

Chapter 3

Deriving Estimators and Their Standard Errors in Dyadic Data Analysis: Examples Using a Symbolic Computation Program

Richard Gonzalez and Dale Griffin

Michael Browne's foundational papers on the analysis of covariance structures (e.g., Browne, 1974, 1982) were influential in our conceptualization of models for dyadic data analysis. He provided a framework for working with covariance structures that makes it relatively easy to derive new results. We began working on dyadic models in the early 1990s, before the explosion of research on multilevel models and random effect models. Motivated by the ubiquity of dyadic data in social psychological research, we recognized that a structural equation modeling (SEM) approach with latent variables representing dyadic similarity would be helpful in modeling multivariate dyadic data. Multilevel models, let alone multivariate multilevel models, were relatively new at that time and we explored their usefulness for dyadic data analysis. But it was the SEM approach and the Browne (1974, 1982) papers that we found most helpful in our initial model formulation.

Our interest was sparked by Kenny and LaVoie (1985), who presented a dyadic latent variable model based on an analysis of variance (ANOVA) approach to the intraclass correlation. They provided an estimator of a latent dyadic correlation but

did not provide significance tests. We wanted to extend the Kenny and LaVoie approach to allow researchers of dyadic and group processes to use a more flexible modeling approach, including both observed and latent variables, covariates, plus tests of mediation, and other standard analytical tools. We wanted closed form estimators based on formal statistical approaches that allowed clear interpretation of the parameters, so we could understand why estimators were sometimes out of bounds, develop better intuition about the parameters, and provide a formal linkage between the latent variable dyadic approach and the well-studied (but often misapplied) aggregate approaches such as the correlation of group means. We wanted our dyadic data analysis contributions to illuminate the analytic issues in a simple manner so the tools could be used to answer research questions in a meaningful manner. We also wanted to provide more intuition about models that partition variance in a multilevel manner into individual-level and group-level sources, so there was a pedagogical aspect to our work as well.

It was in this context of exploration that we came across Michael Browne's very clear theoretical formulation of covariance structures. Both the 1974 article and the 1982 chapter were ahead of their time, and illuminated the deep and magnificent structure that characterizes covariance structures. We studied these papers carefully. We remember the excitement each page provided, and how we spent hours working through each page of the papers. Not only did we learn about covariance structures at a much deeper level, but we also developed an appreciation for the underlying statistical theory and the importance of clear, well-justified recommendations for practicing researchers. This influence continues to run deep in our research and in our thinking, even outside the scope of dyadic data analysis.

The problem we were working on when we learned about the 1974 and 1982 papers was that we had noticed a discrepancy between the standard errors we derived analytically for one of our dyadic models and the numerical standard errors that emerged from structural equation modeling programs. We noticed that the way in which an analyst identifies a latent factor in a structural equation model (i.e., how the scale of a latent variable is set) can influence the standard error in a nonlinear way, thus influencing the Wald Z test for a parameter. We noticed that statistical programs were giving different Wald tests for identical parameters based on how latent variables were identified, that is, whether the latent variance was set to 1 or which of the indicators was given a path weight of 1. These different Wald tests occur even though the goodness-of-fit indices are invariant to where the scale identifying "one" is placed. This confused us because we had derived the standard errors for a special case model by hand and our standard errors were different from those that emerged from the SEM programs. At first we thought we had made a mistake in our derivation, but we eventually discovered that the issue had to do with how most SEM programs approximate the standard errors. Michael Browne's 1974 and 1982 papers were instrumental in giving us the background and understanding to see this relation. This work led to the Gonzalez and Griffin (2001a) paper that illustrated the problem, and provided a solution using the

likelihood ratio test to avoid the curvature approximation (Hessian) that is used in the estimation of the standard error in the Wald test.

It is wonderful to honor Michael Browne's contribution to statistical methodology in psychology. In this chapter we illustrate how we used some of the results in Browne's 1974 and 1982 papers to derive standard errors for particular parameters in standard models of dyadic data analysis. Our goal in this chapter is to illustrate a little of the underlying statistical theory, and to show how one can use a symbolic software package like Maple (<http://www.maplesoft.com/products/maple/>) or Mathematica (<http://www.wolfram.com/mathematica/>) to derive useful results. One of the many things we learned from Michael Browne's scholarly contributions is the importance of analytic derivation to facilitate understanding of the underlying statistical properties. Not all psychology students have the background to perform some of the analytic computations by hand. Computer packages like Maple can go a long way toward empowering students to play with analytic results, provided the students know a few fundamental ideas from calculus and linear algebra. We believe that it is possible to incorporate more theoretical results and properties from statistics into advanced graduate training in psychology, and tools like Maple and Mathematica can provide helpful bridges for students who have not had much quantitative background. If taught in the right way, the student can achieve not only a deeper understanding of the underlying statistical toolbox but also know how to adapt the tools (or know when the tools do not apply) to solve new problems. Given the rapid pace of advances in quantitative psychology, graduate methods training might serve students well by a teaching style that facilitates adaptability to new methods throughout one's career. Textbooks teaching more theoretical statistical concepts using these tools have already begun to emerge (e.g. Rose & Smith, 2002).

