frequently, "interactions" (e.g., similarity) are statistically significant because the confounding effects of actor and partner have not been removed. For example, suppose a study indicates that couples with similar levels of depression are more satisfied, but the main effects are not controlled. The similarity effect may actually be due to actor and partner effects of depression. That is, most people are not depressed, and so if two people are "similar," then it is likely that they are *both* not depressed. However, if one member is depressed, it is likely that the absolute difference in the two depression scores would be large and that the dyad members would be "dissimilar." Thus couple similarity is confounded with individual levels of depression, and these effects should be controlled.



### ESTIMATION OF THE APIM: INDISTINGUISHABLE DYAD MEMBERS

We present three different methods to estimate the APIM: the pooled regression method, MLM, and SEM. We do not discuss the pioneering estimation method presented by Kraemer and Jacklin (1979). Theirs was the first explicit attempt to estimate the model that came to be known as APIM. However, their technique is limited to the case in which there is a single dichotomous independent variable ( $X$ ) and tests of statistical significance are approximate. Also, we do not discuss pairwise estimation of the model, and we refer the reader to Gonzalez and Griffin (2001a).

Although not required, it is advisable to make zero a meaningful value for the predictor variables. This is often done by centering the  $X$  variable (i.e., subtracting the mean). In our first example, the key predictor variable is the number of hours per week each of two roommates spend doing housework. Because zero is a meaningful value (i.e., the person does no housework), we do not center this variable. In our second example, however, the key predictor variable is a measure of neuroticism, which does not have a meaningful zero value. Thus, for the second example data set, we computed the grand mean of  $X_1$  and  $X_2$  combined and subtracted it from all of the  $X$  scores.

### Pooled-Regression Method

The pooled-regression approach to the APIM involves estimating two regression equations and then pooling the results together to estimate the APIM parameters (Kashy & Kenny, 2000). One of the regression equations focuses on the within-dyads effects of the mixed independent variable (referred to as the *within-dyads regression*), and the other regression equation focuses on the between-dyads effects of the mixed independent variable (referred to as the *between-dyads regression*).

#### Analysis with a Single Predictor Variable

As a very simple example, consider a fictitious study of housing satisfaction involving 20 pairs of same-sex roommates. Say that one predictor variable under consideration is a measure of the number of hours each individual spends per week cleaning and maintaining the residence. The fictitious data for this example are presented in Table 7.1. As can be seen in this table, for the first pair of roommates, the first roommate spends a

little over 1 hour per week on housework and has a satisfaction score of 6 (on a 9-point scale). The other roommate works about half as much (0.6 of an hour) and has a satisfaction score of 7. Note that this is a dyad-level data set in which each record is one dyad (see Chapter 1).

To estimate actor and partner effects, two regression equations are computed. The within-dyads regression involves predicting the difference between the dyad members' scores on the outcome variable ( $Y_1 - Y_2$ ) with the difference between the dyad members' scores on the mixed predictor variable ( $X_1 - X_2$ ). The between-dyads regression involves predicting the dyad mean on outcome variable  $[(Y_1 + Y_2)/2]$  with the dyad mean on the mixed predictor variable  $[(X_1 + X_2)/2]$ . The difference in  $X$  predicts the difference in  $Y$ , and the mean of  $X$  predicts the mean of  $Y$ .

Thus, in the within-dyads regression, the data would be derived by subtracting one roommate's score on satisfaction from the other roommate's score on satisfaction. This difference between the two partners' scores on satisfaction (roommate 1's satisfaction score minus roommate 2's satisfaction score) serves as the outcome score in a regression in which the predictor score is the difference (subtracting in the same direction across the two dyad members: person 1 minus person 2) between the two partners' housework scores. Using the data in Table 7.1, for the first three dyads, the outcome and predictor scores, respectively, would be, for Dyad 1, 0.6, -1; for Dyad 2, 2.0, -4; and for Dyad 3, -0.2, -1.

It is important to note that the intercept should not be estimated in the within-dyads regression. We discussed this in Chapter 3 when we

TABLE 7.1. Data for Fictitious Study of Roommates

Dyad	H		S		G		H		S		G	
	$R_1$	$R_2$	$R_1$	$R_2$	$R_1$	$R_2$	$R_1$	$R_2$	$R_1$	$R_2$	$R_1$	$R_2$
1	1.2	0.6	6	7	1	1	11	11	8	5	-1	-1
2	4.3	2.3	5	9	-1	-1	12	12	5	7	-1	-1
3	0.4	0.6	3	4	1	1	13	13	4	6	1	1
4	0.3	0.5	6	8	1	1	14	14	5	6	7	1
5	3.2	1.0	2	6	-1	-1	15	15	5	6	1	1
6	1.1	0.8	3	8	1	1	16	16	7	4	-1	-1
7	2.0	1.5	5	7	1	1	17	17	2	4	-1	-1
8	0.4	1.6	7	4	-1	-1	18	18	4	4	-1	-1
9	0.3	0.5	3	3	1	1	19	19	6	5	-1	-1
10	3.2	2.0	5	8	1	1	20	20	0.5	3	4	-1

Note. H, housework; S, satisfaction; G, gender;  $R_1$ , respondent 1;  $R_2$ , respondent 2. For gender, men = 1 and women = -1.

described the analysis of within-dyads independent variables. The issue is that when we compute the difference between the dyad members' scores, the direction of differencing is arbitrary in the sense that we can either subtract Joan's score from Sue's or vice versa. Because of this arbitrariness of direction, the intercept should not be estimated in the within-dyads regression. As we described in Chapter 3, most computer packages have a no-intercept option for regression analyses. For example, in SPSS, within "Linear Regression," the "Options" box is checked, and then the check is deleted from the box "Include the constant in the equation." For SAS, this is the NOINT option in PROC GLM or PROC REG. Forcing the intercept to be zero ensures that the regression solution is the same even if the direction of differencing was switched for some dyads. Accordingly, the within-dyads equation for dyad  $i$  is

$$Y_{1i} - Y_{2i} = b_w(X_{1i} - X_{2i}) + E_{wi}.$$

Note that there is no intercept in the equation.

In the between-dyads regression, the average of the two satisfaction scores is computed for each dyad, as is the average of the two housework scores. The mean housework score for the dyad is then used to predict the mean satisfaction score for the dyad. Again, using the data in Table 7.1, for the first three dyads the outcome and predictor scores, respectively, for the between-dyads regression would be, for Dyad 1, 0.9, 6.5; for Dyad 2, 3.3, 7.0; and for Dyad 3, 0.5, 3.5. The no-intercept option is not used in this part of the analysis, and the intercept is estimated. The equation for dyad  $i$  would be as follows:

$$(Y_{1i} + Y_{2i})/2 = b_0 + b_b(X_{1i} + X_{2i})/2 + E_{bi}.$$

The unstandardized regression coefficients derived from these two regressions,  $b_b$  from the between regression and  $b_w$  from the within regression, are then used to estimate actor and partner effects as follows (Kenny, 1996b):

Actor:  $(b_b + b_w)/2$

Partner:  $(b_b - b_w)/2$

For the housework and satisfaction data in Table 7.1, the between regression equation yields a  $b_b$  of 0.297, and the within regression yields a  $b_w$  of -1.479, and thus the actor effect is estimated to be -0.591 and the partner

ner effect, 0.888. Both the actor and partner effect estimates can be interpreted as unstandardized regression coefficients. The actor effect in this example indicates that each 1-hour increase in housework done by an individual yields a drop in housing satisfaction of 0.591 points—people who do more housework are less satisfied. The partner effect estimate indicates that each 1-hour increase in housework done by a person's roommate corresponds to a 0.888 increase in the person's satisfaction—so people whose roommates do more housework are more satisfied.

To test whether the actor and partner effects differ significantly from zero, the standard errors associated with the between and within regression coefficients ( $s_b = 0.347$  and  $s_w = 0.359$ , respectively) must be pooled. (The standard error for each of these regression coefficients can be derived by taking the  $t$  value associated with the regression coefficient and dividing it by the regression coefficient.) The formula for calculating the pooled standard error is

$$\text{Pooled standard error} = \sqrt{\frac{s_b^2 + s_w^2}{4}}.$$

For the example of roommate housing satisfaction, the pooled standard error is estimated as 0.250. The estimate of the actor effect is divided by this pooled standard error to yield a  $t$ -test indicating whether the actor effect differs significantly from zero. In the example, the  $t$  value for the actor effect is  $-0.591/0.250 = -2.36$ . Similarly, the partner-effect  $t$  value is  $0.888/0.250 = 3.55$ . The degrees of freedom (Satterthwaite, 1946) for both of these tests are estimated as

$$df = \frac{(s_b^2 + s_w^2)^2}{\frac{s_b^4}{df_b} + \frac{s_w^4}{df_w}}.$$

Surprisingly, the degrees of freedom can be fractional. For the  $t$  distribution, critical values and exact  $p$  values can be determined for fractional values of degrees of freedom. If a fractional answer is obtained and a critical value is to be sought from a  $t$  table, to be conservative, one rounds down. For the example, the degrees of freedom are estimated to be 37.07, and so both the actor and partner effects are statistically significant.

An additional relevant statistic in this analysis is the amount of variance explained by the actor and partner effects combined. To compute this statistic, the total variance that can potentially be explained by the actor and partner effects must be computed. To compute this total variance, two additional regressions are conducted: a between-dyads regression without

the averaged mixed predictor variable and a within-dyads regression without the differenced mixed predictor variable. As before, the within-dyads regression should be estimated without the intercept. For the data in Table 7.1 from our fictitious roommate study, this involves computing a between-dyads regression in which average satisfaction is predicted only by an intercept and a within-dyads regression in which the difference in satisfaction is the criterion and in which there are no predictors (again suppressing the intercept).<sup>4</sup>

The between regression without the mixed predictor yields a sum of squares error,  $SSE_b$ , and a mean square error,  $MSE_b$ ; for the example, these values are 119.00 and 5.95, respectively. The within-dyads regression without the mixed predictor also yields a sum of squares error,  $SSE_w$ , and a mean square error,  $MSE_w$ ; 33.24 and 1.75, respectively. The total variance that can be explained is then  $(MSE_b' + MSE_w')/2$ , or 3.85. The total variance left unexplained by the mixed independent variable is computed from the regressions that include the mixed variable as a predictor and equals  $(MSE_b + MSE_w)/2$ , where  $MSE_b$  is the mean square error from the between regression that included the averaged mixed independent variable as a predictor and  $MSE_w$  is the mean square error from the within-dyads regression that included the differenced mixed independent variable as the predictor. The variance left unexplained by housework is  $(1.77 + 3.31)/2 = 2.54$ . To compute the proportion of variance explained by the actor and partner effects combined, the value of  $R^2$  is

$$R^2 = 1 - \frac{SSE_b + SSE_w}{SSE_b' + SSE_w'}$$

where  $SSE_b$  is the sum of squares error from the between regression that included the averaged mixed independent variable and  $SSE_w$  is the sum of squares error from the within regression that included the differenced mixed independent variable. (Recall that a mean square is defined as a sum of squares divided by its degrees of freedom.) For the example,  $R^2 = .378$ , and so the actor and partner effects for housework combined account for almost 38% of the variance in satisfaction scores.

#### *Analysis with Multiple Predictor Variables*

The pooled-regression approach to analyzing the APIM has a natural extension in which multiple predictor variables can be examined simultaneously. In the multiple predictor variable extension, the two regressions

(within and between dyads) are conducted as multiple regressions. For the within-dyads regression, each of the mixed predictor variables, as well as the outcome variable, are differenced (always in the same direction across partners). For the between-dyads regression, each of the mixed predictor variables and the outcome variable are averaged across partners. Actor and partner effects for each mixed predictor can then be estimated using the formulas presented previously.

Some analyses may contain purely between-dyads or purely within-dyads predictor variables. In our example study with roommates, some of the roommate dyads are both men, and others are both women, and so gender is a between-dyads variable. (Although not included in this fictitious example, a within-dyads variable might be a categorical variable denoting which of the two roommates has greater financial resources.) For both purely within- or purely between-dyads variables, separate actor and partner effects cannot be estimated; however, such variables can be included in the model when estimating actor and partner effects for mixed variables, and their effects can be controlled. Purely between-dyads variables would be included only in the between-dyads regression. Thus, in the example, the gender variable (coded 1 for men and -1 for women) would be included in the between-dyads regression. If a within-dyads variable were to be included as a predictor variable, the difference between the dyad members' scores (always differencing in a way consistent with how the outcome variable was differenced) would be included as a predictor only in the within-dyads regression. Such analyses would provide estimates of the general effects of the purely between-dyads and purely within-dyads variables, as well as estimates of both actor and partner effects for all mixed variables.

In addition to including the main effects of purely between- or within-dyads variables, this approach can also accommodate interactions between mixed variables and between- or within-dyads variables. Such interactions allow researchers to test whether actor and partner effects for the mixed variables vary significantly across the levels of a between-dyads or a within-dyads variable. In our example study of roommates, we could estimate and test the interaction between the hours of housework done (the mixed variable) and gender (the between-dyads variable). The actor component of this interaction tests whether the relation between a person's time spent on housework and his or her own satisfaction differs for men and women. The partner component tests whether the relation between a person's time spent on housework and his or her partner's satisfaction (the

partner effect) differs for men and women. Similar effects can be computed for the interactions between within-dyads variables and mixed variables. A detailed multivariate example using the pooled regression approach can be found in Kashy and Kenny (2000).

Although the pooled regression method provides estimates of the APIM, it has three serious drawbacks. First, it is awkward and cumbersome, because it requires piecing together several analyses. This awkwardness can easily lead to computational errors. Second, unsaturated models (e.g., models in which some variables may have only actor effects, and other variables may have only partner effects) cannot be estimated. Third, the pooled regression method does not allow for missing data, whereas both MLM and SEM allow for missing data for one member of the dyad (although if  $Y_1$  is missing,  $X_1$  should not be missing, because  $X_1$  is in the  $Y_2$  equation.) For these reasons, MLM and SEM are now the preferred methods for estimating the APIM.

### Estimating the APIM with Multilevel Analysis

In this section we discuss estimating the APIM effects using four multilevel programs, SAS, SPSS, HLM, and MLwiN. We do not discuss the use of some SEM programs (e.g., LISREL and Mplus) that directly estimate multilevel models.

#### SAS and SPSS

To estimate the APIM using a multilevel approach, the data set needs to be arranged as a pairwise data set (see Chapter 1). For such a data set, each individual is an observation (i.e., each individual has his or her own data record), and each individual's outcome score is associated with both his or her own predictor scores and his or her partner's predictor scores. Thus each person's predictor score is entered twice, once as an actor predictor score associated with that person's outcome and once as a partner predictor score associated with the partner's outcome. For the data in Table 7.1, the first two dyads' data would be entered as:

Dyad	Person	Gender	Satisfaction	Actor Housework	Partner Housework
1	1	1	6	1.2	0.6
1	2	1	7	0.6	1.2
2	1	-1	5	4.3	2.3
2	2	-1	9	2.3	4.3

The data are input for each individual independently, such that each individual is treated as one case and there are two cases for each couple. Therefore, with 20 couples in the example, the input statement would read in 40 individual cases. It is advisable, though not always necessary, to sort the data set by dyad and to create a dummy record if a case is missing.

The SAS code for a multilevel analysis that estimates the basic APIM parameters (i.e., actor and partner effects for housework predicting satisfaction) is

```
PROC MIXED COVTEST;
CLASS DYADID;
MODEL SATISFACTION = ACT_HOUSE PART_HOUSE / SOLUTION
      DDFM=SATTERTH;
REPEATED / TYPE=CS SUBJECT=DYADID;
```

Alternatively, instead of the "REPEATED" statement, we could have the following record:

```
RANDOM INTERCEPT / SUBJECT = DYADID;
```

As explained in Chapter 4, the latter approach treats the nonindependence as a variance and not as a correlation, and so it cannot handle negative nonindependence. The "REPEATED" statement, which we prefer, allows for negative nonindependence.

As we discussed in Chapter 4, the default estimation method used by SAS's PROC MIXED is REML, and the estimates derived from this default exactly replicate those given using the pooled-regression approach. The COVTEST option requests that SAS provide tests of the variance components. The CLASS statement indicates the variable that identifies dyad membership (DYADID). SATISFACTION is the individual's outcome score, ACT\_HOUSE is the number of hours per week that the person does housework, and PART\_HOUSE is the number of hours per week that the person's partner does housework. The SOLUTION option in the MODEL statement requests that SAS print the estimates for the intercept and the actor and partner effects for housework. The DDFM = SATTERTH option requests the Satterthwaite (1946) approximation to determine the degrees of freedom for the intercept and slopes (Kashy & Kenny, 2000). The degrees of freedom for mixed predictor variables using the Satterthwaite approximation is some value between the number of dyads less 1 and the number of individuals in the study less 2. The REPEATED statement treats the individual scores as repeated measures in the dyad, and CS implies

compound symmetry, which means that the variances of the intercepts for the two dyad members are equal. As we said earlier, nonindependence is estimated as a correlation and not as a variance.

A subset of the results from the SAS's PROC MIXED analysis is presented in Table 7.2. Turning to these results, in the solution for the fixed effects, the effect estimate for ACT\_HOUSE is the actor effect for housework,  $b = -0.59$ ,  $t(37) = 2.37$ , and the effect estimate for PART\_HOUSE is the partner effect for housework,  $b = 0.89$ ,  $t(37) = 3.56$ . The value of the intraclass correlation in satisfaction scores, controlling for the effects of housework, can be computed by taking the ratio of the CS covariance parameter estimate (0.95) to the sum of the CS parameter and the residual variance estimate (0.95 + 1.65). Thus the partial intraclass correlation is .36, indicating that after controlling for the effects of housework, satisfaction levels for the two roommates were somewhat similar.

