17

Dyadic Data Analysis Using Multilevel Modeling

David A. Kenny
Department of Psychology
University of Connecticut

Deborah A. Kashy
Michigan State University

Multilevel modeling (MLM) is a method for analyzing hierarchically nested data structures such as over-time data from individuals (i.e., observations are nested within individuals) or group data (i.e., individuals are nested within groups). One way to conceptualize multilevel analyses is as a two-step process: A "lower-level" regression is computed separately for each "upper-level" unit (e.g., the relationship between a person's score on X and the person's score on Y is computed across individuals for each group), and then the lower-level regression estimates are pooled across groups. In the prototypical multilevel case, both the intercepts and slopes from these regressions are treated as random effects. Therefore, in addition to estimating the average intercept and average Y-X slope, both of which are fixed effects, the variance of the intercepts and variance of the slopes are also estimated.

In this chapter we consider multilevel analyses for dyadic data. Dyads are a special case of groups in that the number of individuals nested within each group equals two. From a multilevel perspective, this small group size represents a potential difficulty because, with only two data points, the two-step analysis conceptualization does not work. That is, if there are only two data points, it is not possible to compute a lower-level regression for each dyad because any two data points fall exactly on a line. Nonetheless, as we discuss in this chapter, MLM represents a powerful tool for the analysis of dyadic data, both for simple dyadic designs in which each person is a member of only one dyad as well as for complex designs in which individuals may participate in more than one dyad.

Dyadic data are very common in the social and behavioral sciences. The prototypical dyad study involves heterosexual married couples. However, there are many other possibilities: roommates, dating couples, friends,

coworkers, patient and caretaker, siblings, and parent and child. Dyads need not be preexisting but can be created in the laboratory. Members of the dyad need not even interact, as in the case of yoked controls. Moreover, the two observations may not even be from two people; they might be from two animals, two eyes, two arms, or the left and right side of the brain.

Recently Kenny, Kashy, and Cook (2006) have extensively discussed the analysis of dyadic data, and MLM is an important tool in many of the analyses they describe. In this chapter, we review and extend their discussion. We begin with key definitions in dyadic analysis concerning types of dyads, types of dyadic designs, and types of variables. We then consider the use of MLM for the three major types of dyadic designs. Finally, we consider the analysis of over-time data for one of those designs. Most dyadic researchers would benefit from reading the entire chapter even if they have an interest in only one of these designs. Topics that are discussed in one section (e.g., coding of variables, computer syntax, and interpretation of results) are relevant for the other sections of the chapter.

Because we cover many topics, we are limited in the amount of space that can be dedicated to illustrations; however, we have included example data sets, syntax files, and outputs on the website for the book. Here we focus primarily on the syntax of computer applications, but space limitations preclude discussion of all programs, and so we present syntax for use with SAS software (SAS Institute Inc., 2002–2003). We chose SAS because of its flexibility and its ability to link to data transformations and other procedures. However, at times we discuss the use of SPSS (SPSS for Windows Rel. 16.0) and other programs. To help interpret the

syntax, we adopt the convention that syntax commands are denoted by upper-case terms whereas variable names are denoted by lower-case terms in bold. We presume that the reader already has some familiarity with the concepts of MLM.

17.1 DEFINITIONS

17.1.1 Distinguishability

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. In heterosexual dating relationships, dyad members may be distinguishable by gender: Each couple has one man and one woman. In sibling dyads, the two siblings may 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 roommates or 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 or exchangeable, then there is no systematic or meaningful way to order the two scores.

The issue of distinguishability is both conceptual and empirical. The examples of gender in heterosexual couples and birth order in siblings highlight the conceptual component of distinguishability: There must be a categorical variable that can be used to systematically classify dyad members. However, even when dyad members are conceptually distinguishable, they may not

be empirically distinguishable. Empirical distinguishability occurs when there are detectable differences between dyad members as a function of the distinguishing variable. That is, if there are no differences in the means, variances, and covariances as a function of the distinguishing variable, then dyad members are not empirically distinguishable and a simpler and more parsimonious model (i.e., the model for indistinguishable dyads) can be estimated.

We shall see that whether dyad members are distinguishable has important implications for analyses, and we discuss procedures that can be used to test distinguishability. If there is no evidence of differences as a function of the distinguishing variable, we recommend that researchers use methods that are appropriate for indistinguishable dyads because such methods allow researchers to pool estimates both within and across dyads, which increases precision and statistical power. Even in cases in which there are compelling conceptual reasons for treating dyad members as distinguishable, one can statistically evaluate whether distinguishability makes a difference.

17.1.2 Typology of Dyadic Designs

In this chapter we discuss variants of three types of dyadic designs at length: the *standard dyadic design,* the *over-time standard dyadic design,* and the *one-with-many design.* We also provide a brief introduction to the Social Relations Model (SRM) design. The factor that differentiates these designs is the number of dyads in which each person participates. In addition, each of these designs can be reciprocal or nonreciprocal. A design is reciprocal when both dyad members provide outcome scores, and a design is nonreciprocal when only

TABLE 17.1

Possible Dyads for the Three Designs With Six People (1 Through 6)

Standard:	{1,2} {3,4} {5,6}
One-with-many:	{1,2} {1,3} {4,5} {4,6}
SRM:	{1,2} {1,3} {1,4} {1,5} {1,6} {2,3} {2,4} {2,5} {2,6} {3,4} {3,5} {3,6} {4,5} {4,6} {5,6}

one of the two persons is measured on the outcome.

In the standard design, each person is a member of one and only one dyad. As seen in Table 17.1, the six persons, 1 through 6, form 3 dyads: {1,2}, {3,4}, and {5,6}. The standard design is by far the most common dyadic research design, and it is typified by studies of marital relationships because (in most studies of marriage at least) each husband is paired with only one wife. For the standard design we consider only reciprocal designs in which both persons are measured on the outcome, because if only one person provides an outcome score the data would not be multilevel.

Likewise, in the over-time standard design, each person is a member of one and only one dyad. However, in this design, both partners are measured at multiple occasions. Most commonly, the two individuals are measured at the same points in time. For example, in a study of the transition to parenthood, both partners' relationship satisfaction might be measured one month prior to birth, and again 1, 3, and 6 months after birth. Thus, in this case individuals are nested within dyads, but time is crossed with individuals. Less commonly, the two partners are measured at different occasions, and so occasions are nested within the individual, and individuals are nested within dyad. We limit our discussion of the over-time standard design to the more common crossed structure.

In the one-with-many design each person is paired with multiple others, but these others are not paired with any other persons. As seen in Table 17.1, for the one-with-many design, the dyads are {1,2}, {1,3}, {4,5}, and {4,6}. Note that the "ones" are persons 1 and 4, and the "many" are 2, 3, 5, and 6. 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 primary care physician (so that there are multiple patients rating the same physician). In this design one person is linked to many others and the others are not linked to each other. In most cases, the one-with-many design is not reciprocal: Data are just from either the "one" or just from the "many."

In a Social Relations Model (SRM) design, each person is paired with multiple others and each of these others is also paired with multiple others. The prototypical SRM design is a round-robin design in which a group of persons rate or interact with each other. As seen in Table 17.1, all possible dyads are formed for the SRM design. For example, in a four-person round-robin design (e.g., persons A, B, C, and D), each person interacts with or rates three partners, and so a four-person round-robin results in a total of 6 dyads (i.e., AB, AC, AD, BC, BD, and CD). Because round-robin designs are inherently reciprocal, these 6 dyads generate a total of 12 outcome scores (i.e., the AB dyad generates two scores—A's score with B and B's score with A).

17.1.3 Types of Variables

In multilevel data, variables are traditionally denoted as varying at either the upper level (level 2) or the lower level (level 1), whereas in dyadic data, variables are typically denoted as between-dyads, within-dyads, or mixed. Not surprisingly, these two classification systems are related. Between-dyads variables vary from dyad to dyad, but do not vary within dyads, and so are upper level variables. 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 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 such that some dyads would be in the high stress condition and others would be in the low stress condition.

Alternatively, both within-dyads and mixed variables vary from person to person within the dyad, and so both of these types of variables can be viewed as lower-level variables from an MLM perspective. 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. A prototypical within-dyads variable is gender in heterosexual couples in that every couple is comprised of both a man and a woman. A less obvious example of a within-dyads variable is the 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 roommates.

Mixed variables vary both within and between dyads such that the two partners' scores can differ from each other and there

are differences in the dyad averages from dyad to dyad. Age is an example of a mixed independent variable in marital research because the two spouses' ages may differ from one another and some couples may be older on average than others. Because they can differ across the two dyad members, mixed variables fall into the lower level classification for MLM.

first semester (i.e., *Y*; in the table and syntax denoted as **adjust**), and the key predictor is a mixed variable that is an estimate of the average number of days per week each person drank during that semester (**act_drink**). Because the study includes only same-sex roommates, **gender** is a between-dyads variable that is represented as *Z* in this example. Note that the data, syntax, and outputs can be downloaded from http://davidakenny. net/dyadmlm/downloads.htm.

17.2 STANDARD DESIGN

In Table 17.2, we present a fictitious data set based on 10 pairs of same-sex roommates that we use as an example. In this data set the outcome variable is a measure of the person's college adjustment at the end of the

17.2.1 Indistinguishable Dyads

In the standard design, each person is paired with only one other person, and both dyad members are measured on the same variables. As a running example, we use the data presented in Table 17.2 in which

TABLE 17.2

Data set for the Fictitious Dyadic Study of Roommates

Dyad	Person	(Y) Adjust	(X) Act_Drink	Part_Drink	(Z) Gender	Citizen	D1	D2
1	1	5.5	0.2	−0.8	1	1	1	0
1	2	7.0	−0.8	0.2	1	−1	0	1
2	1	3.5	−0.8	0.2	−1	1	1	0
2	2	5.5	0.2	−0.8	−1	−1	0	1
3	1	2.5	2.2	3.2	1	1	1	0
3	2	0.3	3.2	2.2	1	−1	0	1
4	1	5.0	−1.8	−0.8	−1	1	1	0
4	2	5.5	−0.8	−1.8	−1	−1	0	1
5	1	3.0	0.2	1.2	1	1	1	0
5	2	2.0	1.2	0.2	1	−1	0	1
6	1	6.5	−0.8	−0.8	1	1	1	0
6	2	0.0	−0.8	−0.8	1	−1	0	1
7	1	5.5	−1.8	−0.8	1	1	1	0
7	2	6.0	−0.8	−1.8	1	−1	0	1
8	1	7.0	−0.8	0.2	−1	1	1	0
8	2	4.8	0.2	−0.8	−1	−1	0	1
9	1	4.5	1.2	0.2	−1	1	1	0
9	2	5.8	0.2	1.2	−1	−1	0	1
10	1	6.8	−0.8	1.2	−1	1	1	0
10	2	7.5	1.2	−0.8	1	−1	0	1

Note: act_drink and part_drink have been grand-mean centered around their mean of 1.80.

each person provides an outcome score, *Y*, a score on a lower-level predictor variable, *X* (note that *X* could be either mixed or within-dyads but the *X* depicted in Table 17.2 is mixed), and a score on an upper-level (i.e., between-dyads) predictor variable, *Z*.

If one takes the two-step analysis perspective for MLM, the level-1 model for person *i* in dyad *j* with a single lower-level predictor variable, *X*, would be

$$Y_{ij} = b_{0j} + b_{1j}X_{ij} + e_{ij},$$

where b_{0j} represents the predicted *Y* when *X* equals zero for person *i* in dyad *j*, and b_{1j} represents the coefficient that estimates the relationship between *X* and *Y* for dyad *j*. In the example, assuming that *X*, the person's drinking score (i.e., **act_drink**), has been grand-mean centered, b_{0j} represents the predicted college adjustment when drinking is average, and b_{1j} represents the change in adjustment as drinking increases by one day.[1] 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 level-2 analyses, the regression coefficients from the first step are assumed to be a function of a dyad-level predictor *Z*, and the equations would be

$$b_{0j} = a_0 + a_1Z_j + d_j$$

$$b_{1j} = c_0 + c_1Z_j.$$

The first level-2 equation treats the first-step intercepts as a function of the *Z* variable, and its form is similar to the standard MLM case that specifies that the first-step

intercepts are comprised of both fixed and random effects components. Specifically, a_0 estimates the grand mean (assuming that *X* and *Z* were either effect coded or grand-mean centered), a_1 estimates the overall effect of *Z* on *Y*, and d_j represents that part of the intercepts for dyad *j* that is not explained by *Z* (also called the residual). The variance of these residuals captures the nonindependence of the *Y* scores for the two dyad members, and the proportion of variance of these residuals, $s_d^2/(s_d^2 + s_e^2)$, measures the level of nonindependence and is commonly called the *intraclass correlation*.

The second level-2 equation treats the first-step slopes as a function of the *Z* variable. In this model c_0 estimates the average effect of *X* on *Y* and c_1 estimates the effect of *Z* on the *X-Y* relationship (i.e., the interaction between *X* and *Z*). This equation reflects the one major restriction that is necessary to apply MLM to dyads that need not be made for groups with more than two members: The first-step slopes are not allowed to vary randomly from dyad to dyad, and therefore are comprised of only fixed effects. This is because dyads do not have enough lower-level units (i.e., dyad members) to allow the slopes to vary randomly from dyad to dyad.

One additional modification of the standard MLM formulation is useful in dyadic data analysis. In the standard MLM formulation, nonindependence is modeled as a variance, but an alternative is to treat the scores from the two dyad members as repeated measures such that each dyad member would have an error and the errors would be correlated. Such a formulation models the nonindependence between dyad members as a covariance rather than a variance. This is particularly important when the outcome measure is structured such that when one dyad member has a higher score, the other

---

[1] Because the number of days drinking variable has a meaningful zero value, it is unnecessary to grand-mean center. Ordinarily, to have meaningful zero values, it is necessary to grand-mean center.

person's score tends to be lower (e.g., variables involving compensation, competition, division of a resource). Nonindependence in dyadic data is often negative and negative nonindependence can be captured by a covariance, but not by a variance. Thus, the standard formulation can be problematic and we strongly urge researchers to use the repeated measures formulation rather than the random intercept approach when analyzing dyadic data. Note that if the nonindependence is positive, then this covariance in the residuals equals the variance of the intercepts (s_d^2) described earlier.

The SAS syntax for a MLM that specifies this basic dyadic model in which the outcome variable is adjustment, the lower-level predictor is the person's drinking, the upper-level predictor is gender, and negative nonindependence is possible, would be:

> PROC MIXED COVTEST;
> CLASS **dyad**;
> MODEL **adjust = act_drink gender act_drink*gender**/S DDFM = SATTERTH;
> REPEATED/TYPE = CS SUBJECT = **dyad.**

The COVTEST option in the PROC MIXED statement requests that the random effects in the model be tested for statistical significance, thereby providing a test of the partial intraclass correlation for the outcome (partialling out the effects of the person's *X*). The CLASS statement defines **dyad** as a classification variable. The MODEL statement specifies that **adjust** is a function of the person's *X*, **act_drink**, the person's value on *Z*, **gender**, and the *XZ* interaction; the S (or SOLUTION) option requests that SAS print out the estimated fixed-effect coefficients, and the DDFM = SATTERTH option requests that the Satterwaithe approximation be used to compute the degrees of freedom.

Finally, the REPEATED statement is used to model the residual variance and covariance, and the SUBJECT = **dyad** option specifies that individuals at the same level of **dyad** are related. The TYPE = CS option requests a residual structure known as compound symmetry, which constrains the residual variances to be equal across the dyad members and specifies that there is a covariance between the residuals as well. The equal variance constraint is particularly important because the dyad members are indistinguishable, and so their residuals are sampled from the same underlying population.

The comparable SPSS syntax is:

> MIXED
> **adjust** WITH **act_drink gender**
> /FIXED = **act_drink gender act_drink*gender**
> /PRINT = SOLUTION TESTCOV
> /REPEATED = **person** | SUBJECT(**dyad**) COVTYPE(CS).

The **person** variable in the last syntax statement arbitrarily denotes the two dyad members as a "1" and "2."

Using these models with the example data set (i.e., *Y* = **adjust**, *X* = grand-mean centered **act_drink**, and *Z* = **gender**), $a_0 = 5.24$ and is an estimate the grand mean for adjustment, and $a_1 = -0.38$ and estimates the effect of gender on adjustment. Given that gender is coded men = 1 and women = –1, this value suggests that women's average college adjustment scores are higher than men's. The average effect that a person's drinking has on his or her college adjustment is $c_0 = -0.51$, indicating that each one unit increase in average weekly drinking corresponds to a predicted decrease of 0.51 points on adjustment. The gender difference in the relationship between drinking and adjustment is $c_1 = -0.63$. These

coefficients together suggest that the strong negative relationship between drinking and college adjustment is primarily true of men, because the coefficients predicting adjustment from drinking would be –1.14 for men and 0.12 for women. Finally, the residual variance is $s_e^2 = 1.653$, and the residual covariance (which in this case would equal the variance of the intercepts) is estimated at 0.855. So the partial intraclass correlation that estimates the similarity in the two roommates' adjustment scores after controlling for the effects of drinking and gender is $r_I = .52$, indicating that adjustment scores for roommates are quite similar.

In sum, the MLM model for the standard dyadic design with indistinguishable dyads, one X variable, and one Z variable has four fixed effects and two random effects (variation in the intercepts and error variance). Although it may seem that constraining the slope variance to zero might bias the other multilevel estimates, this is not the case. Instead, variance in the slopes is modeled as one component of the error variance. Tests of the null hypotheses are not biased when the slopes do in fact vary across dyads.

One common practice in MLM is to compute the mean of the level-1 predictor variable and use it as a level-2 predictor. This can only be done if the level-1 variable is a mixed variable, because within-dyads variables do not vary at level 2. For example, in addition to using a person's drinking as an X variable to predict his or her adjustment, the average drinking score for the dyad can be used as a Z variable in the level-2 equations. This would result in an estimate of (a) whether a person who drinks more is lower in adjustment, and (b) whether living in a room in which both roommates drink more on average moderates the effect of a person's drinking on his or her adjustment.

For dyadic data we recommend a somewhat different approach. Instead of using the dyad average of the X variable as a level-2 predictor variable, we suggest including both the person's X and his or her roommate's X as level-1 predictors of the person's adjustment. In the drinking example, this would allow us to estimate both the effect of a person's own drinking on his or her college adjustment as well as the effect of the roommate's drinking on the person's adjustment. This approach of using one's own and partner's X as predictors has been called the Actor-Partner Interdependence Model (APIM; Kenny et al., 2006). Note that in Table 17.2 each person's drinking score variable appears twice in the data file, once in the person's own record as an actor effect (i.e., **act_drink,** denoted in the following MLM equations as XA) and again in the partner's record as a partner effect (i.e., **part_drink** denoted in the following MLM equations as XP). Data files of this format are sometimes called *pairwise* data sets. The lower-level model that depicts both actor and partner effects is:

$$Y_{ij} = b_{0j} + b_{1j}XA_{ij} + b_{2j}XP_{ij} + e_{ij}.$$

The upper-level models, assuming that there is one upper-level predictor variable, Z, would be:

$$b_{0j} = a_0 + a_1Z_j + d_j$$

$$b_{1j} = c_0 + c_1Z_j$$

$$b_{2j} = h_0 + h_1Z_j.$$

The new parameters in these upper-level models, h_0 and h_1, can be interpreted as the average effect that the partner's X has on the person's Y, and the degree to which Z moderates the relationship between the partner's

X and the person's Y, respectively. Thus, the partner effect, h_0, would estimate the effect of the roommate's drinking on the person's college adjustment, and h_1 would estimate gender differences in the partner effect (e.g., having a roommate who drinks heavily may have more impact on men's adjustment than on women's). The SAS syntax for a basic actor-partner model that does not include a dyad-level (i.e., level 2) predictor would be:

```
PROC MIXED COVTEST;
CLASS dyad;
MODEL adjust = act_drink part_drink
   /S DDFM = SATTERTH;
REPEATED/TYPE = CS SUBJECT = dyad;
```

and the corresponding SPSS syntax would be:

```
MIXED
adjust WITH act_drink part_drink
/FIXED = act_drink part_drink
/PRINT = SOLUTION TESTCOV
/REPEATED = person | SUBJECT(dyad)
   COVTYPE(CSR).
```

To extend these models so that they include a Z variable (e.g., **gender**), the MODEL statement in SAS would be extended as follows:

```
MODEL adjust = act_drink part_drink
   gender gender*act_drink gender*
   part_drink/ S DDFM = SATTERTH;
```

and a similar change would be used for SPSS. The coefficients from these APIM models that include gender interactions using the example data set are $a_0 = 5.27$, $a_1 = -0.39$, $c_0 = -0.42$, $c_1 = -0.60$, $h_0 = -0.05$, $h_1 = -0.18$, and $r_I = .53$. The effects for actor are quite similar to those already described. The new partner effect results suggest that having a roommate who drinks more often has a negative effect on men's adjustment (combining h_0 and h_1 for men, the partner effect coefficient is -0.23), but not women's (the coefficient is 0.13).

One additional feature of the actor–partner model is that actor and partner effects can interact to create a new level-2 variable. We can form the interaction in the usual way by computing a product; in the roommate example this interaction might suggest that when both roommates drink a great deal, their adjustment scores are especially low. Alternatively, it may be more appropriate to form the interaction by computing the absolute difference between the person's X and the partner's X scores to create a measure of dissimilarity. Such an interaction might indicate that dissimilarity (i.e., when one person drinks a great deal but the other person does not) has a particularly detrimental effect on the two roommates' adjustment scores. We refer the reader to Chapter 7 of Kenny et al. (2006) for a discussion of these interactions.

17.2.2 Distinguishable Dyads

The standard MLM approach works well when dyad members are indistinguishable, and the model can be adapted to handle cases in which dyad members are distinguishable as well. We outline three different strategies for handling distinguishable dyads. As a running illustration, we amend our example by noting that for each roommate pair, one person is an international student and the other is a U.S. citizen. Thus, **citizen** (or C) is a within-dyads variable that can be used to systematically distinguish between the two roommates. In the data set in Table 17.2, the **citizen** variable is coded such that U.S. students are coded as 1 and non-U.S. students are coded as -1.