**TABLE 7.2.** APIM Results from SAS's PROC MIXED for the Artificial Roommate Data in Table 7.1

Covariance Parameter Estimates				
Cov	Subject	Estimate	Standard Error	Z
Parm	DYADID	0.9477	0.6493	1.46
CS		1.6530	0.5363	3.08
Residual				
Fit Statistics				
-2 Res Log Likelihood				148.4
AIC (smaller is better)				152.4
AICC (smaller is better)				152.7
BIC (smaller is better)				154.4
Null Model Likelihood Ratio Test				
DF	Chi-Square	Pr > ChiSq		
1	2.65	0.1035		
Solution for Fixed Effects				
Effect	Estimate	Standard Error	DF	t Value
Intercept	4.7910	0.6387	18	7.50
ACT_HOUSE	-0.5908	0.2494	37	-2.37
PART_HOUSE	0.8878	0.2494	37	3.56
				Pr >  t
				<.0001
				0.0232
				0.0010

We have requested the Satterthwaite degrees of freedom that represent a complicated weighted average of the between and within degrees of freedom (see the earlier formula). Fortunately, SAS and SPSS do provide them. However, some multilevel programs do not give degrees of freedom (e.g., MLwiN), and others give either the between or the within (HLM6) degrees of freedom. We believe that the Satterthwaite estimate of degrees of freedom is more appropriate because it takes into account the mixture of the between and within parts of the estimate. However, when there is a reasonable number of cases, the differences in the  $t$ -test and  $p$  values due to differences in degrees of freedom are trivial.

The estimate of the pseudo  $R^2$  can be obtained by computing

$$R^2 = 1 - \frac{CS + RES}{CS + RES'}$$

where CS is the compound symmetry term and RES is the residual error variance. The prime refers to the unrestricted model, that is, a model without ACT\_HOUSE or PART\_HOUSE in the equation. For the example, the estimate of pseudo  $R^2$  is  $1 - (0.948 + 1.653) / (262 + 2.975) = .196$ . Note that this value differs from the pooled regression value because variance and covariance components are used, not sums of squares. If the coefficients are small, it is possible that the estimated value for  $R^2$  would be negative. In such a case, it would be reported as zero.

The SPSS specification for the APIM is quite similar to that of SAS. One needs version 11 of SPSS or later; we recommend having at least version 12.0. Again, the data set needs to be structured such that each individual's outcome is associated with both that person's predictor score and the person's partner's predictor score (i.e., a pairwise data set; see Chapter 1). The SPSS code for the analysis that includes both actor and partner effects of housework predicting satisfaction is

```
MIXED
  SATISFAC WITH ACT_HOU PART_HOU
  /FIXED = ACT_HOU PART_HOU
  /PRINT = SOLUTION TESTCOV
  /REPEATED = PERSONID | SUBJECT(DYADID) COVTYPE(CS) .
```

The estimates from SPSS are contained in Table 7.3. Note that the values are essentially the same as those from SAS. The only difference is that the  $p$  value for the error variance is twice as large as that from SAS. Because tests of variance should be one-sided, the  $p$  value for the variances from SPSS should be divided by 2.





**TABLE 7.4.** APIM Results from HLM6 Analysis of Artificial Roommate Data in Table 7.1

Summary of the model specified (in equation format)

## Level-1 Model

$$Y = B_0 + B_1(\text{ACT\_HOU}) + B_2(\text{PART\_HOU}) + R$$

## Level-2 Model

$$B_0 = G_{00} + U_0$$

$$B_1 = G_{10}$$

$$B_2 = G_{20}$$

$$\text{Sigma\_squared} = 1.65301$$

Tau

$$\text{INTRCPT1, } B_0 \quad 0.94763$$

Random level-1 coefficient Reliability estimate

$$\text{INTRCPT1, } B_0 \quad 0.534$$

The outcome variable is SATISFAC

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	d.f.	Approx. P-value
For INTRCPT1, B0					
INTRCPT2, G00	4.790978	0.638676	7.501	19	0.000
For ACT_HOU slope, B1					
INTRCPT2, G10	-0.590827	0.249389	-2.369	37	0.023
For PART_HOU slope, B2					
INTRCPT2, G20	0.887773	0.249389	3.560	37	0.001

actor and partner effects on the number of individuals in the sample. Therefore, the significance tests are very slightly more liberal in HLM than in SAS and SPSS.

## MLwiN

Finally, we have used the computer program MLwiN to estimate the APIM. As explained in Chapter 4, MLwiN requires an individual data file, and for the APIM the data file would need to be a pairwise data file in which each person's outcome score is associated with that person's predictor score and the partner's predictor score. As was the case with HLM, there are two levels in the analysis, level 1 being individual and level 2 being dyad. In Table 7.5, we have the output from the analysis using MLwiN. We see that the estimates are identical to those of the other programs. However, the deviance is somewhat different, because MLwiN uses generalized least squares as its estimation method rather than the REML method used by the other programs. As mentioned in Chapter 4, MLwiN does not provide degrees of freedom or statistical tests of the parameter estimates. However, the parameter estimates may be divided by their standard errors of the estimate (the term in parentheses after the estimate), which are provided, to obtain a Z score.

## Comparing the Size of Actor and Partner Effects

Recall our earlier discussion of the relative sizes of actor and partner effects. In this discussion, we suggested that in the couple-oriented model, actor and partner effects are equal such that a person's outcome is equally

**TABLE 7.5.** APIM Results from MLwiN Analysis of Artificial Roommate Data in Table 7.1

$$\text{satisf}_{ij} \sim N(\alpha_B, \Omega)$$

$$\text{satisf}_{ij} = \beta_{0i}\text{intercept} + -0.591(0.258)\text{act\_house}_{ij} + 0.888(0.258)\text{part\_house}_{ij}$$

$$\beta_{0i} = 4.791(0.523) + e_{0ij}$$

$$\{e_{0ij}\} \sim N(0, \Omega_e); \Omega_e = [2.382(0.533)]$$

$$-2 \times \log\text{likelihood(IGLS Deviance)} = 148.232(40 \text{ of } 40 \text{ cases in use})$$



determined by his or her own and his or her partner's predictors. We can conduct a statistical test that determines whether or not the actor and partner effects differ significantly. To do such a test, we would estimate a multilevel model in which two predictors,  $(X_1 + X_2)/2$  and  $X_1 - X_2$ , are used to predict each person's score on  $Y$ . That is, we predict each person's outcome with the dyad average and the dyad difference of the predictor. Note that the difference refers to the person's score minus the partner's score. If the average has an effect and the difference does not (i.e., it is not statistically significant), then we would conclude that actor and partner effects were not significantly different from each other.<sup>2</sup>

We can also conduct a similar test for social comparison effects. Recall that social comparison effects are seen when the actor and partner effects are of similar size but opposite sign. Again, we would enter the dyad average and difference,  $(X_1 + X_2)/2$  and  $X_1 - X_2$ , as predictors of  $Y$  in a multilevel analysis. In this case, if the average had no effect and the difference did, then we would conclude that actor and partner effects were equal but have opposite signs.

### *Analyses with Interactions and Multiple Predictor Variables*

Inclusion of actor and partner effects for more than one mixed predictor variable, purely within- or between-dyads variables, interactions between actor and partner effects, or interactions between actor and partner effects and within- or between-dyads variables is quite straightforward with MLM using SAS or SPSS. In either program, new variables (such as multiplicative interactions or the absolute value of actor and partner differences) can be created that are functions of existing variables. Such variables or interactions between variables are simply added to the MODEL statement in SAS or the statement that specifies the outcome and the predictor variables in SPSS. However, as we noted in Chapter 4, it is important that any variables included in interactions should: (1) have a meaningful zero and, if they do not, should be grand-mean centered; and (2) have their main effects included in the model statement, in addition to the interaction. Moreover, as we discussed earlier in the chapter, the interaction should be created in a manner that is theoretically relevant. For instance, if the interest is similarity, then the absolute difference, not the product, should be formed.

In SAS's PROC MIXED, multiplicative interactions can be formed directly in the model statement itself. For example, the following SAS

MODEL statement could be used to estimate the actor and partner effects of housework, as well as the multiplicative interaction between actor and partner effects:

```
PROC MIXED COVTEST;
  CLASS DYADID;
  MODEL SATISFACTION = ACT_HOUSE PART_HOUSE
    ACT_HOUSE*PART_HOUSE / SOLUTION
    DDPM=SATTERTH;
  REPEATED / TYPE=CS SUBJECT=DYADID;
```

In the example, the interaction estimates the degree to which the impact that the partner's housework contribution has on a person's satisfaction is moderated by the amount of housework the person does. For the example data, the interaction coefficient is only  $-0.05$  and is not statistically significant.

Similarly, the more recent version of SPSS also allows users to specify multiplicative interactions directly in the statement of the model:

```
/FIXED = ACT_HOU PART_HOU ACT_HOU*PART_HOU
```

In older versions of SPSS, interaction variables must first be created via a "compute" statement. For example, if we wanted to test a model that allowed for interactions between gender and actor or partner effects, we would first compute the interaction variables:

```
COMPUTE ACTGEND = ACT_HOU*GENDER
COMPUTE PARTGEND = PART_HOU*GENDER
```

We would then include the main effect variables, as well as the interaction variables, with gender in our model:

```
SATISFAC WITH ACT_HOU PART_HOU GENDER ACTGEND PARTGEND
/FIXED = ACT_HOU PART_HOU GENDER ACTGEND PARTGEND
```

This model allows us to estimate the degree to which actor and partner effects for housework on satisfaction vary for men and women.

Specifying interactions with HLM6 and MLwiN requires some forethought. Interactions between actor and partner effects, or between within-dyads variables and actor (or partner) effects, cannot be created while running HLM6 or MLwiN. Instead, they must exist as already com-

puted values within the appropriate data set. For example, to create an actor-partner interaction for housework, a variable that represents the interaction between the two partners' housework scores would need to be created prior to the analysis. Note that interactions between actor and partner effects are dyad-level effects and would need to be included in the level-2 data set. Interactions between a level-1 variable (i.e., a mixed variable such as ACT\_HOU or a within-dyads variable) and a between-dyads variable, which is a level-2 variable, such as GENDER in the example, can be estimated within the HLM or MLwiN program itself and need not be included in the data sets.

### Summary

MLM is perhaps the most flexible estimation approach for the APIM. Unlike the pooled-regression solution, MLM provides direct estimates of actor and partner effects, as well as tests of these effects. It also allows researchers to specify constrained models in which some mixed predictor variables have only actor effects, whereas others may have only partner effects. A third approach offers some advantages under certain situations: SEM.

### Estimating the APIM with SEM

SEM with indistinguishable dyads is not easy to do, and we recommend that its use be restricted to two situations. It is the approach of choice for researchers who are interested in estimating latent variable models or models in which the  $X$  variables have measurement error. We should also note that one advantage of SEM is that the entire model is estimated. Thus the correlation between  $X_1$  and  $X_2$  is estimated within the same model and does not have to be estimated separately.

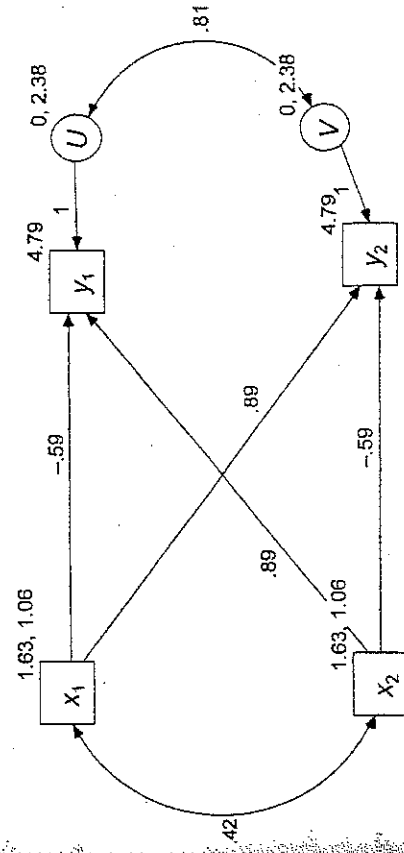
The model depicted in Figure 7.1 is used to estimate the actor and partner effects. Dyad is the unit of analysis, and we treat the dyad as if members were distinguishable. In Chapter 5, we presented the method developed by Olsen and Kenny (2006) using equality constraints. To estimate the APIM parameters for indistinguishable dyads, we estimate means and intercepts forcing the following constraints— $a_1 = a_2$ ,  $p_{12} = p_{21}$ ,  $V(X_1) = V(X_2)$ ,  $V(E_1) = V(E_2)$ ,  $M_{X1} = M_{X2}$ —and also forcing the intercepts of  $E_1$  and  $E_2$  to be equal. Thus a total of six equality constraints are made. The resulting model is invariant under different assignment of person to 1 and 2.

However, the chi-square value varies. This model is the baseline model with 6 degrees of freedom.

We used the housework example in Table 7.1, and we present the estimates in a path diagram in Figure 7.2. Note that paths, means, intercepts, and variances have been set equal. We obtained essentially the same estimates as we obtained using multilevel programs. (They would be exactly the same if we use the maximum likelihood option within MLM.) We find a  $\chi^2(6) = 10.83$ . If we reversed assignments of person 1 and person 2 for some of the dyads, the chi-square would change, but the parameter estimates would remain the same. As explained in Chapter 5, the chi-square test for the base model should not be interpreted. Rather, this model serves as the I-SAT model and is used to adjust models' chi-squares and degrees of freedom.

Consider a model that forces the actor and partner effects to be equal. We obtain a  $\chi^2(7) = 22.97$ . We take this chi-square and adjust it by the I-SAT one to obtain  $\chi^2(1) = 12.06$ ,  $p < .001$ . Thus we conclude that actor and partner effects are statistically significantly different.

Clearly, the SEM solution for indistinguishable dyads is awkward and may be difficult for researchers to implement. As we have noted earlier, we strongly recommend that most researchers with indistinguishable dyads use the MLM approach described earlier in this chapter. However, SEM is relatively easy to implement for estimating the APIM when dyad members are distinguishable, the topic that we now consider.



**FIGURE 7.2.** Estimation of the APIM parameters using structural equation modeling.

### ESTIMATION OF THE APIM: DISTINGUISHABLE DYADS

Distinguishable dyad members, by definition, vary on a within-dyads variable. For example, heterosexual couples are distinguishable by their gender: Each couple has one man and one woman. Other distinguishable dyads might be mother-daughter dyads or supervisor-subordinate dyads. When dyad members are distinguishable on a nonarbitrary variable, the APIM can again be estimated using either pooled regression, MLM, or SEM. Relatively minor changes are required for the pooled-regression and MLM approaches to accommodate distinguishable dyads. The SEM approach, however, changes substantially, and as a result of the changes, this approach becomes dramatically easier to implement.

In describing the APIM for distinguishable dyads, we use data gathered from 98 heterosexual dating couples (Campbell, Simpson, Boldry, & Kashy, 2005). In this example, both members of each couple completed a measure of neuroticism at the beginning of the study. Then, following a 2-week-long diary study, in which individuals reported on the daily conflicts and supportive interactions that occurred with their partners, the couples returned to the laboratory. During a videotaped interaction, they were to try to resolve one of the most serious conflicts that had occurred during the diary data collection period. The videotapes were then coded for a range of behaviors, including five indicators of distress: appearing upset, disappointed, unhappy, satisfied (reversed), and positive (reversed). Our example examines the degree to which a person's own neuroticism predicts that person's distress (an actor effect) and the degree to which a person's partner's neuroticism predicts that person's distress (a partner effect). Because the dyads are distinguishable with respect to gender, we also examine whether the actor and partner effects differ for men and women. An APIM diagram with these variables is presented in Figure 7.3.

#### Pooled-Regression Method

In the pooled-regression approach, the data are organized with dyad as the unit of analysis (see Chapter 1). With this data organization, each record contains two measurements of the outcome variable, one for the man and one for the woman. Similarly, each record also contains two measurements of the mixed predictor variable. Thus there are the scores  $X_M$  and  $X_F$  for each predictor variable and  $Y_M$  and  $Y_F$  for the outcome variable. Recall that in the pooled-regression solution, two regressions are computed: a between-dyads regression in which the dyad averages on the outcome are

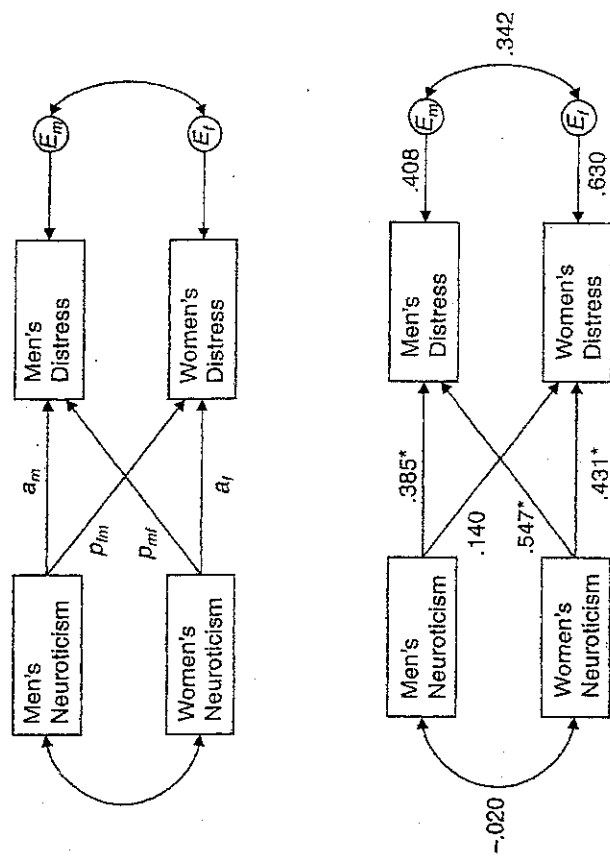


FIGURE 7.3. APIM for Simpson et al. example.

predicted by the dyad averages on the predictor and a within-dyads regression in which the dyad differences on the outcome are predicted by the dyad differences on the predictor. Recall, too, that the intercept is suppressed (i.e., fixed to zero) in the within-dyads regression. To compute the APIM using the pooled-regression solution for distinguishable dyads, several additional variables need to be formed and added to the within- and between-dyads regressions.