The first strategy is identical to the one presented above for handling indistinguishable dyads, but a coded variable (using either 1 and 0 dummy coding or 1 and –1 effect coding as is the case for the citizenship variable) is added to the model to code for the distinguishing variable. For the example, we would add the C variable into the level-1 equation with actor and partner effects (we initially do not include gender to simplify the model):

$$Y_{ij} = b_{0j} + b_{1j}XA_{ij} + b_{2j}XP_{ij} + b_{3j}C_{ij} + e_{ij}.$$

Additionally, we would likely want to allow for interactions between the other variables in the model and the distinguishing variable. For the example, such interactions would specify that the actor and partner effects may differ as a function of citizenship, and are included by multiplying the distinguishing variable times each level-1 variable in the model:

$$Y_{ij} = b_{0j} + b_{1j}XA_{ij} + b_{2j}X.P_{ij} + b_{3j}C_{ij}$$
$$+ b_{4j}XA_{ij}C_{ij} + b_{5j}XP_{ij}C_{ij} + e_{ij}.$$

The level-2 equations that include a dyad-level predictor, Z, would then be:

$$b_{0j} = a_0 + a_1Z_{1j} + d_j$$
$$b_{1j} = c_0 + c_1Z_{1j}$$
$$b_{2j} = h_0 + h_1Z_{1j}$$
$$b_{3j} = k_0 + k_1Z_{1j}$$
$$b_{4j} = m_0 + m_1Z_{1j}$$
$$b_{5j} = p_0 + p_1Z_{1j}.$$

In these models, k_0 estimates the average citizenship difference on Y or college adjustment, k_1 estimates the degree to which Z (gender of the two roommates) moderates the citizenship difference on Y, m_0 estimates the degree to which actor effects differ as a function of the person's citizenship, m_1 estimates whether the actor by citizenship interaction varies as a function of Z, and likewise, p_0 estimates the degree to which partner effects differ by citizenship and p_1 estimates the degree to which the partner by citizenship interaction varies as a function of Z.

Our discussion of distinguishable dyads thus far presumes that the residual variances are the same for both types of members (this is akin to the homogeneity of variance assumption in ANOVA). However, because the dyads are distinguishable, we would probably want to allow for heterogeneity of variance across levels of the distinguishing variable. In the example, this would allow the residual variances in college adjustment to differ for United States versus international students (perhaps there would be more unexplained variance in adjustment for international students relative to U.S. students).

The SAS syntax below specifies a model that includes actor and partner effects for the mixed predictor (i.e., **act_drink** and **part_drink**), a distinguishing variable, **citizen**, and a between-dyads variable, **gender**. (Note that the distinguishing variable can equivalently be treated as a class variable or simply as a dichotomous predictor—the only difference is in the appearance of the output.) By including interactions between the distinguishing variable and the actor and partner effects, the model allows the actor and partner effects to differ as a function of the distinguishing variable. Moreover, by changing the TYPE to CSH (heterogeneous compound symmetry) this syntax also allows for heterogeneous variances as a function of the distinguishing

variable. Note finally that the model also includes interactions with the upper-level or Z variable (i.e., dyad gender), and so there may be three-way interactions between gender, citizenship, and either actor or partner effects.

```
PROC MIXED COVTEST;
CLASS dyad;
 MODEL adjust = citizen gender act_
    drink part_drink citizen*gender
    citizen*act_drink citizen*part_drink
    gender*act_drink gender* part_
    drink citizen*gender*act_drink citi-
    zen*gender*part_drink / S DDFM =
    SATTERTH;
 REPEATED/TYPE = CSH SUBJECT =
    dyad.
```

The corresponding SPSS syntax is:

```
MIXED
adjust WITH citizen gender act_drink
    part_drink
/FIXED = citizen gender act_drink part_
    drink citizen*gender citizen*act_ drink
    citizen*part_drink gender*act_drink
    gender*part_drink citizen*gender*act_
    drink citizen*gender*part_drink
/PRINT = SOLUTION TESTCOV
/REPEATED = person|SUBJECT(dyad)
    COVTYPE(CSH).
```

A second strategy for the analysis of distinguishable dyads is the two-intercept model, which was originally suggested by Raudenbush, Brennan, and Barnett (1995). We presume again here that each dyad contains one U.S. citizen and one international student. Consider the empty model for member i of dyad j:

$$Y_{ij} = b_{1j}D_{1ij} + b_{2j}D_{2ij},$$

where D_{1ij} is 1 for the U.S. citizen and 0 for the international student, whereas D_{2ij} is 0 for a U.S. citizen and 1 for an international student. (The correlation between D_1 and D_2 is –1.) In this model there is no intercept, at least not in the usual sense, but rather there are two intercepts, b_1 and b_2. The intercept for U.S. students is estimated as b_1 and the intercept for international students is estimated as b_2. In addition, there is no error term in the model, making this a very unusual model. Importantly, in this model both b_1 and b_2 are random effects, and so the model has a variance-covariance matrix of b_1 and b_2 with three elements: the variance of b_1 or s_1^2 (the error variance for U.S. students), the variance of b_2 or s_2^2 (the error variance for international students), and the covariance between the two or s_{12} (the degree of nonindependence).

If there were any X or Z variables, they are added to the model, but any X variables need to be multiplied by each of the two D dummies. We add here an actor and a partner effect for X:

$$Y_{ij} = b_{1j}D_{1ij} + b_{2j}D_{2ij} + b_{3j}D_{1ij}XA_{ij}$$
$$+ b_{4j}D_{2ij}XA_{ij} + b_{5j}D_{1ij}XP_{ij}$$
$$+ b_{6j}D_{2ij}XP_{ij}.$$

Thus, the actor effects for U.S. students are given by the b_{3j}'s, and the actor effects for international students are given by the b_{4j}'s, and likewise for the partner effects: b_{5j} and b_{6j}.

With SAS there are two ways to estimate the two-intercept model. The first method is relatively simple. In this method, the distinguishing variable is treated as a classification variable. It is then entered into the MODEL statement, and the NOINT

option is used to suppress the intercept. By suppressing the intercept and including the distinguishing variable as a categorical or classification variable, we force the program to compute two intercepts, one for **citizen** = 1 and the other for **citizen** = –1. Similarly, we obtain estimates of separate actor and partner effects by including interactions between the distinguishing variable and the actor and partner variables. The syntax below estimates the two intercept model that also includes separate actor and partner effects for United States versus international students. (Note that for simplicity, we have not included **gender**.)

```
PROC MIXED COVTEST;
CLASS dyad citizen;
MODEL   adjust = citizen  citizen*act_
   drink citizen*part_drink/S DDFM =
   SATTERTH NOINT;
REPEATED / TYPE = CSH SUBJECT =
   dyad.
```

The comparable syntax is SPSS is:

```
MIXED
adjust BY citizen WITH act_drink
   part_drink
/FIXED = citizen        citizen*act_drink
   citizen*part_drink | NOINT
/PRINT = SOLUTION TESTCOV
/REPEATED = person | SUBJECT(dyad)
   COVTYPE(CSH).
```

The more direct, but less simple way to estimate the two-intercept model with SAS is to actually create the two dummy variables and then use them in the syntax. For example, we could define **d1** = 1 if **citizen** = 1 and **d1** = 0 otherwise and **d2** = 1 if

citizen = –1 and **d2** = 0 otherwise. The SAS syntax would then be:

```
PROC MIXED COVTEST;
CLASS dyad;
MODEL  adjust = d1  d2  d1*act_drink
   d2*act_drink   d1*part_drink   d2*
   part_drink
/S DDFM = SATTERTH NOINT;
RANDOM  d1  d2 /  SUBJECT = dyad
   TYPE = UN;
PARMS 2, 1, 2, 0.000001 / HOLD = 4.
```

The last statement sets the starting values for the random effects, and it is required because in the two-intercept model, the error variance must be constrained to zero. The specific values in the PARMS statement refer to the variables in the RANDOM statement, and because **d1** and **d2** have a UN or unstructured variance-covariance matrix, there are four random effects being estimated: UN(1,1), which is the variance of the U.S. students' intercepts; UN(2,1), which is the covariance between the United States and international students' intercepts; UN(2,2), which is the variance of the international students' intercepts; and a residual variance. The first three numbers in the PARM statement can be almost any value, as long as the two variances are positive and the covariance is less than the absolute value of the product of the two standard deviations. The last value, 0.000001, specifies that the residual or error variance has a starting value that is virtually, but not exactly, zero (i.e., it is 0.000001), and the HOLD = 4 instructs the program to fix the residual variance, the fourth parameter, to its starting value. So far as we know, there is no way within SPSS to fix the error variance to zero. MLwiN does allow zero error

variance whereas HLM allows the fixing of the error variance to a very small value.

Finally, it is sometimes useful to use two of the methods we have described to estimate distinguishable models. In particular, the first method we described provides estimates and tests of the interactions between *X* variables and the distinguishing factor (e.g., the test of whether the effect of one's own drinking differs for U.S. students relative to international students). If such interactions emerge, then estimating and testing the simple slopes separately for each level of the distinguishing variable (e.g., what is the effect of one's own drinking for U.S. students, and what is the effect of one's own drinking for international students) is a natural way to break down the interaction. These estimates and tests of the simple slopes can be provided directly by either of the two-intercept models we have described.

17.2.3 Test of Distinguishability

MLM can be used to test whether conceptually distinguishable dyads are actually empirically distinguishable. Kenny et al. (2006) present such a test using structural equation modeling. To conduct this test using MLM, two models must be estimated, and both of these models should use maximum likelihood estimation (ML) rather than the typical program default of restricted maximum likelihood (REML). The ML option should be used because the distinguishable model generally differs from the indistinguishable model in its fixed effects. In SAS this is accomplished by adding METHOD = ML to the PROC MIXED statement.

In the first model, dyad members are treated as distinguishable both in terms of their fixed and random effects.

PROC MIXED COVTEST METHOD = ML;
CLASS **dyad**;
MODEL**adjust = citizenact_drinkpart_ drinkcitizen*act_drinkcitizen*part_ drink** / S DDFM = SATTERTH;
REPEATED/TYPE = CSH SUBJECT = **dyad;**

In the second model, dyad members are treated as indistinguishable.

PROC MIXED COVTEST METHOD = ML;
CLASS **dyad**;
MODEL **adjust** = **act_drink part_drink** / S DDFM = SATTERTH;
REPEATED/TYPE = CS   SUBJECT = **dyad;**

A chi-square difference test can then be computed by subtracting the deviances (i.e., the -2*log likelihood values). For example, if we want to compare the indistinguishable actor-partner model (dropping the between-dyads, or Z, variable for simplicity) with the actor-partner model that treats the dyad members as distinguishable, there are three additional fixed effects (**citizen, citizen*act_ drink, citizen*part_drink**) and one additional random effect because heterogeneous compound symmetry (CSH rather than CS) allows the two variances to differ but homogeneous compound symmetry does not. If the χ^2 difference with 4 degrees of freedom were not statistically significant, the data would be consistent with the null hypothesis that the dyad members are indistinguishable. If, however, χ^2 were statistically significant, then there would be support for the alternative hypothesis that dyad members are distinguishable. In the example,

AU:Please add a reference with full publication information for Kenny et al. (2006) or delete this citation.

the distinguishable model has a deviance of 60.920 and the indistinguishable model has a deviance of 64.507. Thus, the test of distinguishability is $\chi^2(4) = 3.587$, $p = .46$, and so in the example data set, there is not much evidence that the roommates are empirically distinguished by citizenship (the sample size is very small, resulting in low power for this test).

17.3 OVER-TIME STANDARD DESIGN

Here we consider the standard design with the complication that each member of the dyad is measured at multiple times. By multiple, we mean more than twice and preferably each dyad member is measured more than five times. In over-time data from dyads, there are three factors that define the structure of the data: 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. As pointed out by Laurenceau and Bolger (2005), the problem is that time and person are usually crossed, not nested. That is, for a given dyad, the time point is the same for the two persons at each time point.

There are two potentially undesirable consequences if the three-level nested model is mistakenly assumed. First the correlation between the two partner's intercepts, r_{cc}, is constrained to be positive because it is estimated as a variance. As we have described, there are several types of outcome measures for which this dyadic correlation would likely be negative. The second consequence of mistakenly conceptualizing

crossed over-time dyadic data as a three-level nested structure is that the correlation between the two members' errors at each time, r_{ee}, is assumed to be zero. This correlation measures the time-specific similarity (or dissimilarity) in that part of Y that is not explained by the predictors for the two partners. Such time-specific similarity might arise because of arguments or other events that occur immediately prior to an assessment occasion.

An additional issue with over-time data is the necessity of modeling the nonindependence that arises because variables are measured over time, which is commonly called *autocorrelation* (see e.g., Hillmer, 2001). 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.

We can divide over-time models into two major types. First, there are stochastic models in which a person's or dyad's score is a function of past scores plus a random component. Second, there are deterministic models in which the person or dyad is assumed to be on some sort of trajectory. In this chapter, we focus on linear deterministic growth models. In these models, the explanatory variable is time of measurement, and each person has a slope that estimates his or her rate of change, as well as an intercept that measures the person's level at time zero. As discussed in many treatments of growth models (e.g., Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004), choice of time zero is an important one; however, in this chapter we simply assume that time zero is the initial observation. We begin by discussing distinguishable dyads. We then

turn our attention to the more complex case of indistinguishable dyads.

17.3.1 Distinguishable Dyads

Consider as a simple example, an over-time study of marital satisfaction (**satisf**) in which satisfaction is measured yearly for 5 years. The data for two couples from this fictitious over-time study are presented in Table 17.3, and the distinguishing variable, **gender**, is coded 1 for husbands and –1 for wives. As can be seen in the table, there are 10 records for each dyad (five for each person) and so the data set is structured in a time-as-unit format, sometimes called a *person-period* data set. Thus if 60 married couples were measured at five times, there would be 600 records.

In this data set **time** is coded as zero at the initial assessment and then increases by one each year. On each data record, we recommend creating one additional variable, what we call **timeid,** which simply equals **time**. We do so because we need two different time variables: one of which is continuous time (**time**) and the other is treated as categorical (**timeid**). One key idea in the analysis of data from the over-time standard design is to use the distinguishing variable (e.g., **gender**) to create two dummy variables that represent the two "classes" of individuals. One dummy variable, what we denote as *H* in the MLM equations (**husband** in Table 17.3), is set to one when the scores are from the husband and to zero otherwise. The other, what we call *W* (**wife** in Table 17.3) is set to one when the scores are from the wife and zero otherwise.

TABLE 17.3

Data from Two Couples in a Fictitious Over-Time Study of Marital Satisfaction

Dyad	Time	Person	Husband	Wife	Satisf	Gender
1	0	1	1	0	5	1
1	1	1	1	0	6	1
1	2	1	1	0	8	1
1	3	1	1	0	7	1
1	4	1	1	0	4	1
1	0	2	0	1	4	–1
1	1	2	0	1	5	–1
1	2	2	0	1	3	–1
1	3	2	0	1	6	–1
1	4	2	0	1	7	–1
2	0	1	1	0	8	1
2	1	1	1	0	6	1
2	2	1	1	0	4	1
2	3	1	1	0	7	1
2	4	1	1	0	6	1
2	0	2	0	1	6	–1
2	1	2	0	1	7	–1
2	2	2	0	1	8	–1
2	3	2	0	1	3	–1
2	4	2	0	1	6	–1

Note:

The basic idea is that we create a two-level model in which level 1 is time or observation for both persons, and level 2 is the dyad. Through its use of the *H* and *W* dummy codes, this single model actually represents two growth curves, one for each member of the dyad. Having the two persons represented by one model allows us to model the nonindependence of the intercepts as a covariance and estimate a time-specific correlation between the residuals as well. The level-1 equation for person *i* in dyad *j* at time *t* would be

$$Y_{ijt} = b_{01j}H + b_{02j}W + b_{11j}HT_t$$
$$+ b_{12j}WT_t + He_{1jt} + We_{2jt}.$$

Note that the only predictor variable is time or T_t. There are four random variables at the level of the dyad: the two intercepts, b_{01j} and b_{02j}, and the two slopes, b_{11j} and b_{12j}. For the example, b_{01j} estimates the husband's satisfaction at the beginning of the study (**time** = 0) and b_{02j} estimates the wife's initial satisfaction. The slopes estimate the degree to which the husband's and wife's satisfaction increases or decreases each year on average.

The variance-covariance matrix for these intercepts and slopes contains four variances and six covariances. Two key covariance parameters for the dyadic growth model are the covariances between the two members' intercepts and two slopes. For the example, the covariance between the intercepts estimates whether husbands and wives are similar in their level of satisfaction at the initial assessment. The covariance between the slopes measures whether the rate of change in a husband's satisfaction is similar to his wife's. The four remaining covariances concern correspondence between the intercepts and slopes, including two within-person covariances (e.g., when wives start the study with lower satisfaction do they change more slowly?) as well as two between-person covariances (e.g., when wives start the study with lower satisfaction do their husbands change more slowly?)

There are also two residual variances, one for husbands and one for wives, and these two residuals may have a covariance. In the example, the correlation between the residuals, what we denoted as r_{ee} earlier, measures the degree to which the husband's and wife's satisfaction scores are especially similar at a particular time point, after taking the intercepts and slopes into account.

Dyadic growth models for distinguishable dyads can be estimated using either SAS or SPSS.

For distinguishable dyads, the SAS code is:

```
PROC MIXED COVTEST;
    CLASS dyad timeid gender;
MODEL      satisf = husband      wife
    husband*time wife*time / NOINT S
    DDFM = SATTERTH ;
    RANDOM      husband      wife
    husband*timewife*time/SUB = dyad
    TYPE = UN;
    REPEATED  gender /  TYPE = CSH
    SUBJECT = timeid(dyad);
```

Because there is no intercept in the model (NOINT), the program estimates separate intercepts and separate slopes for husbands and wives. Using TYPE = UN in the RANDOM statement specifies that there are no equality constraints on the variance or covariance estimates, and so separate values are estimated across the distinguishing variable (e.g., across husbands and wives). The TYPE = CSH option in the repeated statement allows for different

residual variances across the distinguishing variable, and it also specifies that there may be a time-specific correlation between the residuals. It is important to realize that although this specification of the error structure allows for different variances across the distinguishing variable, it does fix the error variances to be the same value at each time. Likewise, it estimates a single time-specific covariance rather than different values for each time point.

The SPSS syntax is:

```
MIXED
satisf BY dyad timeid gender WITH
    husband wife time
/FIXED = husband wife husband*time
    wife*time | NOINT
/PRINT = SOLUTION TESTCOV
/RANDOM husband wife husband*time
    w i f e * t i m e | S U B J E C T ( d y a d )
    COVTYPE(UNR)
/REPEATED gender | SUBJECT(timeid*
    dyad) COVTYPE(CSH).
```

We did not achieve a solution with the SPSS but did with SAS.

We refer to this model as the fully *saturated model* in the sense that we have treated all of the fixed effects (i.e., the two intercepts and two slopes) as random. In some cases, there can be difficulties in the estimation of multilevel models when either one or more of the variances is small or two or more of the terms are highly collinear. The researcher may need to explore simpler models, for example, a model in which intercepts are random and slopes are fixed or a model in which both dyad members share a common slope or intercept. The fixed and random effects can be constrained to the same value across the distinguishing variable by replacing the MODEL and RANDOM statements with

```
MODEL satisf = time/S DDFM =
    SATTERTH ;
RANDOM INTERCEPT time / SUB =
    dyad TYPE = UN;
```

This basic linear growth model is typically only the starting point of an over-time dyadic analysis. Researchers might want to see whether the levels or rates of change differ as a function of person-level predictors such as personality scores. For instance, we could include the actor's and partner's neuroticism in the analysis, and in our marital satisfaction example we would be able to determine whether the change in wife's satisfaction over time differs depending on the wife's own neuroticism and her husband's neuroticism. Dyad-level predictors such as length of relationship or experimental condition can also be treated as moderators of the effect of time. Moreover, time-varying variables such as employment status at each data collection period can also be included as predictors. Finally, growth models need not be linear, and so nonlinear functions of time such as quadratic time variables can be included in the analysis.

17.3.2 Indistinguishable Dyads

Although the fundamental principle with indistinguishable dyads is that they cannot be systematically distinguished, the analysis of data from an over-time standard design for indistinguishable dyads requires that we create a variable (which we call **person)** that distinguishes between the dyad members. Thus, for each dyad, one member is coded **person** as 1 and the other is coded **person** as 2. Using **person**, we create two dummy variables. One dummy, what we call P1, is set to one when the scores are from person 1 and to zero otherwise; and the other, what

we call P2, is set to one when the scores are from person 2 and zero otherwise.

The level-1 model is very similar to that for distinguishable dyads,

$$Y_{ijt} = b_{01j}P1 + b_{02j}P2 + b_{11j}P1T_t$$
$$+ b_{12j}P2T_t + P1e_{1jt} + P2e_{2jt}.$$

However, when dyad members are indistinguishable, a series of equality constraints need to be included for both the fixed and random effects. For example, the two intercepts, b_{01j} and b_{02j}, should be equal, as should the two slopes, b_{11j} and b_{12j}. Indeed, the analysis becomes more complicated because we need to place similar constraints on the 4-by-4 variance-covariance matrix of slopes and intercepts.

This analysis can be conducted using SAS and MLwiN but not the current versions of SPSS and HLM. To accomplish this analysis, a data set (referred to as the **g** data set) that defines the equality constraints on the variance-covariance matrix must be created. The values in this data set refer specifically to the RANDOM statement in the PROC MIXED procedure. As shown below, the RANDOM statement in SAS syntax for this analysis defines four random effects: The two dummy variables, **p1** and **p2**, represent the two persons' intercepts, and **p1*time** and **p2*time** represent the two persons' slopes. These random effects create a 4-by-4 variance-covariance matrix where the diagonal represents the variances of the random effects.

Because the dyad members are indistinguishable, a certain structure needs to be imposed on this matrix. Specifically, the intercept variances (**p1** and **p2**) need to be fixed to the same value, and the slope variances (**p1*time** and **p2*time**) need to be

fixed to the same value. In addition, the within-person intercept-slope covariances (**p1** with **p1*time** and **p2** with **p2*time**) need to be equated, as do the between-person intercept-slope covariances (**p2** with **p1*time** and **p1** with **p2*time**). There are two other elements in this matrix: the cross-person intercept covariance (**p1** with **p2**) and the cross-person slope covariance (**p1*time** with **p2*time**). Thus, although there are potentially 10 elements in the variance-covariance matrix, there are only six unique parameter estimates due to the equality constraints.

The **g** data set has a very specific format and must include the following variables: PARM, ROW, COL, and VALUE. These variables are linked to the ordering of variables in the RANDOM statement in the PROC MIXED syntax. PARM represents parameter number. Thus, because we are specifying six variance-covariance parameters between the intercepts and slopes (the intercept variance, the slope variance, the intercept covariance, the slope covariance, the within-person intercept-slope covariance, and the between-person intercept-slope covariance), PARM takes on six different values. The nature of the variance-covariance structure for the residuals is specified in the REPEATED statement, and does not play a role in **g**.