First, two variables denoting the gender of the men and women (or whatever the distinguishing variable is) need to be created. Be aware that these are actually constants, because each dyad contains one man and one woman, and for each dyad there is a variable  $M\_GENDER$  that equals 1.0 for every dyad and a variable  $F\_GENDER$  that equals  $-1.0$  for every dyad. These two variables are then used to create a gender-difference variable (e.g., in SAS  $GENDDIFF = M\_GENDER - F\_GENDER$ ;  $GENDDIFF$  is a constant that always equals 2.0). In the example, the mixed predictor variable is neuroticism, and there is a score for men ( $M\_NEURO$ ) and a score for women ( $F\_NEURO$ ). A difference score of this variable must be created:  $NEURODIFF = M\_NEURO - F\_NEURO$ . Finally, the interaction

between gender and the mixed predictor variable is also a mixed variable for which actor and partner effects can be estimated. These values are formed by multiplying the men's neuroticism score by the men's gender and the women's neuroticism score by the women's gender:  $M\_INTER = M\_NEURO * M\_GENDER$ ; and  $F\_INTER = F\_NEURO * F\_GENDER$ . A difference score for the interaction is then created:  $INTERDIFF = M\_INTER - F\_INTER$ . The within-dyads regression has three variables predicting the dyad difference in distress ( $DISTRESSDIFF$ ), including the difference in neuroticism scores, the gender difference variable, and the difference in the interaction scores (i.e.,  $NEURODIFF$ ,  $GENDERDIFF$ , and  $INTERDIFF$ ). The resulting difference score equation is

$$DISTRESSDIFF = .065 (NEURODIFF) - .035 (GENDERDIFF) + 0.180 (INTERDIFF) + E_w$$

For each of the mixed predictor variables (neuroticism and the neuroticism-by-gender interaction), a dyad-average variable must be created [e.g.,  $NEUROAVG = (M\_NEURO + F\_NEURO)/2$ ;  $INTERAVG = (M\_INTER + F\_INTER)/2$ ]. The between-dyads regression involves predicting the dyad average on distress ( $DISTRESSAVG$ ) as a function of the two-dyad average variables. Note that gender is not included in the between equation because it does not vary between dyads (i.e., all dyads have one man and one woman). The resulting "average" equation is

$$DISTRESSAVG = 3.647 + .751 (NEUROAVG) - 0.227 (INTERAVG) + E_b$$

The results of the between- and within-dyads regressions are then pooled using the formulas provided earlier. For the example data, the neuroticism actor effect is  $(0.751 + 0.065)/2 = 0.408$ ; the neuroticism partner effect is  $(0.751 - 0.065)/2 = 0.343$ . The standard error for these two effects, again using the formula presented earlier, is  $\sqrt{(0.139^2 + 0.061^2)/4} = 0.076$ . The degrees of freedom are estimated to be 130.00. Thus, both the actor effect,  $t(130) = 5.39$ ,  $p < .001$ , and the partner effect,  $t(130) = 4.54$ ,  $p < .001$ , are statistically significant, indicating that individuals higher in neuroticism appeared to be more distressed during the discussion and that individuals whose partners were higher in neuroticism also appeared to be more distressed.

The gender difference for distress was not statistically significant,  $b = -0.035$ ,  $t(95) = 0.90$ . However, examination of the actor and partner effects

(again using the earlier formulas) shows some interesting interactions between neuroticism and gender. The actor effect for the interaction was small,  $(-0.227 + 0.180)/2 = -0.023$ , but the partner effect was relatively large,  $(-0.227 - 0.180)/2 = -0.204$ , and statistically significant,  $t(135) = -2.76$ . This partner effect indicates that the neuroticism partner effect described earlier is qualified by gender such that men with highly neurotic female partners showed a great deal of distress (the partner effect for women's neuroticism on men's distress, given how gender is coded, is  $0.343 + 0.204 = 0.547$ ), but women with highly neurotic male partners were not as affected by their partner's neuroticism (the partner effect for men's neuroticism on women's distress is  $0.343 - 0.204 = 0.139$ ).

As was the case with indistinguishable dyads, the pooled-regression solution for distinguishable dyads is awkward. It has one additional drawback in the distinguishable case: We must assume homogeneity of variance for the two levels of the distinguishing variable. That is, we must assume that the men and women in our example have the same variance in their distress scores. The MLM and SEM approaches for the APIM do not necessarily require such an assumption, and these methods are preferred over the pooled-regression approach.

### Multilevel Modeling

We discuss two different approaches to estimation of the APIM with distinguishable dyads using MLM. The first is an extension of the approach with indistinguishable dyads and uses interaction terms to factor in the effects of the distinguishing variable. The second is the two-intercept approach that was introduced in Chapter 4.

#### Interaction Model

Recall that for MLM the data set needs to be structured so that each individual has his or her own data record on which that person's outcome score is recorded (e.g.,  $DISTRESS$ ). In addition to the person's outcome score, the person's own score on the mixed predictor variable is also recorded (e.g.,  $ACT\_NEURO$ ), as is his or her partner's score on the mixed predictor variable (e.g.,  $PART\_NEURO$ ). Three additional variables need to be included in the distinguishable case: a variable denoting the actor's score on the distinguishing variable (e.g.,  $GENDER$ ) and interactions between the distinguishing variable and the actor and partner scores on

the predictor. In this example, GENDER, our distinguishing variable, is coded as 1 for men and -1 for women. The two interaction variables are ACT\_INTER = ACT\_NEURO\*GENDER and PART\_INTER = PART\_NEURO\*GENDER. Researchers using HLM6 or MLWIN are advised to form these interaction scores within the level-1 data set itself, whereas researchers using SAS or qPSS can use the statistical program to form the interaction variables.

Only two alterations need to be made to the general MLM approach we described earlier to apply it to the case of distinguishable dyads. The first change is that the distinguishing variable and its interactions with the mixed predictor variable need to be added to the model. For example, using PROC MIXED in SAS, the syntax would be

```
PROC MIXED COVTEST;
CLASS DYADID;
MODEL DISTRESS = ACT_NEURO PART_NEURO GENDER
ACT_NEURO*GENDER PART_NEURO* GENDER/
SOLUTION DDFM=SATTERTH;
```

Notice that in this model, in addition to including the actor and partner effects for the mixed variable and the interactions between the distinguishing variable and the mixed variable, we have included the main effect of the distinguishing variable (i.e., the person's GENDER). This main effect simply estimates and tests whether there are mean-level differences in DISTRESS for men and women.

The second change that can be made to the analysis to accommodate the distinguishability factor is that the specification of compound symmetry can be changed to *heterogeneous compound symmetry*. In SAS this is done in the REPEATED statement, with the TYPE being set to CSH rather than CS:

```
REPEATED / TYPE=CSH SUBJECT=DYADID;
```

This change to heterogeneous compound symmetry removes the homogeneity-of-variance assumption and allows the error variances to differ for the two types of dyad members. A partial section of the results from PROC MIXED using the example data are presented in Table 7.6. The actor and partner effect estimates exactly replicate those from the pooled-regression approach. The significance tests differ slightly, however, because the MLM solution allows for heterogeneous variances.

The parallel statements for SPSS are as follows:

```
MIXED
DISTRESS WITH ACT_NEURO PART_NEURO GENDER
/FIXED = ACT_NEURO PART_NEURO GENDER
ACT_NEURO*GENDER PART_NEURO*GENDER
/PRINT = SOLUTION TESTCOV
/REPEATED = GENDER | SUBJECT(DYADID) COVTYPE(CSH).
```

Recall that tests of variances within SPSS are two-tailed when they should be one-tailed, and so  $p$  values should be divided by 2.

**TABLE 7.6.** Results from the Neuroticism Example Using PROC MIXED with Heterogeneous Compound Symmetry

Cov	Covariance Parameter Estimates			
	Subject	Estimate	Standard Error	Z
Parm	DYADID	0.6503	0.09436	6.89
Var(1)	DYADID	0.4211	0.06110	6.89
Var(2)	DYADID	0.6735	0.05606	12.01
CSH				<.0001

Fit Statistics			
-2 Res Log Likelihood			387.6
AIC (smaller is better)			393.6
AICC (smaller is better)			393.7
BIC (smaller is better)			401.5

Null Model Likelihood Ratio Test			
DF	Chi-Square	Pr > ChiSq	
2	61.87	<.0001	

Solution for Fixed Effects					
Effect	Estimate	Error	DF	t	Value Pr >  t
Intercept	3.6474	0.08625	95	42.29	<.0001
ACT_NEURO	0.4079	0.07438	128	5.48	<.0001
PART_NEURO	0.3432	0.07692	123	4.46	<.0001
GENDER	-0.03541	0.03918	95	-0.90	0.3684
ACT_NEURO*GENDER	-0.02340	0.07241	133	-0.32	0.7471
PART_NEURO*GENDER	0.2036	0.07502	127	2.71	0.0076

Note. GENDER is coded men = 1 and women = -1.

*Two-Intercept Model*

An alternative approach for MLM estimation with a distinguishable variable is to estimate the two-intercept model that we described in Chapter 4. Two dummy variables must be created. One might be called MALE (MALE = 1 if the person is male, 0 otherwise) and the other called FEMALE (FEMALE = 1 if the person is female, 0 otherwise). The correlation of these two variables is  $-1$ , and, ordinarily, they could not both be included in the same equation. However, they can both be included if we drop the intercept from the model.

We illustrated in Chapter 4 how such a model can be estimated using MLWIN, and we discussed how it can be estimated using HLM. Fortunately, a version of the model can also be estimated using SAS and SPSS. The basic code in SAS is:

```
PROC MIXED COVTEST;
CLASS DYADID;
MODEL DISTRESS = MALE FEMALE ACT_NEURO*MALE /
  NOINT SOLUTION DDFM=SATTERTH;
REPEATED / TYPE=CSH SUBJECT=DYADID;
```

An alternative, equivalent, and simpler<sup>3</sup> code is:

```
PROC MIXED COVTEST;
CLASS DYADID GENDER;
MODEL DISTRESS = GENDER ACT_NEURO*GENDER
  PART_NEURO*GENDER/ NOINT SOLUTION DDFM=SATTERTH;
REPEATED / TYPE=CSH SUBJECT=DYADID;
```

Results from the first version of this SAS code are presented in Table 7.7. This table shows an advantage of the two-intercept model: The actor and partner effects for men and women can be read directly from the output, and no hand computations are required. There is, however, one substantial disadvantage to the two-intercept model—that is, that there is no direct test of whether the actor or partner effects differ significantly for men versus women. To perform that test, we need to use the earlier model that included gender interactions.

The parallel two-intercept code for SPSS is:

```
MIXED
  DISTRESS WITH MALE FEMALE ACT_NEURO PART_NEURO |
  NOINT
```

```
/FIXED = MALE FEMALE ACT_NEURO*MALE
  ACT_NEURO*FEMALE PART_NEURO*MALE
  PART_NEURO*FEMALE
/PRINT = SOLUTION TESTCOV
/REPEATED = GENDER | SUBJECT (DYADID) COVTYPE(CSH).
```

The alternative specification is:

```
MIXED
  DISTRESS WITH GENDER ACT_NEURO PART_NEURO | NOINT
/FIXED = GENDER ACT_NEURO*GENDER PART_NEURO*GENDER
/PRINT = SOLUTION TESTCOV
/REPEATED = GENDER | SUBJECT (DYADID) COVTYPE(CSH).
```

**TABLE 7.7.** Results from the Neuroticism Example Using PROC MIXED with the Two-Intercept Model and Heterogeneous Compound Symmetry

Cov	Covariance Parameter Estimates				
	Subject	Standard Estimate	Error	Value	Pr > Z
Var(1)	DYADID	0.6503	0.09436	6.89	<.0001
Var(2)	DYADID	0.4211	0.06110	6.89	<.0001
CSH	DYADID	0.6735	0.05606	12.01	<.0001

Fit Statistics		
-2 Res Log Likelihood		383.4
AIC (smaller is better)		389.4
AICC (smaller is better)		389.6
BIC (smaller is better)		397.3

Null Model Likelihood Ratio Test			
DF	Chi-Square	Pr > ChiSq	
2	61.87	<.0001	

Solution for Fixed Effects					
Effect	Estimate	Error	DF	t Value	Pr >  t
MALE	1.1663	0.3581	95	3.26	0.0016
FEMALE	2.1838	0.4451	95	4.91	<.0001
MALE*ACT_NEURO	0.3845	0.1009	95	3.81	0.0002
FEMALE*ACT_NEURO	0.4313	0.1066	95	4.05	0.0001
MALE*PART_NEURO	0.5468	0.0858	95	6.38	<.0001
FEMALE*PART_NEURO	0.1396	0.1254	95	1.11	0.2687



### Structural Equation Modeling

The SEM solution with distinguishable dyads is perhaps the simplest data-analytic method for estimating the APIM, in the sense that the model can be directly estimated using a standard application of a well-known data-analytic method. (See Bui et al., 1996; Kenny & Acitelli, 2001; and Murray, Holmes, & Griffin, 1996, for illustrations of this approach.) Essentially, the SEM approach involves estimating the APIM parameters as they appear in the model presented in Figure 7.3. The data have a dyad-level structure. Written in the form of two linear equations, where  $Y_m$  is the man's distress,  $Y_f$  is the woman's distress,  $X_m$  is the man's neuroticism (centered around the grand mean across both men and women), and  $X_f$  is the woman's neuroticism (also centered around the grand mean across both men and women), this model can be summarized as:

$$Y_m = a_m X_m + p_{mf} X_f + E_m,$$

$$Y_f = p_{fm} X_m + a_f X_f + E_f.$$

Note that the couple is the unit of analysis, and thus the sample size for this analysis is the number of couples (which is 98 in the example). The model in Figure 7.3 is identical to that in Figure 7.1, but now separate actor and partner effects are estimated for both members of the dyad.

Interpretation of the actor and partner effects is straightforward. For this example,  $a_m$  refers to the effect of the man's neuroticism on his own level of distress. The partner effect,  $p_{fm}$ , is the effect of the man's neuroticism on his partner's distress, and  $p_{mf}$  is the effect of the woman's neuroticism on her partner's distress. Recall that the usual convention is to have the effect first and the cause second.

The SEM solution allows model constraints to be placed and tested. For example, one can test whether the actor effects differ significantly for men and women by constraining the two actor parameters to be equal and then assessing the degree to which this constraint significantly worsens the model fit. When this constraint is placed for the example data, the actor effect is estimated to be .407 (for both men and women), and the chi-square test with 1 degree of freedom indicates that this constraint does not significantly worsen fit,  $\chi^2(1) = 0.107$ ,  $p = .744$ . On the other hand, constraining the partner effects to be equal does significantly worsen the model fit,  $\chi^2(1) = 7.311$ ,  $p = .007$ , indicating that there is a statistically significant difference between the two partner effects. The effect from the

woman's neuroticism to the man's distress is larger than the effect from the man's neuroticism to the woman's distress.

Finally, we must add one note of caution. When using SEM, it is essential that the coefficients not be standardized separately for each dyad member type (e.g., for men and women separately), because such a procedure renders the coefficients incomparable across dyad member type. The safest course of action is not to standardize at all but, instead, to report the unstandardized coefficients (as we have done in Figure 7.3). Alternatively, one can standardize the variables before computing the model. Such an approach would involve standardizing the data using the mean computed across men and women, as well as the standard deviation computed across the entire sample (see Chapter 6).

### POWER AND EFFECT SIZE COMPUTATION

As discussed in Chapter 3, determining effect sizes and power when predictor variables are either entirely between or within dyads is relatively straightforward. To determine power, the effect size measure based on the nonindependent data (either  $d$  or  $r$ , what we called  $d_p$  and  $r_D$  in Chapter 3) is determined, and then that value is adjusted by the degree of nonindependence in the data to create an adjusted  $d$  (or  $r$ ). Recall that this adjusted value is the estimate of the effect size for independent units. With the adjusted  $d$  or  $r$ , the level of power is determined using the appropriate sample size.

We adopt the same strategy here, but it is quite a bit more complicated. In fact, it is so complex that we suggest the following approximation. We first compute the intraclass, or Pearson, correlation for the predictor variable, or  $X$ , and we denote that value as  $r_x$ . If  $r_x$  is  $> .5$ , treat the variable as if it were a between-dyads variable. If  $r_x < -.5$ , treat it as if it were a within-dyads variable. For all other values, ignore the nonindependence when estimating power (i.e., use the unadjusted effect size measure). This general strategy yields a good approximation of the power.

The more complicated and exact strategy is to combine the between- and within-dyads adjustments that we presented in Chapter 3, weighting by the intraclass correlation of the mixed variable. We first need to determine the adjustment factor that corrects for the nonindependence between dyad members' scores. This value is:



$$\sqrt{\frac{(r_x + 1)^2 (1 - r_x)^2}{2(1 + r_x) + 2(1 - r_y)}}$$

where  $r_x$  is the intraclass, or Pearson, correlation for the mixed variable, and  $r_y$  is the residual intraclass, or Pearson, correlation for the outcome variable.

To determine power, we estimate  $d$  or  $r$  from the nonindependent data, and then we multiply that value<sup>4</sup> by the preceding adjustment factor to obtain a modified measure of effect size. However, there are two additional complications. First, note that in an APIM analysis we generally measure actor effects controlling for partner effects, and vice versa. To control for multicollinearity between the two effects, we further need to adjust the effect size by  $\sqrt{1 - r_x^2}$ . The resulting modified formula that includes both adjustments is

$$\sqrt{\frac{1 - r_x^2}{2} \left[ \frac{(r_x + 1)^2 (1 - r_x)^2}{(1 + r_x)(1 - r_y)} \right]}$$

Of course, if the design is either between or within ( $r_x$  equals 1 or -1), this adjustment should not be made.

Finally, in determining power, we need to correct the sample size by the degree of nonindependence. If we denote the total sample size as  $2n$  (where  $n$  is the number of dyads), the effective sample size in terms of a power computation can be shown to be approximately equal to  $2n/(1 + r_x^2)$ . This corrected value is less than or equal to  $2n$ , and this value is used as the sample size to determine power.