The ROW and COL values refer to the variables in the random statement, and so a 1 for ROW represents the effect of the **p1** dummy code (the intercept for **p1**), and having ROW = COL = 1 implies that the estimated parameter is the variance of the intercepts for **p1** (i.e., it is the covariance of **p1**'s intercept with itself). Although the second line in this data set has a PARM value of 1, it also has ROW = COL = 2. This information together specifies that the variance of the intercepts is based on the variance of

both the **p1** intercept and the **p2** intercept. The next line identifies the second parameter that is the covariance between the two person's intercepts. The third and fourth lines define the variance of the slopes, and the fifth line defines the covariance between the two slopes. The remaining two parameters are the within person intercept-slope covariance (PARM = 5) and the between person intercept-slope covariance (PARM = 6).

The full SAS syntax for creating the G matrix would be

```
DATA g;
INPUT PARM ROW COL VALUE;
DATALINES;
1 1 1 1
1 2 2 1
2 3 3 1
2 4 4 1
3 1 2 1
4 3 4 1
5 1 3 1
5 2 4 1
6 1 4 1
6 2 3 1
;
```

The SAS syntax for the analysis is then:

```
PROC MIXED COVTEST;
CLASS dyad timeid person
MODEL y = time/NOINT S DDFM =
    SATTERTH ;
RANDOM p1 p2 p1*time p2*time /G
    SUB = dyad TYPE = LIN(6) LDATA = g;
REPEATED person/TYPE = CS SUB =
    timeid(dyad);
```

Note on the REPEATED statement that TYPE is CS rather than CSH. This specifies that the residual variances are the same both across time and across person.

To test for distinguishability we could estimate two models using ML. One allows for full distinguishability and the other allows for indistinguishability. The difference in deviances is a chi-square test that, for the model presented, would have 7 degrees of freedom: the four constraints made in the **g** matrix (two sets of variances equal and two sets of covariances equal), the two equality constraints of the two intercepts and slopes, and the equality of the two error variances.

17.4 ONE-WITH-MANY DESIGN

In the one-with-many design, a person is in multiple dyadic relationships, but each of the person's partners is in a relationship with only that one person. For instance, a doctor might interact with many patients. Alternatively, adolescents might provide information about their relationships with their mothers, fathers, romantic partners, and best friends. We refer to the person who has multiple partners (the "one") as the *focal person* and to the multiple others (the "many") as the *partners*. In the first example, doctors would be focal persons and the patients would be their partners; in the second example the adolescents would be the focal persons and their mothers, fathers, romantic partners, and best friends would be the partners. As is illustrated by these two examples, the partners in the one-with-many design can be distinguishable or indistinguishable.

The one-with-many design is a blend of the standard dyadic and Social Relations Model (SRM) designs; it is similar to the standard design in that each partner is paired with only one focal person, and it is like an SRM design in that the focal

person is paired with many partners. The one-with-many design is nonreciprocal if only one of the two members of each dyad provides outcome scores. Thus, if the adolescents rated their relationship closeness with their mothers, fathers, romantic partners, and friends, but these partners did not rate their closeness with the adolescent, the design would be nonreciprocal. Similarly, if the partners rated their closeness with the adolescent, but the adolescent did not make ratings, it would again be nonreciprocal. On the other hand, if both the focal person and the partners are measured, the design would be reciprocal. We first consider the nonreciprocal design for indistinguishable and distinguishable partners, and then we discuss the reciprocal design.

17.4.1 Nonreciprocal One-With-Many Designs: Indistinguishable Partners

As an example, we use the small data set presented in Table 17.4 from a fictitious study of intimacy in friendships. In this table we see that our data set contains nine focal persons (5 women and 4 men; **focalsex** is coded 1 = men and -1 = women), each of whom report on the intimacy of their friendships (1 = not at all intimate, 9 = very intimate) with varying numbers of male and female friends (**partsex** is also coded 1 = men and -1 = women). The data in Table 17.4 actually include two measures of intimacy: One variable denotes the intimacy scores as rated by the focal person (**f_intimacy**), and the second denotes the intimacy scores as rated by the partners (**p_intimacy**). Thus, the data set is actually reciprocal (although we do not treat it as such at this point in our discussion and so **p_intimacy** will be ignored). Say that one question to be addressed in the

study is whether the partners' relationship self-esteem (**rsepart**) predicts the focal person's perceptions of intimacy, and whether such a relationship is moderated by the focal person's gender.

Two variables must be included in the data set to apply MLM to the one-with-many design. First, a variable that identifies the focal person that is involved in each dyad should be created. In this chapter, we use **focalid** to denote the focal person. Second, a variable that identifies the partner that is involved in the dyad must be created, and we use the variable **partid** for this purpose. We assume that both **focalid** and **partid** start at the number one and continue to as many as needed. For **partid**, we have a choice. We can either number them consecutively from 1 to the total number of partners in the study, or for each focal person we use the same numbers. That is, if each focal person has three partners, they would be numbered 1, 2, and 3. When partners are distinguishable, the latter strategy is preferable. When partners are indistinguishable, either method can be used, but as we show, the SAS syntax would need to change.

Data from the one-with-many design are hierarchically structured because partners are tied to a focal person. In some ways, the one-with-many design is more closely linked to standard MLM designs because partners are nested within "groups" that are defined by the focal person. Thus, in the one-with-many design, the upper-level or level 2 is the focal person and the lower-level or level 1 is the partner. The variable X (e.g., partner's relationship self-esteem) is a level-1 variable because it is assumed to vary across partners within focal person. The level-1 equation for partner i with focal person j is:

$$Y_{ij} = b_{0j} + b_{1j}X_{ij} + e_{ij}.$$

TABLE 17.4

Data Set for the Fictitious Indistinguishable One-With-Many Design

Focalid	Partid	F_Intimacy	P_Intimacy	Focalsex	Partsex	Rsepart	Rsepartc
1	1	5	1	1	1	5	−5.06
1	2	8	6	1	1	4	−6.06
1	3	5	3	1	−1	5	−5.06
1	4	2	5	1	−1	13	2.94
2	1	7	6	−1	1	12	1.94
2	2	6	4	−1	−1	11	0.94
2	3	9	3	−1	−1	14	3.94
3	1	5	5	−1	1	8	−2.06
3	2	4	6	−1	1	12	1.94
3	3	2	7	−1	1	9	−1.06
3	4	6	6	−1	−1	15	4.94
3	5	5	7	−1	−1	17	6.94
4	1	7	3	1	−1	8	−2.06
4	2	5	2	1	−1	7	−3.06
4	3	8	4	1	−1	13	2.94
4	4	5	5	1	1	5	−5.06
5	1	1	6	1	1	12	1.94
5	2	2	6	1	1	16	5.94
5	3	5	7	1	−1	15	4.94
6	1	6	8	−1	−1	17	6.94
6	2	5	5	−1	−1	9	−1.06
6	3	8	5	−1	1	14	3.94
7	1	4	6	1	1	12	1.94
7	2	6	7	1	1	8	−2.06
7	3	7	4	1	1	4	−6.06
7	4	8	4	1	−1	3	−7.06
8	1	6	8	−1	−1	1	−9.06
8	2	7	7	−1	1	8	−2.06
8	3	9	8	−1	1	9	−1.06
9	1	7	7	−1	−1	15	4.94
9	2	5	4	−1	1	10	−0.06
9	3	7	4	−1	1	11	0.94

Note: focalsex and partsex is coded boys = 1, girls = −1.

To make the intercept, b_{0j}, more interpretable, it is generally advised to center the X variables by subtracting the overall partner mean (i.e., the grand mean computed across all partners in the data set). For the example data, this would mean that the grand mean relationship self-esteem, $M = 10.062$), would be subtracted from each partner's **rsepart** score. The term e_{ij} represents the error or residual for partner i with focal person j. The level-2 models are:

$$b_{0j} = a_0 + a_1 Z_j + d_j$$

$$b_{1j} = c_0 + c_1 Z_j + f_j,$$

where Z_j is a level-2 variable (e.g., focal person gender, or **focalsex**) such that it takes on the same value for all partners of the same

focal person. Unlike the standard design, there can be a random component for the slopes as well as for the intercepts. If there are relatively few partners per focal person, allowing such a variance may not be possible. If *k* refers to the number of partners per focal person (assumed to be equal only for the purposes of this calculation), to treat *all* the level-1 slopes as random for *p* lower-level predictor variables, *k* must be at least *p* + 2. In the small example data set, each person reports on at least three friendships, and so we could (in principle, although we do not do so) include a random slope for one lower-level predictor variable.

Although the slopes can be constrained to be equal for all focal persons (i.e., the slopes would be modeled as a fixed effect component only), it is almost always advisable to allow for the possibility that the intercepts vary across focal persons. As was true for the standard design, the variation of the intercepts models the nonindependence in the data—but in this case the nonindependence refers to similarity in scores for individuals who are paired with the same focal person. We can compute the variance of the intercepts or s_d^2 and the error variance or s_e^2. The ratio of the variance due to the intercept to the total variance or $s_d^2/(s_d^2 + s_e^2)$, provides an estimate of the intraclass correlation, the measure of nonindependence. When there are lower-level predictor variables, or *X*s, the intraclass correlation based on the intercepts is a partial intraclass that represents the proportion of variance due to the focal persons after controlling for the effects of the predictor variable(s).

The interpretation of this measure of nonindependence depends on whether the data come from the focal person or the partners. If the data come from the focal person (e.g., **f_intimacy**), then the variance in the

intercepts refers to the consistency in how the focal person sees or behaves with the partners. (It is analogous to the actor variance in an SRM design, see below.) In the example analysis of the focal person's ratings of intimacy, the variance in the intercepts measures the degree to which focal persons tend to report similar levels of intimacy across all of their friends. If the data come from the partners, then the variance in the intercepts refers to the consistency in how the partners see or behave with the focal person. (It is analogous to the partner variance in an SRM design, see below.) If we treated the partner-rated intimacy as the outcome measure, the variance of the intercepts would measure the degree to which friends experience similar levels of intimacy with the focal person.

For the example data, we estimated a model predicting the focal person's intimacy ratings that allowed for random intercepts, random slopes for the effect of the partner's relationship self-esteem, and a covariance between the intercepts and slopes. In this model we treated grand-mean centered relationship self-esteem (**rsepartC**) as a lower-level predictor and focal-person gender as an upper-level predictor. The SAS syntax for this analysis is

```
PROC MIXED COVTEST;
CLASS focalid;
MODEL  f_intimacy = rsepartC focal-
    sex  rsepartC*focalsex/S DDFM =
    SATTERTH;
RANDOM  INTERCEPT/SUBJECT =
    focalid;
```

and the corresponding SPSS syntax is

```
MIXED
f_intimacy WITH rsepartC focalsex
```

/FIXED = **rsepartC focalsex focalsex* rsepartC** | SSTYPE(3)
/PRINT = SOLUTION TESTCOV
/RANDOM INTERCEPT | SUBJECT (**focalid**) COVTYPE(VC).

Based on the data in Table 17.4, $a_0 = 5.45$, indicating that average intimacy scores were somewhat above the scale midpoint. The effect of focal-person gender was $a_1 = -0.64$, and given the coding scheme, this indicates that men reported lower average intimacy across friends than did women. The average effect of relationship self-esteem on intimacy was relatively small, $c_0 = -0.05$, but there was evidence of a focal-person gender difference for the effect of relationship self-esteem on intimacy, $c_1 = -0.21$. Thus, women reported higher intimacy with friends who had higher relationship self-esteem, and men reported lower intimacy with friends that had higher relationship self-esteem. The random effects yielded the intercept variance of $s_a^2 = 1.127$, and a residual or error variance of $s_e^2 = 2.500$. Thus, the partial intraclass for intimacy ratings, controlling for the effects of self-esteem is .311, and so there is some evidence that focal persons reported similar levels of intimacy across their friends.