As an example, assume that the effect size based on nonindependent data,  $d$  (what we referred to as  $d_0$  in Chapter 3), is 0.5; the correlation between the dyad members' mixed predictor scores,  $r_x$ , is .3; and the correlation between the outcome scores, or  $r_y$ , is .5. These values result in a corrected  $d$  of 0.490 (the adjustment for nonindependence is 1.026 and the adjustment for multicollinearity is 0.954, and therefore the overall adjustment is 0.979). The effective sample size using 100 persons, or 50 dyads, is 91.7. The estimated power is .64. If, instead, we were to use the simple approximation of not adjusting the effect size, and using 100 as the sample size, the power, assuming independent data for a  $d$  of 0.5 and 100 cases, would be .697, not all that different from the value of .64.

Returning to the fictitious roommate data that we considered earlier in the chapter, we can use the adjustment factor to compute the effect size. We compute  $d$  or  $r$  using the standard formulas. Then we take that  $d$  or  $r$  and divide by the adjustment factor to determine the effect size. Spe-

cifically, consider the estimate of the partner effect for the roommate data in Table 7.1. We obtained  $t(37) = 3.56$ , with  $r_x = .415$  and  $r_y = .364$ . Using the  $t$  to  $r$  formula of  $t/\sqrt{df + 1}$ , we obtain an estimated  $r$  for the nonindependent data of 0.505. The adjustment for multicollinearity is .910, and the adjustment for nonindependence is 1.001, resulting in an overall adjustment of 0.910. Thus the estimated effect size is 0.460. Note that the adjustment factor results in a relatively small change in the effect size measure.

## SPECIFICATION ERROR IN THE APIM

The basic APIM model in Figure 7.1 specifies a particular pattern of causation. If that model were not correctly specified, then the parameter estimates would be misleading. An important type of specification error occurs when the structure of the causal relationship is erroneous. Chapter 15 considers two models that are alternatives to the APIM: the mutual-feedback model and the common-fate model. In the mutual-feedback model,  $Y_1$  causes  $Y_2$ , and vice versa. In the common-fate model, the causal effect from  $X$  to  $Y$  occurs between latent variables.

A fundamental assumption of the APIM is that  $X$  causes  $Y$ . We could treat the APIM as a prediction model (i.e.,  $X$  is used to predict  $Y$ ), but, typically, the assumption is that  $X$  causes  $Y$ . Thus, if there were actor and partner effects, a change in  $X$  would lead to a change in both  $Y_1$  and  $Y_2$ . Rarely (in fact, we know of no case in which it has happened) is  $X$  a manipulated variable. If  $X$  were manipulated, then the direction of causality would be known:  $X$  causes  $Y$ , and not vice versa.

Measurement error is also an important issue with the APIM. Measurement error in a causal variable biases not only the coefficient of that causal variable but also the coefficients of the other variables in the equation (Kenny, 1979). Three approaches can be used to address the measurement error problem: disattenuation, latent variables, and instrumental variable estimation. We discuss each in turn. We note that each can relatively easily be accomplished using SEM and cannot be done within either pooled regression or MLM.

For the disattenuation strategy, the reliability of the  $X$  variable must be known. If  $X$  is a scale, its internal consistency estimate may be available. Following the strategy of Williams and Hazer (1986), an SEM program can be used, and we set the path from true  $X$  to measured  $X$  to 1 and fix the

error variance to  $s_x^2(1 - \alpha)$ , where  $\alpha$  is the reliability of  $X$  and  $s_x^2$  is the variance of the measure  $X$ . It is possible that  $X_1$  and  $X_2$  might have different values of reliability, and those different values should each be used.

A latent variable strategy is the most common approach to the estimation of models with measurement error. For example, for  $X_1$ , there would be multiple measures. Normally, just two measures per latent variable would be needed, but because of the likelihood of correlated error across dyad members (see Chapters 5 and 6), three measures per  $X$  variable would be needed to have an identified model. Also, one would want to test that the measurement model is the same for  $X_1$  and  $X_2$ , so that their units of measurement would be comparable. Ferrer and Nesselroade (2003) illustrate a version of the APIM with latent variables in a study of emotions in heterosexual couples. They found effects from prior negative husband emotion to later wife emotion, but not vice versa.

The instrumental variable solution to estimating models with measurement error in the  $X$  variables is not commonly used. In instrumental variable estimation, the model must specify that the variable measured with measurement error,  $X_1$ , correlates with another variable,  $Z_1$ , which is the instrumental variable. However,  $Z_1$  does not cause either  $Y_1$  or  $Y_2$ , and it is the absence of the paths from  $Z_1$  to  $Y_1$  or  $Y_2$  that allows estimation of the error variance in  $X_1$ .

Finally, if we allow for measurement error in  $X_1$  and  $X_2$ , estimation of interaction effects becomes problematic. Kenny and Judd (1984) have provided a general strategy for the estimation of multiplicative interaction effects (i.e.,  $X_1X_2$ ). In principle, the method they advocate could be adapted to test interactions that are specified as the absolute value of the difference between  $X_1$  and  $X_2$ , as well as interactions that are specified as the maximum or minimum value of the two partners' scores. However, very large sample sizes are needed to estimate these models with any precision.

## SUMMARY AND CONCLUSIONS

This chapter presents the Actor-Partner Interdependence Model (APIM), which can be used to analyze the effects of mixed predictor variables. This model of dyadic data suggests that a person's standing on a predictor variable affects his or her partner's outcomes (i.e., partner effects), as well as their own outcomes (i.e., actor effects). We noted that this general model can be applied to data from groups larger than dyads and that it can incor-

porate other types of independent variables and interactions among variables. Four different configurations or models of actor and partner effects were considered: actor-oriented (the partner's standing has little effect on the person's outcomes), partner-oriented (only the partner's standing predicts the person's outcomes), couple-oriented (both actor and partner effects occur and are similar in size and sign), and social comparison (actor and partner effects are similar in size but opposite in sign).

Our discussion then considered estimation of the APIM effects for both indistinguishable and distinguishable dyads. For each we discussed three different methods: pooled regression, MLM, and SEM. The pooled regression method can be accomplished by pooling together results from regressions that use simple ordinary least squares. Note, however, that this method is now outdated, and researchers should consider MLM or SEM as better options. Although both MLM and SEM can be used for either indistinguishable or distinguishable dyads, MLM is clearly the estimation method of preference for indistinguishable dyads. However, for distinguishable dyads, SEM is the most straightforward method. We detailed methods for computing effect sizes and estimating power for the APIM. Finally, we presented a brief discussion of specification errors and their impact on APIM estimates.

The APIM is a simple yet compelling model of dyadic behavior. When two people interact or are involved in a relationship, each person's outcomes are affected by both his or her own inputs and his or her partner's inputs. In fact, the presence of partner effects can be used as an operational definition of a relationship. As we noted, actor and partner effects may also interact, and sometimes that interaction is the major research focus (e.g., when researchers are interested in the effects of the similarity between dyad members). Although we did not extensively discuss the estimation of interaction effects, they should be considered. We also limited our attention to outcome variables measured at the interval level of measurement (see Chapter 1), and we have not discussed outcomes measured at the nominal level of measurement. Such models can be estimated by several different programs (e.g., HLM6, MLWIN, and SAS's PROC NMIXED). Also, Thomson (1995) has developed a method that tests the effect of a mixed dichotomous variable on a dichotomous outcome. We do discuss multilevel models with a dichotomous outcome in Chapter 14.

In the next four chapters, we leave the standard dyad design (see Chapter 1), and we consider more complicated designs. In Chapters 8, 9, and 11, we consider SRM designs. In Chapter 10, we consider the one-with-many design.

## NOTES

1. Note that in the difference equation there are no predictors or intercept. Some computer programs do not allow the estimation of an equation with no predictors or intercept.
2. The sum and difference approach can be used when there is a couple-level outcome variable (e.g., relationship breakup) and distinguishable dyad members. For such outcomes, we can examine the extent to which one member has more of an effect than the other (Attridge, Berscheid, & Simpson, 1995).
3. We thank Joseph Olsen, who suggested this method to us.
4. As we noted in Chapter 3, the adjusted value of  $r$  could be greater than 1 in absolute value after multiplying by  $\sqrt{2/(1+r)}$ . If such were the case, the investigator has chosen much too optimistic a value of  $r$  and should lower that value.

## 8

## Social Relations Designs with Indistinguishable Members

Some research questions can be addressed only by designs in which persons participate in more than one dyad (Kashy & Kenny, 2000). For example, consider the question of whether affective evaluations are primarily determined by the unique relationship between two individuals. If individuals participate in only one dyad, then a measure of liking of the partner may not necessarily represent unique liking. A high liking score from one member of the dyad could be due to the fact that the individual is a "liker" and that he or she likes everyone. Similarly, it may be that the partner is simply "likable" and is liked by everyone. Finally, it may be the case that liking is, in fact, relationally determined and that ratings of liking within dyads are unique. To separate these three factors, each individual would have to participate in more than one dyad. With multiple interactions, the degree to which a person is a "liker" can be assessed by looking at whether that person generally likes everyone with whom he or she interacts. Similarly, the degree to which a person is "likable" can be assessed by observing whether everyone who interacts with that person likes him or her. Finally, if there is evidence that one person likes the other, over and above the person's tendency to be a "liker" and over and above the partner's tendency to be liked, then there is evidence that liking is uniquely determined by the specific relationship between the two individuals.

In the previous chapter we introduced the Actor-Partner Interdependence Model (APIM). In that model, each individual participates in a single dyad, and one person's outcomes are affected both by properties of that individual (actor effects) and by properties of the person's

## Over-Time Analyses Interval Outcomes

- Cycles in the intensity of mother–infant interactions (Lester, Hoffman, & Brazelton, 1985).
- Changes in marital satisfaction over years of marriage (Kurdek, 1998).

The analysis of over-time data from individuals is very complex, and entire books have been written on this topic. Adding the complications of dyadic data analysis makes a difficult topic even more challenging. Moreover, methodologists have not yet come to a complete consensus on which data-analytic techniques are most appropriate. Devoting only two chapters to covering this topic necessarily means that the discussion is brief and preliminary.

We adopt the following notation. For dyad  $i$ , members 1 and 2 are each measured up to  $T$  times. It need not be the case that all persons are measured at each of  $T$  time points. Generally, we assume that dyad members are distinguishable, but we occasionally consider the case of indistinguishable dyad members.

A fundamental idea in working with over-time data is the concept of lagging. If we have a measure of  $Y_{1it}$  that is a measure of  $Y$  for person 1 of dyad  $i$  at time  $t$ , its lagged value would be  $Y_{1,t-1}$  (the previous observation). Note that there is no lagged value for the first observation,  $Y_{111}$ . If we have  $T$  observations, we have  $T - 1$  observations with complete data if we lag back one time point.

There are many different types of nonindependence in over-time dyadic data. The first is the usual type of nonindependence: dyadic nonindependence. Scores from the two members of the dyad are likely nonindependent, and so the two scores,  $Y_{1it}$  and  $Y_{2it}$ , might well be correlated. Moreover, any parameter that is estimated for each person, for example, a slope or an intercept, might be correlated across the two members. That correlation is usually a positive correlation, but it might be negative.

There is a second type of nonindependence called *autocorrelation*. There is probably no more reliable finding in the social and behavioral sciences than the fact that the best predictor of future behavior is past behavior. Statistically, autocorrelation is the association between a measure taken at one point in time and the same measure taken at another point in time. As with other forms of nonindependence of observations, it can be a problem for statistical analysis. The correlation between these observations must be controlled. However, as with other analyses, nonindependence in a time-series analysis may be interesting in its own right (Warner,

Like our discussion of dyadic indexes in Chapter 12, the methods presented in this chapter are idiographic in nature. However, whereas the dyadic indexes were computed across items or variables for each dyad, in this chapter and the next we consider the analysis of over-time dyadic data. That is, this chapter describes analyses that are appropriate for observations that are gathered across multiple time points from both members of the dyad.

We presume in this chapter that the outcome variable is measured on an interval or ratio scale, such as marital satisfaction. In the next chapter we consider outcome variables that are dichotomous (e.g., positive vs. negative speech acts). For each of the methods described in these two chapters, the behaviors of both members of the dyad are observed repeatedly over time such that for each member there is literally a stream of recorded behavior, commonly called a *time series*. Some examples are:

- The members' responses to their partners' criticism during a conversation (Gottman, Swanson, & Swanson, 2002).
- One person's competitiveness as a response to his or her partner's competitiveness in a game situation (Cook, 1988).
- Level of aggressiveness of one person in response to the partner's level of aggressiveness (Leonard, 1984).

1998). In this chapter and the next, we discuss many different ways to model autocorrelation.

As is true of the majority of research in which the effects of time are considered, we adopt what are called *first-order* models. For these models, any nonindependence between observations separated by two or more time points is assumed to be explained by the intermediate time point. For example, the correlation between  $Y_{1,t-2,t}$  and  $Y_{1,t,t}$  is explained by  $Y_{1,t-1,t}$ . Because the data are dyadic, it is also assumed that the correlation between  $Y_{1,t-1,t}$  and  $Y_{2,t,t}$  is explained by  $Y_{1,t-1,t}$  and  $Y_{2,t-1,t}$ .

The combined presence of dyadic dependence and autocorrelation makes the analysis of over-time dyadic data especially difficult. In this chapter, we consider five types of analyses: cross-lagged regressions, the standard Actor-Partner Independence Model (APIM) growth-curve modeling, cross-spectral analyses, and nonlinear dynamic modeling. However, only the first three topics are covered extensively. We introduce each of the remaining two topics, but the reader is advised to consult other sources for details before attempting such analyses. Three of the models, cross-lagged regressions, standard APIM, and nonlinear dynamic modeling, can be considered to be longitudinal variants of the APIM. Growth-curve modeling emphasizes trends in time series, whereas cross-spectral analysis emphasizes cycles.

We again advise the reader that the models discussed in this chapter are complicated, perhaps the most intricate models discussed in this book. Moreover, our use of multilevel modeling (MLM) in this chapter is much more complicated than it was in Chapters 4 and 7. We also suggest that interested readers consult Laurenceau and Bolger (2005) for an introduction to this topic.

### CROSS-LAGGED REGRESSIONS<sup>1</sup>

The cross-lagged regression model for distinguishable dyad members is contained in Figure 13.1. In this model, we have a person's behavior,  $Y$ , at times  $t$  and  $t - 1$ , for members 1 and 2. For example, the time points may be days, and  $Y$  might be the level of stress experienced by each member of the dyad. The APIM serves as the conceptual model for the cross-lagged regression analysis. The actor effects are the effects from  $Y_{1,t-1}$  to  $Y_{1,t}$  and from  $Y_{2,t-1}$  to  $Y_{2,t}$ , and the partner effects are from  $Y_{1,t-1}$  to  $Y_{2,t}$  and from  $Y_{2,t-1}$  to  $Y_{1,t}$ . The two equations for dyad  $i$  are

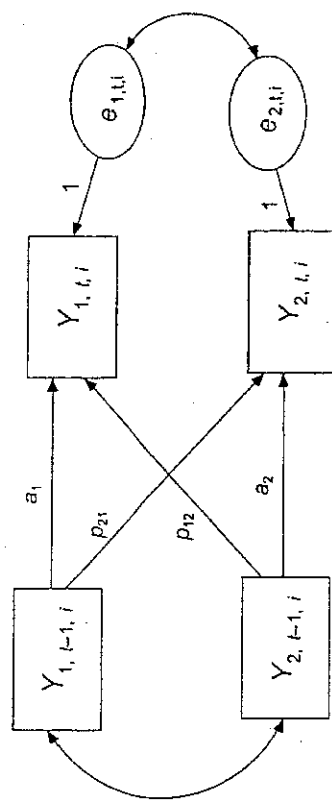


FIGURE 13.1. APIM for over-time data.

$$Y_{1it} = c_{1i} + a_{1i}Y_{1,t-1,i} + p_{12i}Y_{2,t-1,i} + e_{1it},$$

$$Y_{2it} = c_{2i} + a_{2i}Y_{2,t-1,i} + p_{21i}Y_{1,t-1,i} + e_{2it},$$

where  $a_{1i}$  and  $a_{2i}$  represent actor effects and  $p_{21i}$  and  $p_{12i}$  represent partner effects. Actor effects are interpreted as stability effects, and partner effects represent cross-partner influence, or reciprocity (Cook & Kenny, 2005). Because the data are measured over time, the model can be estimated for each dyad, which is the reason the paths in the model are subscripted with an  $i$ . We adopt the convention that  $a_{1i}$  refers to the stability for member 1 in dyad  $i$  and that  $a_{2i}$  refers to the average value of the actor effects for member 1 (e.g., for husbands) across dyads. There are six parameters for each dyad  $i$ :

- Actor effect for person 1, or  $a_{1i}$ .
- Actor effect for person 2, or  $a_{2i}$ .
- Partner effect from person 1 to person 2, or  $p_{12i}$ .
- Partner effect from person 2 to person 1, or  $p_{21i}$ .
- Intercept for person 1, or  $c_{1i}$ .
- Intercept for person 2, or  $c_{2i}$ .

Each of these terms may vary across dyads. Accordingly, if  $a_1$  varies, then that implies that for some 1's (e.g., some husbands), there is more stability than there is for other 1's. Similarly, there might be variance in the partner effects such that individuals in some dyads may be more responsive or reactive to their partners than are individuals in other dyads. Finally, the intercepts may also vary.

It is very important in an over-time model to make certain that we have centered or subtracted off the mean of  $Y_{1,t-1}$  and  $Y_{2,t-1}$ , making sure we use a common value (i.e., the same value is subtracted from person 1's data and person 2's data). Thus we compute a mean for both  $Y_1$  and  $Y_2$  and average them. With centered values for the lagged values of  $Y$ , the intercepts estimate the average or typical value. Centering is also important to make the variance of the six effects interpretable.

These six terms (two actor, two partner, and two intercepts) may also covary with each other. For instance, the intercepts of member 1 and member 2 might be correlated. In principle, there is a six-by-six variance-covariance matrix of terms. Later we discuss different patterns of covariation among these effects.

The two APIM equations for each dyad also include two error terms, one for person 1 and one for person 2, and these two errors might be correlated. This correlation between the errors, or  $r_{ee}$ , tests a key question because it measures the extent to which the two dyad members are especially similar to one another at a particular time point. It is important to distinguish the correlation of the intercepts, or  $r_{cc}$ , from the correlation of the errors. For example, suppose that stress is measured each day for husbands and wives. Then  $r_{cc}$  would measure the extent to which husbands who experience more stress *overall* have wives who also experience more stress *overall*. In contrast,  $r_{ee}$  would measure the extent to which, if the husband experiences more stress on a particular day than one would expect given his prior value and his wife's prior value, then his wife would also experience more stress on that same day than one would expect given her prior value and her husband's prior value. Thus the unit for  $r_{ee}$  is the person, and the unit for  $r_{cc}$  is time within person. Note that either correlation can be negative, and, ideally, the statistical analysis procedure should allow for that possibility.

The saturated cross-lagged regression model is quite complex, having a total of 8 (the 6 effects and 2 errors) variances and 16 covariances (15 for the effects and 1 for the errors). There are also 6 fixed effects: the 2 average actor effects, the 2 average partner effects, and the 2 average intercepts. Thus the total number of potential parameters is 30. However, simpler versions of the model can be estimated. First, it may be the case that the actor, partner, and intercepts do not vary by dyad. For instance, the partner effects may be the same for all persons, and their partner variances would be set to zero. Alternatively, we can set pairs of variances, pairs of covariances, or pairs of fixed effects equal. For example, wife-to-husband

partner variances might be set equal if wives do not differ in the degree to which they influence their husbands. If variances are set equal, it may make sense to force certain covariances to be equal. As discussed in Chapter 4, we can test whether setting a pair of variances or covariances equal is empirically supported by comparing the change in the deviances for the models with and without the equality constraint. In the extreme case, in which we want to treat dyad members as indistinguishable, we would have just 3 fixed parameters (intercept, actor, and partner effects), 4 variances (intercept, actor, partner, and error), and 10 covariances. Thus the complete model with 30 parameters would be reduced to a model with only 17 parameters.

### Analysis Model

In over-time data from dyads, we have three factors: time, person, and dyad. Researchers often make the mistake of considering these data to be a three-level nested model in which time points are nested within persons and persons are nested within dyads. The problem is that time and person are usually crossed, not nested. That is, for a given dyad, the level of time is the same for the two persons at each time point. If the three-level nested model is mistakenly assumed, then the correlation between the two partner's intercepts,  $r_{cc}$ , is constrained to be positive (because it is estimated as a variance), and the correlation between the two members' errors at each time,  $r_{ee}$ , is assumed to be zero.

An alternative approach that does not force these constraints is an extension of the two-intercept model described in Chapters 4 and 7. To apply the two-intercept model in this context, two dummy variables are created. One dummy, what we call M1 (member 1), is set to 1 when the score is obtained from member 1 and set to 0 otherwise. The other, what we call M2 (member 2), is set to 1 when the data are obtained from member 2 and to 0 otherwise. These two terms are then multiplied by the predictor variables (i.e., the actor's and partner's scores for the prior trial). Consequently, for every point in time at which member 1's outcome is the dependent variable ( $Y_{1,t}$ ), member 1's prior score ( $Y_{1,t-1}$ ) is one of the independent variables (representing the actor variable), and member 2's prior score ( $Y_{2,t-1}$ ) is the other independent variable (representing the partner variable). Correspondingly, for every point in time at which member 2's outcome is the dependent variable ( $Y_{2,t}$ ), member 2's prior score ( $Y_{2,t-1}$ ) and member 1's prior score ( $Y_{1,t-1}$ ) are, respectively, the actor and partner vari-

ables. In this way, actor and partner effects can be estimated for the two members.

To allow for correlated errors, we create a variable that takes on the values 1 and -1 to contrast the two dyad members. We call this variable MM. Recall that in Chapter 7, when our distinguishing variable was gender, we created a variable MALE, which was 1 for male partners and 0 for female partners. We also created a similar variable, FEMALE. We then noted that a simpler analysis could be done using the variable GENDER, coded -1 for men and 1 for women. In the current discussion, M1 is analogous to MALE, M2 is analogous to FEMALE, and MM is analogous to GENDER.

### Computer Programs

In this section, we discuss the use of MLM programs to estimate the cross-lagged regression model. We first note that we are creating a data file in which each unit refers to a "person time point." If we have 50 couples measured at 10 time points, we would begin with 1,000 units.

The first step involves creating a variable that we call OBS\_ID, which creates a unique code for each pair of observations obtained from each dyad at each time point. This code is the same for the two dyad members at a particular point in time, but it varies across dyads so that it ranges from 1 (the observation identification number for dyad 1 at time 1) to the number of time points times the number of dyads (i.e., the observation identification number of the last dyad at their last time point). The observation identification code for each pair of observations can be computed using the formula  $OBS\_ID = TIME\_ID + NT(DYAD\_ID - 1)$ , where NT is the number of time points, TIME\_ID is a variable that codes for the time point of the observation (which varies from 1 to T), and DYAD\_ID is the dyad number. Therefore, if 50 couples are measured at 10 times, there would be a total of 1,000 observations, and OBS\_ID would then vary from 1 to 500. However, because one time point for each dyad is lost due to lagging, only 900 observations are actually used in the analysis.

Each data record would contain  $Y_t$ ,  $Y_{a,t-1}$ ,  $Y_{p,t-1}$ , M1, M2, MM, DYAD\_ID, TIME\_ID, and OBS\_ID, where  $Y_t$  is the score at time  $t$ ,  $Y_{a,t-1}$  is the prior score for the person, and  $Y_{p,t-1}$  is the prior score of the person's partner. Thus the data set is structured in a time-as-unit format, sometimes called a *person-period data set*, with one record for each time point for each dyad member. As we stated earlier, the lagged values of Y should be centered, that is, have the grand mean subtracted from them.

There are six variances in the model, and it may be that some of them are very small. Small variances create estimation difficulties for MLM programs. For instance, if there were little or no variance in the partner effects (i.e., the partner effect is essentially the same for all dyads), the computer program may not run. Sometimes the program runs for a very long time but does not converge (i.e., it fails to yield a solution). This is what typically happens with SAS, MLwiN, and HLM. At other times, the program converges, but there are warnings in the solution. This is typically what happens with SPSS. To obtain a meaningful solution, we need to trim the terms that have very small variances out of the model. Typically, problems in estimation are due to the presence of too many terms in the model (Singer & Willett, 2003).

We are faced with a "catch 22." For the program to run, we need to eliminate terms that have zero variance. However, to find those terms, we need to run the program. We suggest the following approach to get around this difficulty: First, the researcher runs a saturated model, a model with all of the terms. If it runs, then there is no problem. However, if there are estimation difficulties, we need to consider running a simpler model. A model with just a few random terms or a model in which the covariances are set to zero (although that option is not currently available with HLM6) can be run.

### SPSS

In Chapters 4 and 7 we discussed the estimation of multilevel models for dyadic data with SPSS. When time is added to the model, the SPSS syntax becomes more complex:

```
MIXED
  Y BY MM WITH TIME_ID YA YP
  /FIXED = MM TIME_ID MM*YA MM*YP | NOINT
  /PRINT = SOLUTION TESTCOV
  /RANDOM MM MM*YA MM*YP | SUBJECT(DYAD_ID)
  COVTYPE(UN)
  /REPEATED = MM | SUBJECT(DYAD_ID*OBS_ID)
  COVTYPE(CSH).
```

where the variables are as follows: Y is the outcome variable, YA is the lagged value of Y for the actor, and YP is the lagged value of Y for the partner. Note that this SPSS code uses MM and is analogous to the syntax for the two-intercept model in Chapter 7 that uses GENDER rather than



MALE and FEMALE. In the fixed section, the MM\*YA term enables estimation of separate actor effects for the two members, and the MM\*YP enables estimation of separate partner effects for the two members. Note also that TIME\_ID is in the model. It represents a linear trend in the data, and it should be centered.

In the RANDOM statement, the corresponding terms (MM\*YA and MM\*YP) enable estimation of variance in the actor and partner effects. The COVTYPE(UN) allows for the six random effects to be correlated. Notice that the time effect is not random. When we discuss growth-curve models later in the chapter, the time effect is random. The final statement is the most complicated statement. It states that for each dyad at each time, the two observations might well be correlated. Using CS allows for that correlation to be negative, and the H of the CSH allows for heterogeneous variances. Instead of the CSH option, we could have also used the unrestricted, or UN, option.

### SAS

Not surprisingly, the code for SAS is similar to that for SPSS. Again, we create an observation identification variable, or OBS\_ID = TIME\_ID + NT(DYAD\_ID - 1), where NT is the number of time points. The SAS code is:

```
PROC MIXED COVTEST;
  CLASS DYAD_ID OBS_ID MM;
  MODEL Y = MM TIME_ID YA*MM YP*MM / NOINT S
         DDFM = SATTERTH ;
  RANDOM MM YA*MM YP*MM / SUB=DYAD_ID;
  REPEATED MM / TYPE=CSH SUBJECT=OBS_ID R;
```

Because there is no intercept in the model (NOINT), the program estimates separate intercepts for the distinguishing variable (MM), as well as separate actor and partner effects.

### HLM

For the computer program HLM, two data sets must be created. First, there must be an observation or a time data set similar to the one created for SPSS and SAS. In addition, there must be a dyad data set that includes two dummy variables for the two dyad members, what we have called M1 and M2. Member would be treated as a repeated measure within this for-

mulation. We can run the two-intercept model that we described in Chapter 4 with HLM. We advise centering the lagged Y variables and time before reading them into HLM and not to allow the program to center the variables.

### MLwiN

One of the main advantages of the computer program MLwiN is that we can place constraints on the covariance matrix. It is easy to set values to zero and to set values equal to each other. The equality constraints are particularly useful in testing differences between distinguishable members. Currently, we see this program as the most flexible program for the analysis of dyadic over-time data. However, as we discussed in Chapter 4, it is not an easy program to use.

### Structural Equation Modeling

Although we have emphasized the use of MLM, SEM can also be used to estimate dyadic over-time models. This estimation technique is particularly useful when the number of time points is relatively small, and it can be used for as few as two (Cook & Kenny, 2005). With this technique, we do not allow for actor and partner effects to vary by person or dyad. However, we can allow for actor and partner effects to vary between different pairs of waves; for example, the actor effect from wave 1 to wave 2 may be greater than that from wave 2 to wave 3. Cook and Kenny (2005) illustrate the possibility that actor effects (i.e., stabilities) increase over time; that is, persons change less as they age.

### Example

We analyze a data set, gathered by Leonard (1984), of dyads that were randomly paired. In the subset of the data considered here, one member of 10 male dyads was given 1.5 ounces of vodka for every 40 pounds of body weight, enough to become legally intoxicated. The other member was not given any alcohol, and thus the dyad members can be distinguished by their alcohol levels. They were then put in a situation in which they chose the level of electric shock to administer to each other, the shock intensity being rated from 1 to 10. There were a total of 25 trials. The data set is rather small, only 10 dyads, and its size creates estimation problems.

First, we create a data set in which the observations are gathered for each person in a dyad at each time; thus, for a data set with 10 dyads, each giving 25 trials, the number of records in the data set would be  $2 \times 10 \times 25 = 500$ . The observed variable is SHOCK level. Next, we create two new variables using the lag transform function. One of these variables is the person's prior behavior at time  $t - 1$ , ASHOCK, and the other variable is the person's partner's prior behavior at time  $t - 1$ , PSHOCK. We then delete from the data set all time-1 observations, because at time 1 there is no prior behavior.

For this analysis we create three dummy variables: INTOX, which is 1 for the intoxicated dyad member and 0 for the sober member; SOBER, which is 1 for the sober person and 0 for the intoxicated person; and INTOXS0B, which is 1 for the intoxicated person and -1 for the sober person. We compute an observation identification number as described earlier (OBS\_ID), and we also have a TRIAL\_ID variable that varies from 1 to 25. The variable DYAD\_ID varies from 1 to 10. As mentioned, the observed variable is shock level, or SHOCK. We computed the mean SHOCK level across all the dyads and all the times ( $M = 5.34$ ) and subtracted that value from the lagged values (i.e., we mean-centered ASHOCK and PSHOCK). Note that we subtracted the same value from both the intoxicated and the sober members. Even though dyad members are distinguishable, we still want the units of measurement of the shock variables to be the same. We centered the TRIAL\_ID variable (i.e., subtracted 13 from each TRIAL\_ID score). Centering the predictor variables is crucial, as we have random intercepts and we want to interpret those intercepts as if they reflect the average response (i.e., the mean) of each person.

We have the traditional APIM framework (see Figure 13.1, as well as Figure 7.1), in which the horizontal paths are actor effects ( $a_5$  and  $a_1$ ) and the crossing paths are partner effects, with  $p_{SI}$  being the path from the intoxicated to the sober member and  $p_S$  being the path from the sober member to the intoxicated member. For the Leonard (1984) data, we can view the partner effect as a measure of retaliation: If the intoxicated member gave the sober member a strong shock, does the sober member respond with a strong shock at the next trial? Because the members are distinguishable, we can test whether the stability effects,  $a_5$  and  $a_1$ , are the same and whether the partner effects, or retaliation effects, are the same,  $p_{SI} = p_S$ . Moreover, we can test whether the intercepts for the sober and intoxicated individuals are equal, a test that evaluates whether the sober or intoxicated member administered more severe shocks. All of these effects are fixed effects.

We are estimating a plethora of random effects for the model. They include:

- Variance of the actor effect of the sober member, or the variance of  $a_5$ .
- Variance of the actor effect of the intoxicated member, or the variance of  $a_1$ .
- Variance of the partner effect of the sober member, or the variance of  $p_{SI}$ .
- Variance of the partner effect of the intoxicated member, or the variance of  $p_S$ .
- Variance of the intercept for the sober member.
- Variance of the intercept for the intoxicated member.

Moreover, the preceding six terms can be correlated, resulting in 15 different covariances. Additionally, there are two error variances:

- Error variance for the sober member.
- Error variance for the intoxicated member.

Finally, these two error variances can be correlated. Accordingly, the complete model contains 8 variances and 16 covariances. The reader should be happy that we did not treat trial as random (but we do so later when we consider growth-curve models).

We first estimated the full model with the 8 variances and 16 covariances. However, we had considerable difficulty fitting such a model. When we ran the model with SAS and HLM6, it never converged, and within SPSS standard errors for some parameters could not be estimated. We also had estimation difficulties with MLwiN and could not get the program to converge. When we estimated a model in which the six effects were uncorrelated, we noted that the variances of the partner effects were smaller than the other effects. Thus it appears that the two partner effects do not vary by dyad. We then estimated a model in which actor effects and intercepts were random (i.e., had a variance) but partner effects were not. When we ran this model we noted that the actor effects were essentially uncorrelated with intercept effects. So in our final model the partner effect was not random, and there were no correlations between intercepts and actor effects. The results of this model, as estimated by SPSS and SAS, are presented in the first column of Table 13.1. The results using MLwiN were very similar, but not exactly identical, to those from SPSS and SAS. The SAS code for the run is:

**TABLE 13.1.** Fixed Effects from the Cross-Lagged Regression Analysis of the Leonard (1984) Study

Fixed effects	Estimate	
	Distinguishable	Indistinguishable
Effect of trial		
Intercept	-0.032*	-0.028*
Sober	5.520*	5.395*
Intoxicated	5.484*	
Actor effect		0.216*
Sober to sober	0.252*	
Intoxicated to intoxicated	0.188†	
Partner effect		0.271*
Sober to intoxicated	0.234*	
Intoxicated to sober	0.316*	

† $p < .10$ ; \* $p < .05$ .

```

PROC MIXED COVTEST;
CLASS DYAD_ID OBS_ID INTOXSOB;
MODEL SHOCK = TRIAL_ID INTOXSOB INTOXSOB*ASHOCK
INTOXSOB*PSHOCK / NOINT S DDFM=SATTERTH;
RANDOM INTOXSOB / SUB=DYAD_ID TYPE=UN GCORR;
RANDOM INTOXSOB*ASHOCK / SUB=DYAD_ID TYPE=UN GCORR;
REPEATED INTOXSOB / TYPE=CSH SUBJECT=OBS_ID R;

```

Note that we include two RANDOM statements, one for the intercepts and one for the actor effects, each with a TYPE=UN, or unstructured covariance matrix. This allows the intercepts to be correlated and the actor effects to be correlated while constraining the intercepts to be uncorrelated with the actor effects. We also add in the RANDOM statements the option GCORR, which gives the correlation of the effects.

To test whether there were differences between the intoxicated and the sober member, we reparameterized the model by creating interaction terms. This required the following SAS syntax:

```

PROC MIXED COVTEST;
CLASS DYAD_ID OBS_ID INTOXSOB;
MODEL SHOCK = TRIAL_ID INTOXSOB ASHOCK INTOXSOB*ASHOCK
PSHOCK INTOXSOB*PSHOCK / SDDFM=SATTERTH;
RANDOM INT INTOXSOB / SUB=DYAD_ID TYPE=UN GCORR;
RANDOM ASHOCK INTOXSOB*ASHOCK / SUB=DYAD_ID TYPE=UN
GCORR;
REPEATED INTOXSOB /TYPE=CSH SUBJECT=OBS_ID R;

```

In this formulation, we do not estimate two intercepts (note that we no longer include the NOINT option), but rather we have all effects interact with the INTOXSOB variable. We note that the two models are the same in that the parameters from one solution can be derived from the other. We also note that the deviances (see Chapter 4) of the two models are identical. We use the first model to obtain estimates for the sober and the intoxicated members, and we use the second model to determine whether there are statistically significant differences for the two types of members.

In all the analyses, we controlled for the linear effect of trial (i.e., the effect of TRIAL\_ID, which was centered and which assesses the tendency to administer more or less severe shocks over time). The shock level declined by about 0.76 units across the 25 trials.