It is important to note that by definition a ratio of the variance of the intercepts to the total variance must be nonnegative, and so this method presumes that the intraclass correlation cannot be negative. Normally, we would not expect the intraclass correlation to be negative for the one-with-many design. However negative intraclass correlations could occur for variables requiring social comparison or compensation across partners; for example, if the outcome variable is structured such that if one partner has a high score, other partners have lower scores, the intraclass might be negative. However, treating nonindependence as a variance precludes the possibility of any negative nonindependence. If a negative intraclass correlation (Kenny, Mannetti, Pierro, Livi, & Kashy, 2002) is likely to occur (e.g., variables for which one partner having a high score constrains other partners to having lower scores), we suggest that partners should be treated as a repeated measure so that nonindependence is modeled as a correlation rather than a variance. This would be accomplished in SAS by substituting the RANDOM statement with a REPEATED statement as follows:

REPEATED/SUBJECT = **focalid**
TYPE = CS;

and similarly for SPSS:

/REPEATED **partid** | SUBJECT(**focalid**) COVTYPE(CS).

It is possible to treat the mean of the X variable for each focal person, or M_X, as a predictor in the second-stage of the multilevel analysis. For instance, in the example where X is relationship self-esteem, then the effect of M_X on a focal person's intimacy would estimate whether individuals whose friends have higher self-esteem on average report higher levels of intimacy on average. For the standard design, we suggested using the partner's X as a level-1 predictor (e.g., the partner's drinking), but there is not a direct extension of this approach to the one-with-many design. Nonetheless, as we have shown with the example of focal-person gender, characteristics of the focal person may be relevant predictors.

Perhaps the most interesting case occurs when the same predictor is measured for

the focal person as well as the partners. In the example this would mean that we have a measure of the focal-person's relationship self-esteem (e.g., **rsefocal**) in addition to the partners' scores on this variable. In this case the lower-level predictor, or *X*, would be the partner's relationship self-esteem, and the upper-level predictor, *Z*, would be the focal-person's relationship self-esteem. In such a case, the effect of focal-person self-esteem on focal-person intimacy would be an actor effect (i.e., If I have higher self-esteem, do I tend to see my relationships as having higher intimacy on average?), and the effect of partner self-esteem on focal-person intimacy would be a partner effect (i.e., If my partner has higher self-esteem, do I have higher intimacy scores with that partner?). In a parallel fashion, if the outcome measure is the partner-rated intimacy, then the effect of partner self-esteem on partner-reported intimacy would be an actor effect, and the effect of focal-person self-esteem on partner-reported intimacy would be a partner effect.

17.4.2 Nonreciprocal One-With-Many Designs: Distinguishable Partners

One major difference between the distinguishable and indistinguishable cases is that when partners are distinguishable, both the fixed and the random effects may vary by the distinguishing variable. Differential fixed effects are modeled by including partner role as a predictor in the model. In cases in which the effects of other *X* or *Z* variables are examined, interactions between these variables and partner role should be included as well.

Differential random effects as a function of the distinguishing variable can take two forms. The most general format does not place any constraints on the variance-covariance matrix of the random effects. In this model (which we will term the unconstrained random effects model) separate variances are computed for each role and separate covariances are estimated for each combination of roles. Because the covariances can differ across partners, this specification suggests that the focal-person effect may vary across partner roles. The alternative specification, which we refer to as the *constrained random effects model*, estimates a random focal-person effect in the form of an intercept variance, and then allows for differential residual variances for the different partner roles. In effect, this model constrains all of the covariances across partner roles to the same value—which is the variance of the intercepts.

As an example, consider the data in Table 17.5 in which the focal person is an adolescent child and the partners are the child's mother (**partrole** = 1), father (**partrole** = 2), and home-room teacher (**partrole** = 3). The key outcome variable is a measure of the child's cooperativeness, and each partner reports on this measure (**cooperate**; 1 = not at all cooperative, 9 = very cooperative). The data set also includes the child's gender (**focalsex;** boys = 1, girls = -1), which would be an upper-level predictor. Although for simplicity our example does not include any *X* variables, these could be easily added into the model as main effects and in interactions with partner role or other predictors. Finally, to make the fixed effects results more readily interpretable, we create two dummy coded variables: **teacher** = 1 if **partrole** = 3, and **teacher** = 0 otherwise; **father** = 1 if **partrole** = 2, and **father** = 0 otherwise. This coding scheme makes the mothers' ratings serve as the comparison group.

TABLE 17.5

Data Set From the Fictitious Distinguishable One-With-Many Study

Focalid	Focalsex	Partrole	Cooperate	Teacher	Father
1	1	1	6	−1	−1
1	1	2	5	0	1
1	1	3	4	1	0
2	1	1	4	−1	−1
2	1	2	3	0	1
2	1	3	5	1	0
3	−1	1	6	−1	−1
3	−1	2	6	0	1
3	−1	3	3	1	0
4	−1	1	7	−1	−1
4	−1	2	6	0	1
4	−1	3	6	1	0
5	−1	1	8	−1	−1
5	−1	2	7	0	1
5	−1	3	6	1	0
6	1	1	5	−1	−1
6	1	2	3	0	1
6	1	3	3	1	0
7	1	1	6	−1	−1
7	1	2	4	0	1
7	1	3	5	1	0
8	−1	1	6	−1	−1
8	−1	2	5	0	1
8	−1	3	5	1	0
9	1	1	9	−1	−1
9	1	2	7	0	1
9	1	3	6	1	0
10	−1	1	5	−1	−1
10	−1	2	6	0	1
10	−1	3	4	1	0

Note: focalsex is coded boys = 1, girls = −1; partrole is coded 1 = mothers, 2 = fathers, 3 = teachers.

The following SAS syntax includes the child's gender (**focalsex**) as an upper-level predictor, and partner role (**partrole**) is included as a categorical lower-level predictor. This approach provides overall *F*-tests that test whether there are mean differences as a function of partner role, and whether gender interacts with partner role. In addition to allowing for differential fixed effects, this syntax specifies the unconstrained random effects model in which all of the random effects may differ as a function of partner role:

PROC MIXED COVTEST;
CLASS **focalid partrole**;
MODEL **cooperate** = **partrole focalsex partrole*focalsex**/S DDFM = SATTERTH;
REPEATED   **partrole**/TYPE = UN SUBJECT = **focalid** RCORR;

The REPEATED statement estimates the variance-covariance matrix for the different partner roles, and because TYPE is set to UN (unspecified), there are no equality constraints on this matrix. Thus, the variances for each partner role can differ, as can the covariances between the different pairs of roles. Adding RCORR to the REPEATED line gives the correlation matrix across partner roles. The syntax for SPSS is:

```
MIXED
cooperate BY partrole WITH focalsex
/FIXED = partrole focalsex partrole*
    focalsex
 /PRINT = SOLUTION TESTCOV
/REPEATED = partrole|SUBJECT
    (focalid) COVTYPE(UNR).
```

The analysis using the small data set in Table 17.5 shows evidence of partner role differences in both the fixed and random effects. For example, the F-test of the partner role main effect is $F(2,8) = 13.46$, $p < .01$, and the F-test for the interaction between partner role and child gender is $F(2,8) = 3.62$. The effect estimates from these sets of syntax can be difficult to interpret because both SAS and SPSS create their own dummy codes for variables that are treated as categorical (or as factors in SPSS terms). As a result, it can be useful to estimate a second model that uses our dummy coded variables for teachers and fathers. For SAS this would be:

```
PROC MIXED COVTEST;
CLASS focalid partrole;
MODEL cooperate = father teacher focal-
    sex father*focalsex teacher*focalsex/S
    DDFM = SATTERTH;
REPEATED partrole/TYPE = UN
    SUBJECT = focalid RCORR;
```

This analysis indicates that both fathers and teachers rated the children's cooperativeness lower than did mothers (i.e., the coefficient for fathers was $b = -1.0$, $t(8) = 4.26$, $p < .01$ and for teachers was $b = -1.5$, $t(8) = 3.81$, $p < .01$). In addition, it appears that the interaction between partner role and child gender is largely due to fathers ($b = -0.6$, $t(8) = 2.56$, $p = .03$). Thus, given the coding for **focalsex**, this suggests that the fathers tended to see boys as more cooperative than they saw girls.

The random effects from this unconstrained model suggest that the variance for mothers was somewhat larger than those for fathers or teachers (mothers' $s^2 = 2.4$, fathers' $s^2 = 1.65$, teachers' $s^2 = 1.5$). The correlations suggest that the two parents' perceptions were more similar to one another ($r = .88$) than they were to the teacher's ratings (for mothers $r = .62$, and for fathers $r = .46$).

The SAS syntax for the constrained random effects model that allows for a constant level-2 (i.e., focal person) effect, while specifying heterogeneous error variances for the different partner roles would be:

```
PROC MIXED COVTEST;
CLASS focalid partrole;
MODEL cooperate = partrole focal-
    sex partrole*focalsex/S DDFM =
    SATTERTH;
RANDOM INTERCEPT/subject = focalid;
REPEATED partrole / TYPE = VC
    SUBJECT = focalid GRP = partrole;
```

This model in essence allows for heterogeneous error variances, but assumes that the covariances between partners are homogeneous, which is captured by the variance of the intercepts.

Finally, as was the case in the standard dyadic design, it is possible to test whether it

is statistically useful to distinguish the partners. As before we would run two models, both of which will need to use ML rather than the program default of REML. The first model would be the indistinguishable model in which only *X* and *Z* variables predict the outcome, and the error structure is treated as compound symmetry. The second model would be the distinguishable model, and in addition to the effects of *X* and *Z*, this model would include the main effect of partner role as well as interactions between partner role and the other predictors. In addition, this second model specifies a heterogeneous variance-covariance matrix (either using the unconstrained random effects model or the constrained random effects model as described above). For the example data, we estimated the distinguishable model with unconstrained random effects and found a deviance of 76.1 based on a model estimating 12 parameters (6 random effects and 6 fixed effects). We next estimated the model for indistinguishable partners and found a deviance of 101.6 based on a model estimating 4 parameters (2 random effects and 2 fixed effect). The $\chi^2(8) = 25.5$, $p = .001$, suggesting that the model treating dyad members as distinguishable provides a better fit to the data.

17.4.3 Reciprocal One-With-Many Designs: Indistinguishable Partners

In a reciprocal one-with-many design outcome scores are obtained from both the focal person and the partners. For instance, we might ask doctors and their patients if they are each satisfied with one another. A unique and useful aspect of reciprocal one-with-many designs is that they allow estimation of both generalized and dyadic reciprocity. For the doctor-patient satisfaction study generalized reciprocity measures whether doctors who are on average more satisfied with their patients tend to have patients who are on average more satisfied with them. Dyadic reciprocity measures whether a patient with whom the doctor is especially satisfied is also especially satisfied with the doctor.

Table 17.6 shows the data layout required for the analysis of a reciprocal one-with-many design in which the partners are indistinguishable. This table is based on the data presented in Table 17.4, which was a fictitious study of friendship intimacy. To conduct a reciprocal one-with-many analysis using MLM, there would be two records for each focal person-partner dyad, one that contains the focal person's dyad-specific rating (e.g., the score that was presented as **f_intimacy** in Table 17.3, which is the focal person's rating of intimacy with a partner), and one that contains the partner's dyad-specific rating (e.g., **p_intimacy**, which is the partner's rating of intimacy with the focal person). Thus, there is now only one **intimacy** variable in Table 17.6. Three additional variables, which we call **rater**, **focal**, and **partner**, respectively must be created. These variables specify who generated the data—the focal person or the partner. The **rater** would equal –1 if the data are from the focal person, and it would equal 1 if the data are from the partner; the **focal** variable would be 1 if the outcome was generated by the focal person, and 0 if it was generated by the partner; and the **partner** variable would be 0 if the outcome was generated by the focal person, and 1 if it was generated by the partner.

The MLM equations for a reciprocal design are based on the two-intercept approach in which two dummy variables are created to denote the person that provided the outcome score. Thus, we would

TABLE 17.6

Data Set for the Fictitious Reciprocal Indistinguishable One-With-Many Design

Focalid	Partid	Rater	Focal	Partner	Intimacy	Rsepart	Focalsex	Partsex	A_sex	P_sex
1	1	1	1	0	5	5	1	1	1	1
1	1	−1	0	1	1	5	1	1	1	1
1	2	1	1	0	8	4	1	1	1	1
1	2	−1	0	1	6	4	1	1	1	1
1	3	1	1	0	5	5	1	−1	1	−1
1	3	−1	0	1	3	5	1	−1	−1	1
1	4	1	1	0	2	13	1	−1	1	−1
1	4	−1	0	1	5	13	1	−1	−1	1
2	1	1	1	0	7	12	−1	1	−1	1
2	1	−1	0	1	6	12	−1	1	1	−1
2	2	1	1	0	6	11	−1	−1	−1	−1
2	2	−1	0	1	4	11	−1	−1	−1	−1
2	3	1	1	0	9	14	−1	−1	−1	−1
2	3	−1	0	1	3	14	−1	−1	−1	−1
3	1	1	1	0	5	8	−1	1	−1	1
3	1	−1	0	1	5	8	−1	1	1	−1
3	2	1	1	0	4	12	−1	1	−1	1
3	2	−1	0	1	6	12	−1	1	1	−1
3	3	1	1	0	2	9	−1	1	−1	1
3	3	−1	0	1	7	9	−1	1	1	−1
3	4	1	1	0	6	15	−1	−1	−1	−1
3	4	−1	0	1	6	15	−1	−1	−1	−1
3	5	1	1	0	5	17	−1	−1	−1	−1
3	5	−1	0	1	7	17	−1	−1	−1	−1

Note: Rater = 1 if the score was provided by the focal person and rater = −1 if the score was provided by the partner.

have R_1, which is coded 1 if the data are provided by the focal person and 0 if the data came from the partner, and R_2, which is coded 0 if the data are provided by the focal person and 1 if the data are provided by the partner. Using these two variables allows us to specify a model with separate effects for the focal persons and the partners and separate residuals for focal persons' ratings of partners and partners' ratings of the focal persons. The lower-level MLM equation is:

$$Y_{ijk} = b_{01j}R_1 + b_{02j}R_2 + R_1e_{ij1} + R_2e_{ij2}.$$

Where i refers to the partner, j refers to the focal person, and k denotes whether the data was provided by the focal person (i.e., $k = 1$) or the partner (i.e., $k = 2$).

The basic level-2 models are:

$$b_{01j} = a_{01} + d_{1j}$$
$$b_{02j} = a_{02} + d_{2j}.$$

The two random variables at the level of the focal person are d_{1j} and d_{2j}, and they each have a variance. The variance in d_{1j} measures the degree to which the focal person rates all partners in a similar way (i.e., the actor effect from the SRM, see below), and the variance in d_{2j} measures the degree to which partners' ratings of the focal person are similar (i.e., the partner effect from the

SRM, see below). These two random effects also have a covariance that measures generalized reciprocity, or whether there is correspondence between how the focal person generally sees his or her partners and how the partners generally see the focal person. In the intimacy example, this covariance would measure whether individuals who report higher intimacy scores across their friends, are rated as higher in intimacy by those friends. The two errors from the lower-level model also have variances and a covariance between them. Although the error variances may not be simply interpreted, the covariance between the two can address an interesting question concerning dyadic reciprocity: If the focal person reports especially high intimacy with a particular friend, does that friend also report especially high intimacy?

Beyond the variance decomposition aspect, the reciprocal design can also incorporate predictor variables for either the focal person, the partners, or both. For example, if a partner-level variable, *X*, such as the relationship self-esteem measure, **rsepart** were included, the model would be expanded to:

$$Y_{ijk} = b_{01j}R_1 + b_{02j}R_2 + b_{11j}XR_1$$
$$+ b_{12j}XR_2 + R_1 e_{ij1} + R_2 e_{ij2}.$$

This model specifies that intimacy ratings are a function of who makes the rating (e.g., the focal person or the friend; b_{01j} and b_{02j}), and it also specifies that the partner's level of relationship self-esteem may moderate both persons' intimacy ratings (b_{11} and b_{12}).

When the partners are indistinguishable, the SAS syntax to estimate the variance partitioning and covariances (i.e., the model with no partner or focal person predictors) is

```
PROC MIXED COVTEST;
CLASS focalid partid rater;
MODEL   intimacy = focal   partner/
   NOINT S DDFM = SATTERTH;
RANDOM focal partner / SUBJECT =
   focalid TYPE = UN;
REPEATED rater/SUBJECT = partid
   (focalid) TYPE = UN;
```

Here the coding of **partid** is important. If each partner had a unique identification number, we would just need **partid** for the SUBJECT in the REPEATED statement, but **partid(focalid)** also works. If **partid** is not unique, as is the case in Table 17.6, then **partid(focalid)** must be used. Note that the traditional intercept is suppressed (i.e., NOINT) and so there are two intercepts, one for data from the focal person and the other for data from the partners.

The RANDOM statement results in estimates of the variance in the two intercepts, as well as their covariance. For the **focal** variable this variance estimates the degree to which focal persons differ in their average partner ratings (it is akin to the actor variance in the SRM, see below). The variance in the **partner** intercepts estimates the degree to which there are focal person differences in the average ratings they are given by their partners (this is akin to the partner variance in the SRM, see below). The generalized reciprocity covariance is estimated as the covariance between these two level-2 effects.

The REPEATED statement is necessary to specify the error variances and covariances. Again there are two error variances, one for the partners' ratings of the focal person and another for the focal person's ratings of the partners. This statement specifies that rater is repeated across partners, and because the covariance matrix is unspecified (TYPE = UN), it also estimates a

covariance that can be viewed as the dyadic or "error" reciprocity covariance.

The SPSS Syntax to estimate the variance partitioning and covariances is

> MIXED
> **intimacy** BY **rater focalid partid** WITH **focal partner**
> /FIXED = **focal partner** | NOINT
> /PRINT = SOLUTION TESTCOV
> /RANDOM **focal partner** | SUBJECT (**focalid**) COVTYPE(UN)
> /REPEATED = **rater**|SUBJECT(**focalid* partid**) COVTYPE(UN).

When there are predictor variables measured for both the focal person and the partners, researchers have two options. In the intimacy example, we have both focal-person and partner gender (see **focalsex** and **partsex** in Table 17.4). The first option would be to keep these two gender variables as they are currently coded, and we would allow each of these to predict data from the focal person and the partners. The model statement in SAS for this analysis would be

> MODEL **intimacy** = **focal partner focal* focalsex focal*partsex partner* focalsex partner*partsex** / NOINT S DDFM = SATTERTH;

The second option would be to again code two gender variables, but in this case we would code gender of the actor (i.e., the person doing the rating; **a_sex** in Table 17.6) and gender of the partner (i.e., the person being rated; **p_sex**). Thus, when the data point is focal-person rated intimacy, **a_sex** refers to the focal person's gender and **p_sex** refers to the partner's gender. When the data point is the partner's intimacy, **a_sex** refers to the partner and **p_sex** refers to the focal

person. The model statement in SAS for this analysis would be

> MODEL **intimacy** = **focal partner focal* a_sex focal*p_sex partner*a_sex partner*p_sex**/NOINT S DDFM = SATTERTH;

While mathematically equivalent, it is advisable to try out both ways of coding to determine the coding method that yields the simpler and more interpretable results. In either case, the researcher should allow the two gender variables to interact with the **focal** and **partner** dummy coded variables. The meanings of those interactions are very different for the two coding systems.

17.4.4 Reciprocal One-With-Many Designs: Distinguishable Partners

Recall that in the one-with-many design with distinguishable partners, the focal-person is paired with a set of partners who fall into different roles. As an example, we might have adolescents as the focal person, and the partners might be their mother, father, romantic partner, and best friend. In the reciprocal design, the adolescents would rate their relationship closeness with each of these partners, and we would also have the mother, father, romantic partner, and best friend rate their relationship closeness with the adolescent, and so the outcome variable would be **close**. In this example the data set would include a partner role variable, **partrole**, which differentiates these four types of partners (e.g., for mothers **partrole** = 1, for fathers **partrole** = 2, for romantic partners **partrole** = 3, and for best friends **partrole** = 4).

When partners are distinguishable, there are two ways to model the distinguishability. We might just treat the data as if partners

were totally distinguishable. In this analysis, the number of "variables" would be the number of partners times two because the design is reciprocal. In this case we would allow for intercept differences for the different partners (both from the focal person's perspective and from the partner's perspective), different variances, and different covariances between each pair of variables. In essence, this is the saturated model for treating the design as reciprocal. The SAS code for a model of complete distinguishability is as follows:

```
PROC MIXED COVTEST;
CLASS focalid partrole rater;
MODEL    close    =    focal*partrole
    partner*partrole/NOINT S DDFM =
    SATTERTH;
REPEATED partrole*rater /SUBJECT =
    focalid TYPE = UN;
```

The SPSS syntax is:

```
MIXED
close BY rater focalid partrole WITH
    focal partner
/FIXED = focal*partrole partner* par-
    trole | NOINT
/PRINT = SOLUTION TESTCOV
/REPEATED = partrole*rater|
    SUBJECT(focalid) COVTYPE(UN).
```

The other alternative is to estimate separate fixed effects for the different roles, while including a general intercept random effect for the focal person (rather than allowing separate random focal-person intercepts for each role) and likewise including a general intercept random effect for partners. That is, rather than allowing for different intercept random effects for the different types of partners, there is only one focal person intercept and only one partner intercept.

Thus the focal person variance measures consistency in the adolescent's ratings of his or her partners and the partner variance measures similarity in the partner's ratings of the adolescent. In such a model, distinguishability would be specified by allowing the error variances and covariances to differ by partner role. The SAS syntax would be

```
PROC MIXED COVTEST;
CLASS focalid partrole rater;
MODEL    close    =    focal*partrole
    partner*partrole/NOINT S DDFM =
    SATTERTH;
RANDOM  focal  partner/SUBJECT  =
    focalid TYPE = UN;
REPEATED    rater/SUBJECT    =
    partrole(focalid) TYPE = CSH GRP
    = partrole;
```

These MLM have many parameters and they may be slow to converge.

Once again we could test for distinguishability. As before we estimate two models using ML estimation, one of which treats the partners as distinguishable and the other treats the partners as indistinguishable. We would then subtract the deviances of the two models, and that difference has a chi-square distribution given the null hypothesis that dyad members are indistinguishable. The degrees of freedom for the test would be the extra number of parameters in the model when dyad members are treated as distinguishable.