In Table 13.1 we see from the intercepts that the sober and intoxicated members shocked each other at about the same level. There is no statistically significant difference between these two average shock levels,  $p = .90$ . As indicated by the actor effects, there was some degree of stability in both members' behavior: If a member shocked his partner at a high level in the previous trial, then he persisted on the next trial. Again, there is no statistically significant difference in the stability of the intoxicated and the sober persons,  $p = .58$ . The partner effects indicate that both members retaliated against their partners. It appears that the sober members retaliated more, but the difference between the two partner effects is not statistically significant,  $p = .25$ . Note also there is some indication of a couple-oriented model in that actor and partner effects have nearly the same value. Aggression begets aggression, whether the aggression comes from oneself or the other.

In terms of the random components, we see in Table 13.2 that considerable variability exists in the intercepts for both the sober and the intoxicated members. Some people tend to shock their partners more than others. Interestingly, there is a very strong correlation of .975 in these "individual" differences across trials. Thus we conclude that the differences are due to dyad much more than they are due to the person. Having seen these results, we could assume that the two intercepts are equal and then estimate the model with a dyadic intercept and single dyadic variance rather than two individual intercepts and their covariance. To accomplish this in SAS or SPSS, we would not estimate an MM (INTOXSOB) main effect, we would drop the NOINT option, and we would allow the intercept to be a random variable across dyads.

There is some correlation between actor effects: If the sober member is stable, the intoxicated member is also stable. However, the correlation is not statistically significant. Note also that there is little or no correlation

**TABLE 13.2.** Random Effects from the Cross-Lagged Regression Analysis of the Leonard (1984) Study

Random effect	Estimate	
	Distinguishable	Indistinguishable
Variance		
Intercept		2.359*
Sober	1.519†	
Intoxicated	3.035*	
Actor effect		0.058*
Sober to sober	0.082†	
Intoxicated to intoxicated	0.048†	
Error		1.856*
Sober	1.883*	
Intoxicated	1.813*	
Correlations		
Intercepts	.975†	.951*
Actor effects	.433	.603
Errors	-.019	-.036

† $p < .10$ ; \* $p < .05$ .

between the errors. Thus, once the intercepts are controlled, there is no correlation in shock levels for any given trial.

We also present in Tables 13.1 and 13.2 the estimates of the effects of treating the dyad members as if they were indistinguishable, using the program MLwiN to estimate these models. We placed equality constraints where appropriate. When we estimated a model in which there was variance due to the partner effects, that variance was only 0.001; we again set the partner variance to zero. In Table 13.2, we see the estimates of the random effects for the two members. The pattern is similar to that for distinguishable members. The largest systematic variance is in the intercepts: Some persons are more aggressive than others. However, given the strong correlation between dyad members, it is more accurate to say that some dyads are more aggressive than others.

### OVER-TIME STANDARD APIM

In the previous section, we have a version of the APIM in which the prior  $Y$ 's cause the current  $Y$ 's. In the standard APIM only the  $X$ 's, either current or prior, cause the current  $Y$ 's. Consider the following examples:

- Married couples guess each other's emotions at various times (Neyer, Banse, & Asendorpf, 1999; Wilhelm & Perrez, 2004).
- The stress that couples experience during the day determines their moods at night (Thompson & Bolger, 1999).

This model contains actor effects (e.g., Does my stressful day lower my mood?) and partner effects (e.g., Does my stressful day lower my partner's mood?). Such models are very common in diary studies.

There are several issues in analysis of the over-time standard APIM. They include the specification of error structure and measurement of the causal variable. We discuss each in some detail.

### Specification of Error Structure

In some ways, the analysis of the standard APIM is no different from the cross-lagged regression model. We have six random effects of two actors, two partners, and two intercepts. We need to center the  $X$ 's (i.e., subtract off the average of two means across couples and time). We could combine the cross-lagged regression model with the standard APIM, but such a model may be too complex to run. It would have 10 random variables and 45 covariances. There is, however, one key difference between the standard APIM and the cross-lagged model, and that is the problem of autocorrelation in the errors.

In the cross-lagged regression model, prior  $Y$  causes current  $Y$ . The actor effect models the autocorrelation in the data. In the cross-lagged model, the pattern of autocorrelation is called a first-order autoregressive model. In the standard APIM design, the value of  $Y$  at time  $t - 1$  is not a predictor of  $Y$  at time  $t$ . Consequently, the standard APIM needs some way to model the correlation between the current  $Y$  and the  $Y$  that immediately preceded it, and one way of doing that is to allow the errors to be autocorrelated.

One way of modeling the correlations of the error is the model initially suggested by Bolger and Shrouf (in press). In this model, the error structure is defined as a lag 1 autoregressive structure, and in SAS this is accomplished by the TYPE = UN@AR(1) option in the repeated statement. The SAS code for a generic APIM follows. In this code, MM is the variable that codes for the distinguishing factor (e.g., 1, -1 for husbands and wives), XA is the actor's predictor score at a particular time point, XP is the partner's predictor score at that time point, TIME\_ID is the variable that specifies the particular time point, and Y is the outcome for the actor at that same time point.

```

PROC MIXED COVTEST;
CLASS DYAD_ID MM TIME_ID;
MODEL Y = MM MM*XA MM*XP TRIAL_ID / S DDFM=SATTERTH;
RANDOM MM MM*XA MM*XP / TYPE = UN G COORR SUB = DYAD_ID;
REPEATED MM TIME_ID / SUB = DYAD_ID TYPE = UNGAR(1);

```

Note that TRIAL\_ID equals TIME\_ID. So far as we know, this model cannot be estimated using any other MLM program.

### Measurement of the Causal Variable

Consider a study that examines the effect of work stress on mood, both participant mood and partner mood. We might measure the work stress each day and see whether it affects the person's mood that night. However, it is possible, and even likely, that stress from the prior day can carry over to the next day. That is, it might be the accumulated stress that causes mood. Therefore, instead of using today's stress as the predictor, we should use the stress of the previous 3 days, or perhaps a weighted average of 4 days (e.g., .5 times the current day plus .33 times the previous day plus .08 times each of the 2 days preceding). Thus a key concern is how to measure the  $X$  variable. The investigator needs to carefully consider how to compute the  $X$  variable and not just routinely use current or prior  $X$ .

If  $X$  were a lagged composite, we would be lagging back over time. Whenever we lag, we lose observations. Thus, if we have  $T$  observations for a dyad and we lag back 3 days, we would have  $T - 3$  observations for analysis.

### Example

We use the Leonard (1984) data as an example, even though they are not entirely appropriate (because the data really only contain the observations for  $Y$  and there is no  $X$  variable), but it can illustrate the SAS syntax. We estimate the partner effects, the effect of partner's shock level on the previous trial, but not the actor effect. The nonindependence of observations is handled by the autoregressive structure. The SAS code is:

```

PROC MIXED COVTEST;
CLASS DYAD_ID INTOXSOB TIME_ID;
MODEL SHOCK = TRIAL_ID INTOXSOB INTOXSOB*PSHOCK /
NOINT S DDFM=SATTERTH;
RANDOM INTOXSOB / SUB=DYAD_ID TYPE=UN GCOORR;
REPEATED INTOXSOB TIME_ID / TYPE=UN@AR(1)
SUBJECT=DYAD_ID R;

```

The partner effect for the sober member is .269 and for the intoxicated member is .352, and both are statistically significant. The autoregressive coefficient is .088 and is marginally significant. This marginally significant coefficient indicates that the residual for the amount of shock administered at a particular time, controlling for how much the partner shocked the person at the previous time point, is correlated with the residual from the previous time point. The correlation between the intercepts is .920,  $p = .053$ , and between the errors is  $-.045$ ,  $p = .49$ .

### GROWTH-CURVE ANALYSIS

In the prior two sections we examined predictive over-time effects. In this and the next section, we do not consider causal effects, but rather we assume that change occurs in a regular, deterministic fashion. Consider the variable of marital satisfaction, measured on heterosexual couples every 6 months over a period of 2 years, resulting in five measurements. We could fit a straight line to each husband's satisfaction scores. Like any regression line, there are two parameters, a slope and an intercept. The slope represents the rate of change per unit; for example, how much more or less satisfied the husband becomes for each year of marriage. Some husbands may have a negative slope, indicating that they are becoming less satisfied. Others are becoming more satisfied and have a positive slope. Still others may have a zero slope because they showed no change over the course of the study.

The central idea of growth-curve analysis is that we fit a function to a set of observations. The simplest function is a linear, or straight-line, function. With that function, each person has a slope and an intercept parameter. With dyadic data, we can examine the correlation of these parameters. For instance, in couples in which the husband's marital satisfaction is declining, we can determine whether the wife's satisfaction is also declining.

There are two alternative methods of estimating growth-curve models: SEM and MLM. SEM is easiest when dyad members are distinguishable (though see Olsen & Kenny, 2006, for an example of SEM growth-curve model with indistinguishable dyad members), when the temporal spacing is the same for each member, and when there are not too many time points. Before we describe each method, we first discuss two central issues in growth-curve analysis: definition of time zero and functional form.

### Definition of Time Zero

The intercept represents the level of response predicted for "time zero." The researcher must decide how to define the zero point in the time variable. Very often it is set at the time of the first measurement. However, there are alternatives that should be considered. One possibility is to define time zero as the middle of the study; another is to define time zero as the end of the study. Time zero can also be extrapolated to some idealized point before or after measurements were made; for example, the time at which the couple was married. In this case, time zero would be a different point in real time for each couple. Sometimes in developmental research, time zero is defined as the time of birth, or, in other cases, the time of starting school. The researcher has considerable flexibility in determining time zero. The important point to remember is that the first time point need not automatically be time zero.

The choice of time zero affects not only the average intercept but also the variance in the intercepts and the covariance of the slope and intercept. It does not, however, affect the mean of the slopes or the variance of the slopes. Because the choice of time zero affects both the variance of intercepts and the covariance of slope and intercept, it is inadvisable to set either the variance of the intercept or the covariance of slope and intercept to zero. These parameters should be kept in the model even if tests of statistical significance for these two components are nonsignificant. However, if there is evidence that the slopes have zero variance and, accordingly, the slope variance is set to zero, it would be permissible to test the intercept variance and trim it out of the model if it is not statistically significant.

We could choose time zero as the point that minimizes the variance in the intercepts (see Singer & Willett, 2003, pp. 186–187). Note that if there is a time point at which the intercepts have zero variance, we can interpret that point as a point of *common origin*. The time point in which the variance of the intercepts is at a minimum is the time of  $C(a,b)/V(b)$ , where  $C(a,b)$  is the covariance of intercept and slope and  $V(b)$  is the variance of the slope. We take the time measure and subtract  $C(a,b)/V(b)$  from it, and zero becomes the time point at which the intercept variance is at a minimum.

### Functional Form

Typically, growth-curve models are assumed to be linear. That is, the relationship between time and the outcome is a straight line. However, linear-

ity is only an assumption, and the researcher needs to examine this assumption very critically. The researcher should consider alternative functional forms and should statistically evaluate whether there are detectable deviations from linearity. Later in the chapter we discuss how to test statistically for nonlinearity. In this section, we consider alternative functional forms.

There are three ways in which to allow for nonlinearity in the growth curve: (1) to transform the outcome variable, (2) to transform the time variable, and (3) to include nonlinear terms through the use of polynomial. We discuss each in turn.

### Transformation of the Outcome

If the relationship between  $T$ , time, and  $Y$ , the outcome, is not a straight line, one possibility is to turn it into a straight line through the transformation of the outcome variable. Let us consider three examples. First, if the outcome measure is a count (e.g., number of health service visits), one might compute the square root of  $Y$  or take the logarithm of  $Y + 0.5$  to create a functional relationship between  $T$  and transformed  $Y$  that is relatively linear. Second, if the outcome is a proportion (e.g., the number of pills taken divided by the number of pills prescribed), one might want to consider a logit or log odds transform. In this case, we compute  $\ln[p/(1-p)]$ , where "ln" is the natural logarithm and  $p$  is the proportion. A logit removes the floor of 0 and the ceiling of 1 in a proportion. Third, if the outcome is a rate, for example, the number of physician visits during a 1-year period, we might want to consider the reciprocal of the rate measure, or  $1/Y$ . Such a transformation would measure the time interval between visits. Singer and Willett (2003) provide an extensive discussion of possible transformations of the outcome variable.

### Transformation of Time

To straighten a nonlinear relationship, we can transform the predictor variable instead of the outcome variable. For instance, instead of computing the square root of the  $Y$ , we could instead compute the square of  $T$ . Basically, if the function is accelerating (i.e., more change at later time points),  $T$  would need to be raised to a power larger than 1 (i.e., squared), but if the function is decelerating (i.e., more change at earlier time points),  $T$  would need to be raised to a power less than 1 (e.g., square rooted or reciprocal). The potential advantage of transforming  $T$  instead of  $Y$  is that

the results may well be more interpretable because the outcome variable is left in its more "natural metric."

Consider the functional form of a negative exponential<sup>2</sup> in which the outcome variables decelerate toward an asymptote. To model this nonlinearity, we can create exponential time (Kenny et al., 2004). We transform  $T$  by the function  $-p^T$ , where  $p$  is a specific value between 0 and 1. One way of approximating  $p$  would be to take the ratio of the amount of change from time 1 to time 2 to the amount of change from time 0 to time 1. Thus  $p$  measures the rate of deceleration. Greater values of  $p$  correspond to weaker deceleration. The researcher can experiment with different values and determine what value best describes the pattern of change. Note that with exponential time, time 0 becomes the asymptote. Consequently, for this growth model, the intercept refers to the theoretical limit that each person or dyad would eventually reach if they lived forever. The slope would represent the total amount of change from the beginning of the study to the asymptote.

A situation for which we might expect exponential change occurs when regression toward the mean (see Campbell & Kenny, 1999, especially Chapter 8) is plausible. Because scores in many dyadic studies are selected to be extreme, regression toward the mean is all but inevitable. Consider the example of marital satisfaction measured over time. For instance, it is known that for the typical couple, satisfaction declines after marriage (e.g., Kurdek, 1998). Because satisfaction is relatively high at the point of marriage, much of this decline can be attributed to regression toward the mean. As another example, marital satisfaction is also often tracked in marital therapy. Because marital satisfaction is relatively low at the point at which couples enter therapy, much of the improvement of couples in therapy is also due to regression toward the mean and not due to the effects of therapy. In both of these cases, we would expect regression toward the mean—in the first example, downward (the loss of newlywed bliss), and in the latter case, upward (hitting bottom and going to therapy). If regression toward the mean is at all plausible, the researcher should consider exponential growth or decay. Note that regression to the mean over time implies nonlinear growth because there is more change earlier than later.

### *Polynomial Regression*

The standard approach to nonlinearity (e.g., Francis, Fletcher, Stuebing, Davidson, & Thompson, 1991) is to add polynomials to the model, for

example, the square of time and the cube of time. Thus, if we denote time as  $T$ , we enter  $T$ ,  $T^2$ , and  $T^3$  in the model. Such an approach is costly in terms of parameters because we would have to estimate the means of each component, as well as their variances and all of their covariances. For instance, if we allow a squared term, we add three parameters; and we allow for a squared and a cubed term, we add seven parameters.

The difficulty with fitting polynomials, in addition to the large number of parameters, is that it is an empirically based method of estimating a growth curve and not a theoretical one. Because the approach capitalizes on chance, the probability that replication would fail is increased. Given a new sample, the findings may change dramatically when polynomials are used to model nonlinearity.

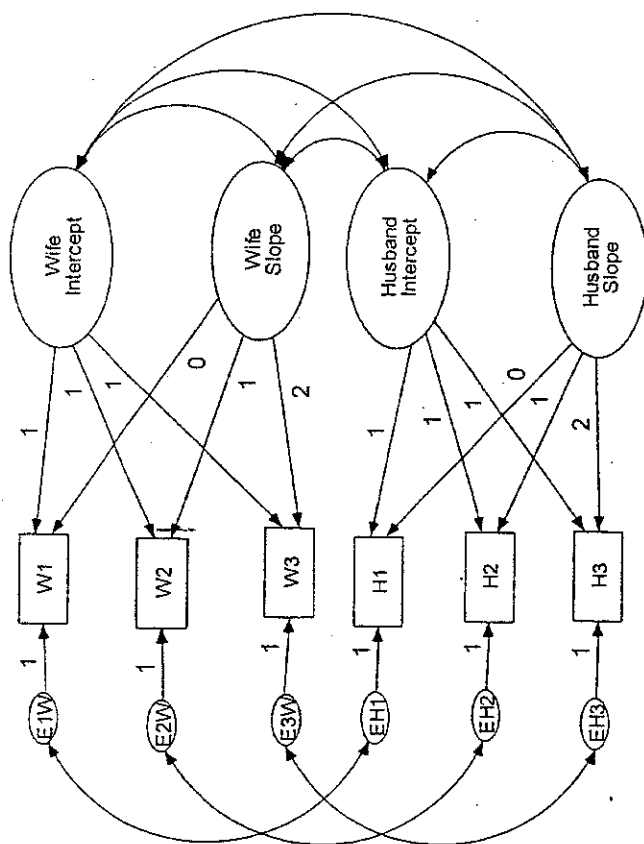
### **Structural Equation Modeling**

Both SEM and MLM have been proposed for the analysis of growth-curve data. The major advantage of MLM is that the number of time points and the spacing between time points can be different for each dyad, whereas SEM performs best when the spacing is the same for each dyad. In this section we introduce growth-curve models within SEM.

Figure 13.2 presents a path diagram for a model of growth in husbands' and wives' marital satisfaction scores measured at three points in time. The essential features of this model are the four latent variables, represented by large circles; the six observed variables, represented by squares; and the six residual factors, represented by the small circles. The model specifies that each of the observed variables is caused by three factors: (1) an intercept, or constant; (2) a rate of change, or a slope; and (3) unknown residual causes. There are several correlations in the model. As is typical in growth-curve models, the intercept and slope are correlated. Also, slope and intercept factors are correlated across members. Finally, with dyadic data, it makes sense to correlate the errors of the same measure at the same time from two different persons. All nine correlations are represented in Figure 13.2 by curved lines.