17.5 SOCIAL RELATIONS MODEL DESIGNS

In SRM designs, each person is paired with more than one partner and each partner

is also paired with multiple others. A full discussion of how to analyze these complex designs using multilevel modeling is beyond the scope of this chapter. For a more complete discussion of the models, we refer interested readers to Kenny (1994) or Kenny et al. (2006), and for SAS syntax to use MLM to analyze SRM data, see Kenny and Livi (2009). Our discussion here is intended as a brief introduction to the designs.[2] In addition, like the other dyadic models we have discussed, the SRM can be used for both indistinguishable and distinguishable dyads. Here we limit the discussion to the indistinguishable case and readers should consult Kenny et al. (2006) or Kashy and Kenny (2000) for more detail on the distinguishable case.

The prototypical SRM design is a round-robin design where a group of persons rate or interact with all the other persons in the group (e.g., Persons A, B, C, and D interact and then A rates B, A rates C, and A rates D. Similarly, B rates A, C, and D for a total of 12 dyadic scores). By convention, the person who generates the measurement is called *actor* and the other person is called *partner*. For instance, if we ask people interacting in small groups how much they like one another, the person reporting on the liking is the actor and person being liked is the partner.

---

[2] In addition to the round-robin design, the SRM can also be estimated using a block design. In a block design, the people are divided into two subgroups and members rate or interact with members of the other subgroup but not with members of their own group. In the symmetric block design the members of the two subgroups are indistinguishable whereas in the asymmetric block design the members of the two subgroups are distinguishable (e.g., one subgroup is comprised of men and the other subgroup is comprised of women). Finally, in the half-block design, we have data from one subgroup with one other subgroup (e.g., men with women but not women with men). The half-block design, unlike the other SRM designs, is not reciprocal: Each person is either an actor or a partner but not both

17.5.1 The SRM Components

The basic SRM equation is:

$$Y_{ijk} = m_k + a_{ik} + b_{jk} + g_{ijk},$$

where Y_{ijk} is the score for person i rating (or behaving with) person j in group k. In this equation m_k is the group mean, a_{ik} is person i's actor effect, b_{jk} is person j's partner effect, and g_{ijk} is the relationship or actor–partner interaction effect. The terms m, a, b, and g, are random variables and each has a variance: σ_m^2, σ_a^2, σ_b^2, and σ_g^2. The SRM also specifies two different correlations between the SRM components of a variable, both of which can be viewed as reciprocity correlations. At the individual level, a person's actor effect can be correlated with that person's partner effect; this covariance assesses generalized reciprocity, and is denoted as σ_{ab}. If the variable being rated is liking, then this covariance measures whether a person who likes everyone in the group is liked by everyone in the group. At the dyadic level, the two members' relationship effects can be correlated; this covariance assesses dyadic reciprocity and is denoted as σ_{gg}. In the example, this covariance measures whether a person who especially likes a particular partner, is especially liked by that partner. There are then seven SRM parameters, one mean, four variances, and two covariances.

As an example, we might have a group of people who rate one another's intelligence. The meanings of the SRM parameters would be as follows:

μ: The overall mean of rated intelligence

σ_m^2: The variance in average ratings of intelligence across groups.

σ_a^2: The variance in how intelligent a person generally sees others.

σ_b^2: The variance in how intelligent a person is generally seen by others.

$\sigma_g{}^2$: Unique variance in the way a particular individual sees a particular partner (i.e., the relationship variance or actor–partner interaction variance plus error .

σ_{ab}: The covariance between how intelligent a person generally sees others with how intelligent other people generally see that person (i.e., generalized reciprocity).

σ_{gg}: The covariance between how one person uniquely sees another with how that other uniquely sees that person (i.e., dyadic reciprocity).

17.5.2 Analysis of Round-Robin Designs

Because of the complexity of the nonindependence, estimating a MLM from round-robin data is problematic. Sometimes researchers treat actor as level 2, and all of the observations within an actor are treated as level 1. Alternatively, researchers may treat partner as level 2, and all of the observations within a partner are treated as level 1. Ironically, both of these are right and wrong at the same time. Actor and partners are levels, but both need to be considered in one model. Moreover, actor and partner are crossed or cross-classified, not nested. Additionally, the analysis must take into account the correlation between the two scores from members of the same dyad.

To estimate the SRM with round-robin data, the structure of the data set is particularly important. For the analyses we present, the data set is structured such that each record is the response of one person in a dyad (e.g., Person A's rating of Person B's intelligence). As seen in Table 17.7, a data set for a round-robin that includes five individuals in

TABLE 17.7

Data From Group 1 of a Fictitious SRM Round-Robin Study

Group	Actor	Partner	Dyad	Y	Act_x	Part_x
1	1	2	1	6	5	7
1	1	3	2	4	5	6
1	1	4	3	2	5	5
1	1	5	4	7	5	7
1	2	1	1	7	7	5
1	2	3	5	5	7	6
1	2	4	6	6	7	5
1	2	5	7	7	7	7
1	3	1	2	3	6	5
1	3	2	5	4	6	7
1	3	4	8	6	6	5
1	3	5	9	5	6	7
1	4	1	3	2	5	5
1	4	2	6	4	5	7
1	4	3	8	3	5	6
1	4	5	10	6	5	7
1	5	1	4	7	7	5
1	5	2	7	6	7	7
1	5	3	9	7	7	6
1	5	4	10	6	7	5

a group would therefore be comprised of 20 records for each group. (Note that this value is 20, not 25 because self-rating data are treated differently by the SRM.) On each record there would be three identification variables: a group identification variable (e.g., **group**), an identification variable that designates the actor who made the rating (e.g., **actor**), and an identification variable that designates the target of the rating (e.g., **partner**). We shall see that sometimes other variables need to be included in the data set.

In this chapter we present a conventional MLM approach. A more complex approach involving dummy variables can be found in Snijders and Kenny (1999) and Kenny and Livi (2009). In this section because of space limitations, we do not explicitly consider fixed variables (what we denoted as *X* and *Z* in prior sections). Such variables would be included in the MODEL statement in SAS and the /FIXED statement in SPSS. Individual-level variables (e.g., the individual's extroversion) would need to be included on all records for which that individual is the rater. Moreover, the researcher should consider including both the actor's variables (e.g., the actor's extroversion) and the partner's variables (e.g., the partner's extroversion) as potential predictors.

17.5.2.1 Conventional MLM

Increasingly, multilevel programs can estimate models with cross-classified variables. However, in these models the actor-partner covariance is assumed to be zero, which is a major limitation of this method. This is particularly problematic when variables such as liking are under consideration because generalized reciprocity is a likely component for this variable (i.e., likeable people tend to like others). We describe this approach in three steps and detail how both SAS and SPSS can be used to estimate the model. We believe that HLM and MLwiN can also estimate the models in this fashion.

To use the conventional MLM approach, identification numbers need to be assigned to each group, each actor, each partner, and each dyad. For SPSS these values must be unique. For example, if there are 15 five-person groups, then the **group** variable would range from 1 to 15; the values for the **actor** variable would range from 1 to 5 for group 1 and from 6 to 10 for group 2; the values for the **partner** variable would be 1 to 5 in group 1 and 6 to 10 in group 2; and finally, because there are 10 dyads in a five-person round-robin, the **dyad** variable will range from 1 to 10 for group 1, 11 to 20 for group 2, and so on. Unique identification numbers are not required for SAS.

We first present the syntax for SAS and then for SPSS. Note again that the actor-partner covariance is not modeled. The syntax for SAS is

```
PROC MIXED COVTEST;
CLASS actor partner dyad group;
    MODEL y = /S DDFM = SATTERTH
        NOTEST;
    RANDOM INTERCEPT/TYPE = VC
        SUB = actor(group);
    RANDOM INTERCEPT/TYPE = VC
        SUB = partner(group);
    RANDOM INTERCEPT/TYPE = VC
        SUB = group;
    REPEATED/TYPE = CS SUB = dyad
        (group);
```

The syntax for SPSS is as follows:

```
MIXED
    y BY group
```

        /PRINT = SOLUTION TESTCOV
        /RANDOM   INTERCEPT|SUBJECT
            (**group**) COVTYPE(VC)
        /RANDOM   INTERCEPT|SUBJECT
            (**actor**) COVTYPE(VC)
        /RANDOM   INTERCEPT|SUBJECT
            (**partner**) COVTYPE(VC)
        /RANDOM   INTERCEPT |SUBJECT
            (**dyad**) COVTYPE(VC).

In SPSS, the REPEATED statement cannot be used for dyad, and so one must presume that the dyadic covariance is positive.[3] Note also that in SPSS the error variance equals the dyad variance plus the error variance, and the dyadic correlation equals the dyad variance divided by the sum of the dyad variance plus the error variance.

The results from SAS and the SPSS would be different if the reciprocity covariance were negative. In that case, it would be incorrectly estimated as zero by SPSS and properly estimated by SAS.

17.6 CONCLUSIONS

We have shown that MLM can be used to estimate a wide range of dyadic models. However, structural equation modeling can also be a useful tool for the analysis of dyadic data, especially when dyad members are distinguishable and when SRM and one-with-many designs are used.

We made a sharp distinction between distinguishable and indistinguishable dyads, but there are variants in-between. For instance, in the one-with-many design, partner can be partly distinguishable and partly indistinguishable. Consider the example in which some partners are men and others are women. We could distinguish between gender, but within gender the partners would be indistinguishable. We limited our discussion to normally distributed outcome variables and we did not consider models for counts and proportions. Although MLM can now be used to estimate models with such outcome scores, these approaches often do not currently allow for correlated errors. In other words, the approaches do not allow for REPEATED statements, which were an important component of nearly every dyadic design that we discussed.

In some sense dyadic models are simple MLMs. Certainly, the basic model for the standard design is a very simple MLM model. However, as we saw, when we allow for distinguishability and more complex designs that are reciprocal, the analysis can become quite complicated. MLM offers the possibility of being able to estimate a wide range of dyadic models and is becoming an important tool for dyadic researchers.

REFERENCES

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9,* 30–52.

---

[3] For SPSS 16 and earlier, tests of variances are two-tailed when they should be one-tailed. Thus, *p* values should be divided by two

Hillmer, S. (2001). Time series regressions. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications*. Mahwah NJ: Erlbaum.

Kashy, D. A. (1992). Levels of analysis of social interaction diaries: Separating the effects of person, partner, day, and interaction. *Dissertation Abstracts International, 53*(1-B), 608–609.

Kashy, D. A., & Kenny, D. A. (1990). Analysis of family research designs: A model of interdependence. *Communication Research, 17,* 462–483.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York, NY: Guilford.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis.* New York, NY: Guilford.

Kenny, D. A., & Livi, S. (2009). A componential analysis of leadership using the social relations model. In F. J. Yammarino & F. Dansereau (Eds.), *Multi-level issues in organizational behavior and leadership* (Vol. 8 of *Research in multi-level issues*). Bingley, UK: Emerald, in press.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, *83*, 126–137.

Laurenceau, J.-P., & Bolger, N. (2005). Using diary methods to study marital and family processes. Special issue on methodology in family science. *Journal of Family Psychology*, *19*, 86–97.

Raudenbush, S. W., Brennan, R. T., & Barnett, R. (1995). A multilevel hierarchical model for studying psychological change within married couples. *Journal of Family Psychology, 9,* 161–174.

Snijders, T. A. B., & Kenny, D. A. (1999). The social relations model for family data: A multilevel approach. *Personal Relationships, 6,* 471–486.

AU:Please add an in-text citation for Kashy & Kenny, 1990.