As is typical of SEM for dyadic data, we analyze a dyadic data set with all six measures in each record (H1, H2, H3, W1, W2, and W3). For growth-curve models, we analyze the variance-covariance matrix and the means. Notice that each of the loadings for the latent variable intercepts are all fixed at 1.0. To interpret this intercept value, we need to know what time zero is. Time zero is determined by the time point that has a zero loading on the slope factor. We see in Figure 13.2 that there are zero load-





**FIGURE 13.2.** Growth-curve model for husbands and wives.

ings on the slope factors at time 1. Under this specification, the mean of the intercept factor refers to the predicted marital satisfaction at the first time point. If we were to change the slope loadings to  $-1$ ,  $0$ , and  $+1$ , the mean of the intercept factor would refer to the expected value of the middle wave. Finally, if we were to change the slope loadings to  $-2$ ,  $-1$ , and  $0$ , the mean of the intercept factor would refer to the expected value for the last wave. Recall that there is a variance associated with the intercept factor that represents the degree to which, for example, husbands (or wives) differ from dyad to dyad in the average value of their satisfaction at time zero. By making changes in the designation of time zero, the variance of the intercept changes, as does the covariance of slope and intercept. In fact, by changing time zero, it is easy to change the sign of the covariance between slope and intercept (see Singer & Willett, 2003).

To interpret the latent variable for slope, it is necessary to examine the factor loadings of that factor. What is not explained by the constant factor (i.e., the intercept) is the degree to which the three observed variables differ from each other, or, in other words, whether husbands' marital satisfac-

tion changes over time. This information is captured by the slope factor. Observe that the loadings for the slope factor go from 0 to 1 to 2, a simple linear pattern. These loadings imply that the level of husbands' marital satisfaction at time 2 is greater (or smaller) than it was at time 1, and husband's marital satisfaction at time 3 is greater (or smaller) than it was at time 2. If most husbands' satisfaction scores decrease, the mean of the slope factor would be negative. The factor mean for the slope is interpreted as the average level of change in marital satisfaction for one unit in time for the average husband and is comparable to an unstandardized regression coefficient and interpreted accordingly. For example, if the average slope for husbands is 1.288, then for each 6 months (the time between measurements), husbands' marital satisfaction increases 1.288 units on average. The variance associated with the slope factor measures the extent to which the slope is the same for all husbands. If the slopes were the same for all husbands, the variance would be zero. If the marital satisfaction of some husbands has grown (or declined) at rates different from that of others, the variance would be nonzero. In fact, growth-curve analysis of marital satisfaction indicates that for both husbands and wives, marital satisfaction declines over the first few years of marriage (Karney & Bradbury, 1997; Kurdek, 1998).

A comprehensive growth-curve analysis should include a test of nonlinearity, which implies either acceleration or deceleration of change. It addresses questions such as: Does the rate of decline in wives' and husbands' marital satisfaction increase or decrease over time? As shown in Figure 13.2, the slope factor was estimated by measures loading 0, 1, and 2 on a latent growth variable. We can test for nonlinearity by freeing one of the loadings—for example, the loading of 2—for both members and observing what happens to the fit of the model. If the fit improves by freeing up the loading, one can interpret the effect as nonlinearity or as a transformation of the time variable. Therefore, if the linear loading that was forced to be 2 before were now estimated as 1.5, then we would conclude that time "slowed down" between the second and third wave of the study. If, however, it were 3.0, change would be accelerating.

Alternatively, a second growth-factor slope could be estimated for each dyad member. This slope factor would measure the quadratic growth function. With four time points, the loadings could be chosen from a table of orthogonal polynomials (e.g., 1,  $-1$ ,  $-1$ , 1) by squaring the linear loadings (e.g., 0, 1, 4, 9) or by centering and then squaring (e.g.,  $-2.25$ ,  $-0.25$ ,  $0.25$ ,  $2.25$ ). Note that although these loading options are different, they are fundamentally the same. More complex curves corresponding to more

complex growth processes (e.g., cubic functions) may also be modeled. The quadratic factor may have a variance, and it can be correlated with the intercepts and linear slope factors. To be able to estimate a simple linear growth curve minimally requires three waves of data. The estimation of curvilinear growth patterns requires four or more waves of data.

With dyadic data, there are two growth curves, one for each dyad member. An interesting empirical question is the degree to which the members of the couples have the same or different growth curves: For example, do husbands and their wives have the same slopes or intercepts? We might ask, for instance, the following questions about marital satisfaction: Is the decline in marital satisfaction, as measured by the slope, greater for husbands or for wives? Additionally, is the level or average of marital satisfaction, as measured by the intercept, greater for husbands or for wives? One should not assume a priori that the growth curves are identical. Rather, this is an assumption that should be tested by determining whether the correlation between the factors is 1. We ordinarily estimate two models, one that constrains the correlation to 1 and another that leaves it free. If the correlation is less than 1, the latter model should fit better than the former.

The slope and intercept factors can be treated as outcome variables. For example, we can examine whether age of the spouses predicts either their slopes or intercepts. As is generally advisable in prediction, the predictors should be centered variables, or, at the very least, zero should be a meaningful value.

We can also use the slopes and intercepts from one variable to predict the slopes and intercepts of another variable. For example, in Raudenbush and colleagues (1995), the growth in husbands' and wives' job-role quality was used to predict growth in husbands' and wives' marital satisfaction. Specifically, it was hypothesized that growth in a husband's marital satisfaction might be caused by either growth in his own job-role quality or growth in his wife's job-role quality. Similarly, it was hypothesized that growth in a wife's marital satisfaction is caused by either growth in her job-role quality or growth in her husband's job-role quality.

The reader may recognize the essential ingredients of the APIM in the Raudenbush and colleagues (1995) study. The extent to which growth in husband's marital satisfaction is predicted by growth in his job-role quality is an actor effect, and the extent to which growth in his marital satisfaction is predicted by growth in his wife's job-role quality is a partner effect. Raudenbush and colleagues found significant actor effects (i.e., growth in job-role quality predicts growth in own marital satisfaction) but no partner

effects. Kurdek (1998) has also combined the APIM and growth-curve analysis to test whether growth in husbands' and wives' depressive symptoms predicts change in their marital satisfaction. He also found significant actor effects for both husbands and wives (i.e., growth in one's own depression predicts decline in one's marital satisfaction) but no reliable partner effects. These studies demonstrate that the APIM and growth-curve modeling are not competing methods of conceptualizing and analyzing interpersonal relationship data but, rather, address different, complementary questions and, as in the preceding cases, can be profitably combined.

### Multilevel Modeling

Due to its multilevel nature, growth-curve analysis is often performed using MLM software such as MLwiN (Rasbash, Steele, Browne, & Prosser, 2004), HLM6 (Raudenbush et al., 2004), SAS, or SPSS. In the context of growth curves, level 1 refers to time, and level 2 refers to dyad. The level-1 models for the two dyad members' scores for dyad  $i$  at time  $t$  would be

$$Y_{1it} = c_{1i} + b_{1i}T_{it} + e_{1it},$$

$$Y_{2it} = c_{2i} + b_{2i}T_{it} + e_{2it},$$

where  $T$  is a time variable,  $c$  is the intercept, and  $b$  is the slope. Note that both slopes and intercepts may be random variables and may have variances and covariances. We use a person-period data set; that is, each observation refers to one time point of one person.

Earlier we introduced a dummy variable  $M1$  that is 1 for person 1 and 0 for person 2. We had a similar variable,  $M2$ . The variable  $MM$  was 1 for person 1 and -1 for person 2. Using these variables, the SPSS code for a model that allows for different slope and intercepts for the two members is

```
MIXED
Y BY MM WITH TIME_ID M1 M2
/FIXED = M1 M2 M1*TIME_ID M2*TIME_ID | NOINT
/PRINT = SOLUTION TESTCOV
/RANDOM M1 M2 M1*TIME_ID M2*TIME_ID | SUBJECT(DYAD_ID)
COVTYPE(UN)
/REPEATED = MM | SUBJECT(DYAD_ID*TIME_ID) COVTYPE(CSH).
```

Note that this is the saturated model that treats the slope and intercepts for the two members as random variables that may be correlated. The last line

allows for correlated errors across dyad members, what we called  $r_{ec}$  earlier in the chapter.

The comparable SAS code is

```
PROC MIXED COVTEST;
CLASS DYAD_ID OBS_ID MM M1 M2;
MODEL Y = M1 M2 M1*TIME_ID M2*TIME_ID
      /S DDFM=SATTERTH NOINT;
RANDOM M1 M2 M1*TIME_ID M2*TIME_ID
      / SUB=DYAD_ID TYPE=UN;
REPEATED MM / SUB=OBS_ID TYPE=CSH R;
```

### General Error Models

With over-time data, responses are nonindependent because the data come from the very same person (i.e., autocorrelation). This nonindependence implies that the elements of the variance-covariance matrix of errors are nonzero. The cross-lagged and growth-curve models have specialized models of the variance-covariance matrix of errors. However, these are only two of many possible models than can be specified. In terms of the extremes, there could be models in which no correlation in errors occurs or models in which the variances of the errors are different at each time and all the pairs of covariances are different.

Singer and Willett (2003) present an extended discussion about the choice of error model. For instance, one classic model, the one assumed by repeated-measures ANOVA, is that of compound symmetry (see Chapters 6 and 10): The variances are equal to the same value, and the covariances are equal to the same value. The most general model of errors is the model that is unstructured (i.e., covariances of different pairs of errors are allowed to be different), the model assumed by MANOVA. For this structure, if there were 10 time points, there would be 10 variances and 45 covariances, and all of them would be free parameters. For a model that places no constraints on the covariance of errors, we cannot estimate any random over-time effects. Consequently, we could not allow for individual differences in growth-curve parameters nor allow for actor effects in the cross-lagged regression model. It is only by placing constraints on the structure of errors that we can estimate these parameters.

Singer and Willett (2003) have claimed that the choice of error structure does not greatly affect the estimation of the fixed effects. We agree that it does not usually have much of an impact on the estimates of the fixed effects, but it can have a dramatic impact on the standard errors. For

instance, it is well known that if the assumption of compound symmetry is violated, standard errors can be grossly inflated (Maxwell & Delaney, 2004).

### Example

We reanalyzed the Leonard (1984) data using MLM of growth curves. As before, we centered the time variable, and thus time zero is defined as the midpoint of the study, or the 13th trial. Note that if we had left time as originally coded (from 1 to 25), the intercept would refer to a point before the study began. The SAS code (variables defined earlier in the chapter) is as follows:

```
PROC MIXED COVTEST;
CLASS DYAD_ID OBS_ID INTOXSOB;
MODEL SHOCK = TRIAL_ID INTOXSOB TRIAL_ID*INTOXSOB /
      NOINT S DDFM=SATTERTH;
RANDOM INTOXSOB TRIAL_ID*INTOXSOB / TYPE = UN
      SUB = DYAD_ID GCORR;
REPEATED INTOXSOB / SUB=OBS_ID TYPE=CSH R;
```

Using MLM, we first estimated separate curves, both slope and intercept, for both the intoxicated and sober partners in the 10 dyads. However, we had difficulty getting the model to run with multilevel programs (e.g., SAS and SPSS). Moreover, when the program did run, we found evidence for variation across dyads in the slope and intercept parameters for both members, and we found a virtually perfect correlation between the two intercept parameters ( $r = .978$ ) and between the two slope parameters ( $r = 1.000$ ). Thus it seemed sensible to treat the data as if there were a common slope and intercept parameter for each dyad.

The code for the analysis in which both partners have the same slope and intercept but in which these common slopes and intercepts are allowed to vary across dyads, using SAS's PROC MIXED, is

```
PROC MIXED COVTEST ;
CLASS DYAD_ID OBS_ID INTOXSOB;
MODEL SHOCK = TRIAL_ID / S DDFM = SATTERTH;
RANDOM INTERCEPT TRIAL_ID / TYPE = UN SUB = DYAD_ID
      GCORR;
REPEATED INTOXSOB / SUB=OBS_ID TYPE=CSH R;
```

We also estimated the same model using SPSS and MLWIN.

The results of the model with a common growth curve for the sober and intoxicated members are as follows: The overall intercept is 5.316. Recall that the range of possible responses is from 1 to 10, which puts the intercept near the scale midpoint of 5.5. The intercept refers to the predicted level of shock given on at the 13th trial. The average effect of time is  $-0.041$ . This implies a decline across the 25 trials of 0.98 points on the shock scale. Although the correlation between the slopes for the sober and intoxicated members was perfect, the mean of the slopes was different for sober and intoxicated members. The decline for intoxicated members' shock level was 1.70, whereas the sober members declined only 0.27 points. It is confusing that the two slopes differ yet their correlation is perfect. Note that a perfect correlation implies relative and not absolute equality.

Both the intercept and slope varied from dyad to dyad, and their variances were statistically significant. Thus some dyads tended to increase shock levels and others tended to decrease them. The correlation between slope and intercept is .572, indicating that those who lowered their shock levels more over time tended to have lower shock scores at trial 13 and those who increased shock levels had higher shock levels at trial 13. This correlation, however, is not statistically significant,  $p = .155$ . The correlation between errors in the shock levels was  $-.054$ , and this small correlation is marginally significant,  $p = .061$ .

The estimate of the variance of the intercepts at trial 13 is 8.50. Note that there is about four times as much intercept variance as error variance, and, accordingly, dyad explains about 80% of the variance in shock level. If we were to change time zero to the first trial, the intercept variance would be 6.26; if we were to change time zero to the last trial, the variance becomes 12.46. These results suggest that the dyad differences in aggression are widening over time. This is what some have called *fan spread* in that dyads differ more later on than they differ initially.

Finally, we also tested for nonlinearity by including a square term for trial (i.e., time). Although there was evidence of a small amount of curvilinearity, such that the shock level went up and then down, for ease of presentation we presented only the simpler linear model.

### CROSS-SPECTRAL ANALYSIS<sup>3</sup>

In cross-lagged regression analysis, the role of time is primarily to establish which behavior came first and which followed; thus temporal precedence

in the causal or influence process is the focus of the analysis. Time plays a much more important role in some processes, however. For instance, an infant may enjoy the playful attention of a caregiver—up to a point—but may eventually become overstimulated and withdraw from interaction. Thus, over time, the caregiver's attention is transformed from something positive to something aversive to the infant (Brazelton, Koslowski, & Main, 1974). If only the infant's behavior were observed, one might find that there is a pattern in the data such that for some period of time the infant is engaged with the caregiver, followed by a period of disengagement, followed by a period of engagement, and so on. In other words, the behavior ebbs and flows, or cycles. Figure 13.3 illustrates such a cyclical pattern in mother-infant interactions. The Y-axis is affective tone of expressive behavior, ranging from negative engagement (e.g., averting gaze) to neutral to positive engagement (direct gaze), and the X-axis is time, measured in seconds. Notice that the Y variable must be measured on an interval scale for the gradual rise and fall (or wave) to be observable.

If we tried to fit a linear regression line to this plot with time as the independent variable, the slope of linear time would be near zero. Every rise (i.e., more positive involvement) is counterbalanced by a fall (i.e., more negative involvement, or disengagement). Thus there is no consistent linear trend in the data because the data are nonlinear. Although it is not obvious, the data are so nonlinear that they cannot be modeled by the squared and cubed (i.e., polynomial) functions. Waves or cycles in over-time data are typically modeled by the sinusoid functions (i.e., the sine and the cosine).

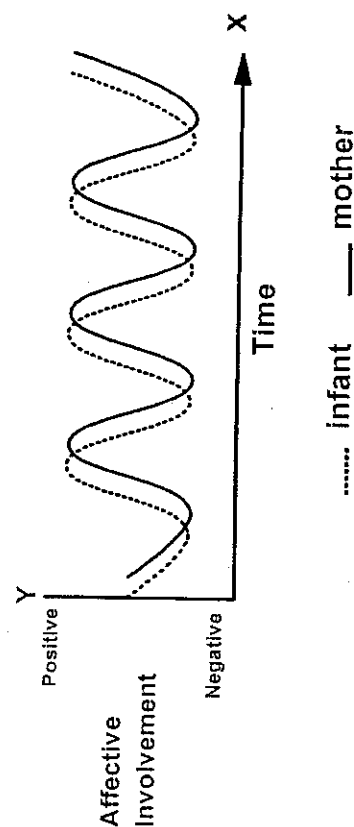


FIGURE 13.3. Cycles in mother and infant behavior.

The presence of cycles in the data implies a very particular correlational structure. Imagine blood pressure data measured every hour, and assume that these data have a strong daily cycle. More temporally adjacent observations—for example, an hour apart—would be very highly correlated. However, as the time lag increases, the correlation weakens and becomes strongly negative, becoming most negative with a lag of 12 hours. Then, for longer lags, the correlation strengthens, and for a lag of 24 hours the correlation would be very strong. Cycles are seen in the correlations, as well as in the raw data.

In dyadic analysis we are not so much interested in the cycles in one person's behavior as in the degree to which the behavior of two persons cycle together. A strong association between the cycles for two people is called *synchrony*. Lester and colleagues (1985) studied synchrony of affective involvement in 40 mother-infant dyads when the infants were 3 and 5 months old. The scale measuring affective involvement ranged from 1 (avoidant behavior) to 13 (positively engaged). In 20 dyads, the infant had been born prematurely, and in the other 20 dyads, the infant was full term. The researchers hypothesized that there would be greater synchrony in the mother-infant dyads with full-term infants than in those with premature infants because of the full-term infants' greater neurological development and information-processing capacity. Cross-spectral analysis is the method used to test their hypothesis. However, before one can investigate the association of two persons' cycles using cross-spectral analysis, one must first have adequate descriptors of each individual's cycling. We therefore discuss the analysis of individual cycles using harmonic analysis and spectral analysis.

Three parameters are involved in the description of a sinusoid. The horizontal distance from the top of one wave (i.e., the peak) to the top of the next is called the *period*. It is the length of the cycle measured either in number of observations or in units of time, and it is symbolized by  $\tau$ . For example, Lester and colleagues (1985) identified a period for infant affective involvement that was approximately 10 seconds long; 10 seconds elapsed from peak of positivity to peak of positivity. The *frequency* is measured by the inverse of the period, or  $1/\tau$ , and is symbolized as  $\omega$ . So if the period for the infant's affective involvement were 10 seconds, the frequency would be  $(1/10)$ , or .10 cycles per second. One cycle per second is called a hertz and is abbreviated as Hz. (The reader probably is familiar with the term megahertz, from audio equipment.) The *phase* of the sinusoid locates the peak relative to the zero point on the time scale. Finally,

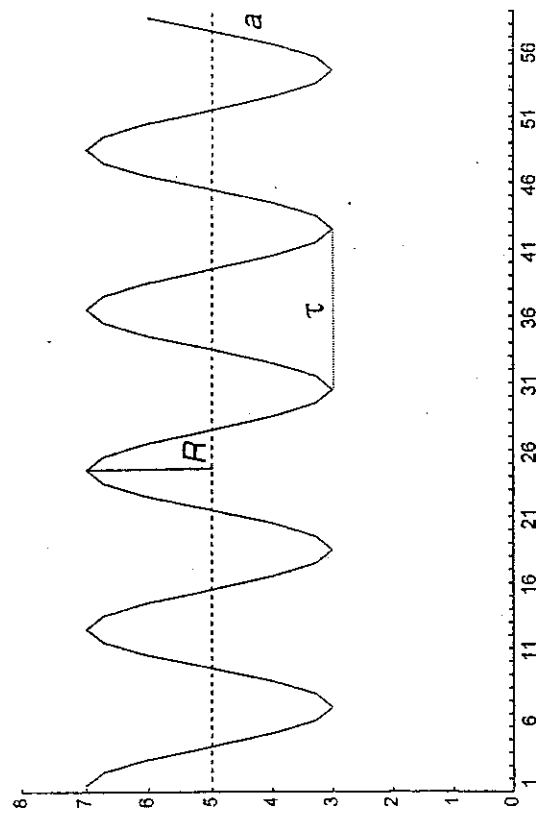
the amplitude is the height of the peak. It is the distance from the mean of the waveform (halfway between the peak and trough) to the peak, measured in the unit or scale of the outcome variable. Amplitude is analogous to the slope in regression analysis. It measures the ability of the sinusoid to explain variance in the response. If the frequency, phase, and amplitude are known, then the sinusoidal cycle can be computed.

Suppose that on the basis of the Lester and colleagues (1985) findings, we hypothesize that there is a 10-second period in infant affective involvement. We can test this hypothesis by fitting a sinusoid function to the infant's behavior observed over time, a procedure called *harmonic analysis*. Harmonic analysis is accomplished by creating predictor variables that represent the sine and cosine of time for a given frequency, denoted as  $\omega$ . The equation (adapted from Warner, 1998) is as follows:

$$Y_t = a + b[\sin(\omega t)] + c[\cos(\omega t)] + e_t$$

The equation is not nearly as complicated as it looks, and we have graphed it in Figure 13.4. In this equation,  $Y_t$  is the value of the infant's affective involvement at different times ( $t = 1, 2, \dots, T$ , or 1 to 60 in Figure 13.4) in the series. It is an ordinary multiple regression equation with an intercept (the typical value of  $Y$ ),  $a$  (see Figure 13.4), and two predictors, the sine and cosine functions. There are two regression coefficients,  $b$  and  $c$ . What is complicated are the two predictor variables, one involving a sine function and one involving a cosine function. They convert the time variable ( $t$ ) into a sine wave and a cosine wave using the frequency, denoted as  $\omega$  (or  $1/\tau$ ), transforming the frequency into an angle (Warner, 1998). The combination of these two waves creates a single wave that accounts for variance in  $Y$ .

By fitting both the sine and cosine waves to the data, one obtains two coefficients,  $b$  and  $c$ . These coefficients are used to compute the amplitude and phase of the cycle. The amplitude of the cycle,  $R$  (see Figure 13.4), is equal to  $b^2 + c^2$ . The phase denoted as  $\phi$  is equal to  $\arctan(-c, b)$ . The phase is important to the analysis of interpersonal influence and is to be discussed later. A large amplitude means that the sinusoid for the period (in this case, 10) explains much of the variability in the person's scores. If the hypothesized period were incorrect (i.e., the behavior cycles at a different frequency or does not cycle at all), the amplitude would be small. Including the sine and cosine wave functions in the equation removes from the outcome variable any variance due to a cycle of the specified length in the



**FIGURE 13.4.** Graph of a sinusoid function with a period ( $\tau$ ) of 12, intercept ( $a$ ) of 5, and amplitude ( $R$ ) of 3.

data (in this case, every 10 seconds), thus correcting for nonindependence of observations due to that cycle. Warner (1998) provides a more detailed explication of a harmonic analysis.

With most time-series data, the length of the cycle is not known. Part of what spectral analysis does, as opposed to harmonic analysis, is to determine empirically what periods (there may be more than one) explain the most variance in the data (i.e., have the largest amplitudes). For a given series with  $T$  time points,  $T/2$  cycles can be estimated. Each of these cycles is tested for statistical reliability, thus identifying potentially important cycles. For example, Lester and colleagues (1985) found reliable cycles in infant affective involvement at three different frequency ranges. We already mentioned one cycle that was approximately 10 seconds long. This is a relatively fast cycle. They also found cycles of approximately 0.05 Hz (midrange cycles with periods of roughly 20 seconds) and cycles of approximately 0.025 Hz (slow cycles with periods of roughly 45 seconds).

As mentioned earlier, the larger goal is understanding the dynamics of dyadic behavior—in this case, associations between partners' cycles. Lester and colleagues (1985), for example, hypothesized about the synchrony of

affective involvement in mother–infant dyads and how it might differ for preterm and full-term infants. If one is interested in whether the behavior of two people is coordinated or coupled, it is necessary to measure the degree to which the two sets of waves rise and fall together, like the waves in Figure 13.3. The estimation of the coupling in the waves is done using cross-spectral analysis. Cross-spectral analysis involves taking the cross-product of the mother's and infant's amplitudes for selected frequencies (viz., those identified as reliable). These cross-products are used in the calculation of covariances, from which we can determine the amount of variance in one person's cycle that overlaps with the other person's cycle. This measure of correspondence between two sets of waves is called *coherence*. If there were high coherence, then the two waves would move together and have the same cycles.

When there is high coherence between the two sets of data (i.e., they are in phase with each other), we say their behavior exhibits *synchrony* (Bakeman & Gottman, 1997). For example, we have already determined that the infant has a 10-second period. If mother and infant behaviors were coupled, the mother would also have a 10-second period. If this were so, then the measure of coherence for the 0.10 Hz frequency would be high. Note, however, that this coherence is specific to the 10-second period. In cross-spectral analysis, coherence is measured for each of the  $T/2$  frequency bands in an exploratory fashion. Thus the investigator must choose which of these frequency bands should be interpreted. As mentioned earlier, interpretation is usually limited to only those frequencies that are relatively powerful predictors (or descriptors) of individual behavior (i.e., that have a high amplitude).

It is also possible to estimate the extent to which one person's waves systematically lead or lag behind those of the partner, suggesting a process of interpersonal influence. If the amount of time the mother makes eye contact with her infant begins to decrease after she notices that the infant has begun to avoid eye contact with her, the cycles describing her behavior would lag slightly behind those of her infant. The formula for phase (earlier called  $\phi$ ) is central to the determination of lag and lead. Lag (or lead) refers to the difference in phase at which the two individual cycles have the greatest coherence. In other words, coherence is estimated while controlling for the difference in the phase of the two cycles. As can be seen in Figure 13.3, the infant's behavior changes first, followed shortly by a change in the mother's behavior. Lag–lead relationships between sets of cyclical data are also analyzed using cross-spectral analysis. Warner (1998) provides a clear and comprehensive introduction to these methods.

We need to be clear about causation in lead-lag relationships. It is not at all clear who is influencing whom. Suppose the mother and child both show a cycle length of 10 seconds, and the peaks in the infant's cycles lead the peaks in the mother's cycles by 3 seconds. Under certain circumstances you might view the infant's behavior as pulling the mother's behavior along, and under other circumstances, you might view the mother's behavior as pushing the infant's behavior along. In the absence of strong theoretical guidance, it is best to say that the two are entrained in a shared cycle. Nonetheless, theory and common sense may dictate that one person be viewed as influencing the other.

A fundamental assumption of cross-spectral analysis is that both time series are stationary. That is, the mean, variance, and covariances do not change over time. Boker, Xu, Rotondo, and King (2002) discuss the relaxation of this assumption.

### NONLINEAR DYNAMIC MODELING

There has been increasing interest in nonlinear dynamic models of dyadic data (Felmlee & Sprecher, 2000). Here we discuss the approach that Gottman, Swanson, and Swanson (2002) have developed, because it has been featured most prominently. As described by Gottman and colleagues, nonlinear dynamic modeling is a qualitative method that uses differential equations to plot the influence two partners have on each other. It is idiographic in that the influence of one person on another is measured for each person. We present a brief description of the technique, and we refer interested readers to Gottman and colleagues for more details. We assume that  $Y_{1t}$  for person 1 and  $Y_{2t}$  for person 2 are measures of positivity–negativity. (We omit here the subscript for dyad.) Lower scores would indicate negative behavior, and higher scores would indicate positive behavior.

There are two equations, one for each partner's outcomes. The outcome variable in each equation (e.g., person A's behavior at time  $t$  or  $Y_{1t}$ ) is a function of the intercept of the regression equation, which is denoted as  $a$ , and is referred to as the person's *uninfluenced steady state* or *emotional inertia*. Additionally, a person is influenced by his or her past behavior,  $Y_{1,t-1}$ . In equation form we have

$$Y_{1t} = a + rY_{1,t-1} + e_t.$$

The intercept and the autoregressive coefficient,  $r$ , are estimated using ordinary regression analysis. Note that the partner's behavior is not entered into the equation.

In the next step, the partner's influence is calculated separately at each point in time over the course of the interaction. This involves subtracting the uninfluenced steady-state component,  $a$ , and the effect of the person's own past behavior,  $rY_{1,t-1}$ , for each of the time-specific observations of person 1. This leaves, for each point in time, a residual effect, or  $e_t$ ; that is, the component of the dependent variable (person 1's score) that is not accounted for by either the uninfluenced steady state (what person 1 brings to the interaction) or person 1's emotional inertia. The dynamic-systems model specifies that this residual effect  $e_t$  is due to the prior behavior of the partner, or  $Y_{2,t-1}$ . However, we take  $Y_{2,t-1}$  and break it up into categories, create dummy variables, and estimate the effects of "discretized"  $Y$  on  $e_t$ . We denote these effects as  $I_{12}$ , or the influence of person 2 on person 1.

For instance, if the scale for the partner's behavior,  $Y_2$ , ranges from 0 to 4, the  $I_{12}$  coefficient can be estimated for every level of  $Y_2$ , producing four different values. We can then graph these effects where the X-axis is the level of  $Y_2$  and the Y-axis is the mean value of the residual. Each of these means is an influence effect,  $I_{12}$ . In this manner, one can qualitatively evaluate whether the degree of influence changes for different levels of the partner's prior behavior. If interpersonal influence is nonlinear, the function on the graph is not a straight line. For example, the effect of the partner may be greater when his or her prior behavior is at higher values (i.e., more negative) than at lower values. It may also be that the partner has no influence at all until he or she manifests a certain level of negativity, which would be indicated by the point at which the line begins to slope upward. This point is called a *negativity threshold*. A second component reflecting sensitivity to positive behavior (called a *positivity threshold*) can also be determined. A parallel analysis can be conducted using person 1 to predict person 2.

In several papers, Gottman and colleagues (2002, 2003) have suggested a simpler version of their model. They propose what is called a *bilinear model*, in which partner's lagged effect is linear but the effect is different when lagged  $Y$  is positive than when it is negative. We discuss the estimation of the bilinear model using an APIM framework.

First, two different predictor variables are created. The variable  $Y_p$  equals  $Y$  when  $Y$  is positive and elsewhere is zero; the variable  $Y_n$  equals  $Y$



when  $Y$  is negative, or else it is zero. To test whether the slopes of  $Y_p$  and  $Y_n$  are different—bilinearity—we estimate a model with lagged  $Y$  and  $Y_p$  (or  $Y_n$ ) as predictors, and we test to see whether lagged  $Y_p$  has an effect. If  $Y_p$  has an effect, bilinearity is indicated.

To measure bilinearity, we create two dummy variables,  $P_{2,t-1,t}$  and  $N_{2,t-1,t}$  where  $P$  equals 1 if the partner's behavior was positive and 0 otherwise, and  $N$  equals 1 if the partner's prior behavior was negative. The dummy variable  $Q$  is 1 if the behavior is positive and -1 if negative. The equations that we gave earlier for the cross-lagged regression model would be modified as follows:

$$Y_{1,t} = c_{11} + a_{11}Y_{1,t-1,t} + p_{211}Y_{2,t-1,t}P_{2,t-1,t} + p_{n21}Y_{2,t-1,t}N_{2,t-1,t} + e_{11t}$$

$$Y_{2,t} = c_{21} + a_{21}Y_{2,t-1,t} + p_{121}Y_{1,t-1,t}P_{1,t-1,t} + p_{n12}Y_{1,t-1,t}N_{1,t-1,t} + e_{21t}$$

The null hypothesis is for the bilinear hypothesis that  $p_{211} = p_{n21}$  and  $b_{11} = 0$ . In other words, the null model states that the effect of the partner's positive behavior is the same (has the same slope) as the effect of the partner's negative behavior.

Although this dynamic-systems model appears to be very different from the cross-lagged regression model, we can view the approach within the APIM. Basically, the dynamic-systems approach postulates that the partner effect is nonlinear and that it varies by dyad. We believe that the dynamic-systems model might be better estimated by MLM. Moreover, we think it problematic that effects are not centered and that the residuals are analyzed, as opposed to simultaneously estimating actor and partner effects.

Again we reiterate that our discussion of dynamic modeling is very brief. To apply this technique, the reader needs to consult the sources that we cited or attempt our APIM adaptation.

## SUMMARY AND CONCLUSIONS

We have discussed five different models for the over-time analysis of dyadic relationships with interval variables: cross-lagged regressions, standard APIM, growth-curve analysis, cross-spectral analysis, and dynamic-systems analysis. Many of the models extend the APIM to over-time data. The cross-lagged regression examines the effect that the person's past behavior and the partner's past behavior have on the person's current behavior. The model of nonlinear dynamic systems presumes that partner

effects are nonlinear. Both cross-lagged regressions and dynamic growth models are stochastic in the sense that a person adjusts to the random changes of the partner.

The growth-curve model and the cross-spectral analysis model make different assumptions about the functional relationship between the dyad members' behavior and time. In growth-curve modeling, the functional form is usually assumed to be linear or exponential. In cross-spectral analysis, the assumption is that the functional form is cyclical (i.e., sinusoidal). Both growth-curve modeling and spectral analysis presume a deterministic growth pattern. Individuals are on track and continue along that track over time. A central question in all analyses is whether both members of the dyad have the same functional relationship; for example, the same rate of growth or the same cycle in their behaviors. If dyad members vary in their functional relationship, then we want to know whether there is any covariation in those functions.

We have seen that both MLM and SEM are used to estimate these models. There are analysis complications of centering, negative variances, and covariation between terms. However, resolving these difficulties is well worth the effort, as we learn a great deal about the structure of change in dyadic processes.

The methods that we have discussed in this chapter are complex, perhaps the most complicated in the book. However, even more complicated methods exist that we have not discussed. For instance, it is possible to combine the cross-lagged regression and growth-curve models (Curran & Bollen, 2001). Additionally, Boker and Laurenceau (2006) present an elaborate approach that blends nonlinear dynamical systems and MLM, and Ferrer and Nesselroade (2003) discuss dynamic factor analysis.

Additionally, one point that we discussed in Chapter 7 is that the standard APIM does not allow for measurement error in the predictors. We do note that the Ferrer and Nesselroade (2003) approach can be viewed as a latent variable, over-time APIM, thus controlling for measurement errors in the predictors.

In the next chapter, we consider over-time data with categorical outcomes: for example, studies of competition versus cooperation. We discuss a method of estimating the cross-lagged regression model with dichotomous outcomes, a procedure more generally known as sequential analysis. We consider estimating the model by both log-linear analysis and MLM. We also consider event-history analysis that treats time as a continuous variable.

## NOTES

1. We wish to acknowledge the extensive feedback and advice that we received from Niall Bolger on this and the next section of the chapter.
2. The standard functional form for negative exponential is  $ae^{-bt}$ , where  $e$  is the transcendental number that approximately equals 2.718. The specification that we have introduced based on regression toward the mean is that  $cd^t$ . The negative exponential formulation is identical to the regression toward the mean formulation given that  $c = a$  and  $d = -\ln(b)$ .
3. The material in this section was adapted from Cook (2003).

## 14

## Over-Time Analyses

### Dichotomous Outcomes

---

In the previous chapter, we described analytic techniques that are appropriate for data from two-person interactions that are measured over time on an interval-level variable. In this chapter, we consider the analysis of categorical outcomes, focusing much of our discussion on outcomes that are dichotomies. Most of this chapter is devoted to sequential analysis, or the study of behavior measured at multiple points in time. Our discussion highlights two different statistical estimation techniques: log-linear analysis and multilevel modeling (MLM).

---

As discussed in Cook (2003), two different approaches for sampling observations from the stream of behavior produced by dyadic interaction have been widely applied. In one approach, the behavior of both persons is coded or rated at specific intervals of time, such as every 15 seconds. In the other approach, the observations are coded at the onset of each new behavior or event. These two approaches are referred to as *interval sampling* and *event sampling*, respectively (Bakeman & Gottman, 1997). Interval sampling requires cutting the stream of behavior into units of equal duration, called *epochs*. An observer codes the behavior of interest for each epoch. Interval sampling of two-person interactions produces two parallel streams of behavior, as illustrated in Figure 14.1.

Rather than defining an observation unit as a specific time interval, event sampling defines an observation unit as a specific event. Conversational data are a good example, because in most cases only one person