3.1 DYADIC DATA ANALYSIS

3.1.1 Intuition, Psychology, and Models

Psychology studies the mind and its related processes, so it becomes especially interesting when we consider two individuals in a social relationship, such as a married couple, a parent and child, or two team members. We can study not just an individual in isolation but also an individual in the context of another person. We can frame this problem in several ways. One framing could be in terms of two levels of analysis: an individual level and a dyadic level. We could explore psychological processes in terms of the dyad's shared variance and refer to what is left over after removing that shared variance as individual-level variance (i.e., a type of residual variance after the shared dyad variance has been removed). This framing of the dyadic analysis problem involves a decomposition of individual- and group-level variance.

A second framing involves a different conceptualization of dyadic interaction, not in terms of shared variance but instead focusing on processes involving "influence" of one person on another. What is the magnitude of one person's influence on the other individual in the dyad? Is the cross-person influence different than a person's influence over his or her own behavior (e.g., a spouse's anxiety influencing her frequency of risky choices versus her own anxiety influencing her own risky behavior)? This second framing of dyadic processes focuses more on the social influence part of the interaction. It pits actor- and partner-influence processes against each other rather than dyadic- and individual-level processes as in the first framing.

The first framing above is called the latent variable interdependence model and the second framing is called the actor-partner interaction model, or APIM (e.g., Gonzalez & Griffin, 2001b; Kenny, Kashy, & Cook, 2006). Both models can be used in the context of modeling a covariance matrix with interdependence. Both models can be expressed either as a structural equation model or as a multilevel model. The latent variable and actor-partner models lead to slightly different generalizations, although they can lead to identical goodness-of-fit under some special cases.

Consider a heterosexual married couple. The two individuals in the couple produce data on the key dependent variable, say, risky behavior. There may also be a predictor variable collected from each dyad member, say, anxiety. The goal is to understand the relation between anxiety and risky behavior in the context of interdependence. The latent variable and actor-partner models approach this goal in different ways. A statistical model can be estimated for each individual but because the two individuals are interdependent, the residuals resulting from such models are correlated. This kind of correlated error is well-known in the econometrics literature, such as seemingly unrelated regression (SUR) models. The goal in those models is to correct the estimation procedure for the presence of such correlated residuals. But the interesting psychological properties may be in the degree of interdependence between the individuals in the dyad. So, rather than treat interdependence as a nuisance that needs to be eliminated from the data to achieve unbiased parameter estimates and tests of significance, as is typically done in SUR models, proponents of dyadic data analysis models conceptualize the correlated error as something of potential or even central psychological interest. The degree of interdependence could inform us about dyadic processes, which could be a theoretically interesting attribute of the data and the model.

The dyadic data structure for this example on two variables (anxiety and risky behavior) for each dyad member has four observations per dyad. The APIM views these as four variables in a two regression framework with correlated residuals and correlated predictor variables. In symbols,

$$Y_w = \beta_{0w} + \beta_{ww} \text{anxiety}_w + \beta_{wh} \text{anxiety}_h + \epsilon_w$$

$$Y_h = \beta_{0h} + \beta_{hw} \text{anxiety}_w + \beta_{hh} \text{anxiety}_h + \epsilon_h$$

where the intraclass correlation $r_{XX'}$ handles the interdependence on the predictor variable anxiety and the residuals are correlated. This model is represented

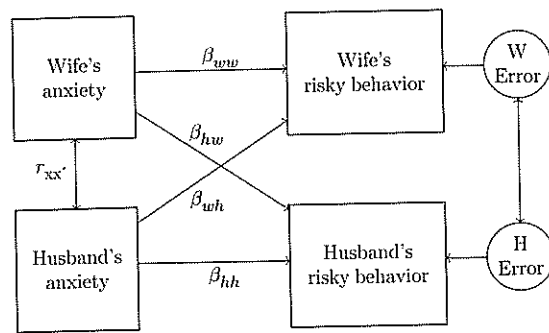


Figure 3.1 Actor-partner model for anxiety and risky behavior. The figure depicts the intraclass correlation $r_{xx'}$ and the correlation between the residuals.

in Figure 3.1. Of course, in the present correlational structure it is not possible to establish causality; it is possible that risky behavior causes anxiety, or that a third variable causes both. Additional information or different designs would be necessary to gain some purchase on the causality question for these variables.

The latent variable model also uses two regression equations, but the usual parameterization involves a partitioning of shared dyadic-level variance and unique individual-level variance. This is accomplished by using two sets of latent variables, or as we will see below, an equivalent formulation in terms of a multi-level model using random effects. One set of latent variables involves individual-level (i.e., residual) variance and the second set of latent variables involves dyadic-level (i.e., factor) variance. In symbols, the variance of each observed variable Y (each square in Figure 3.2, using the convention that observed variables are depicted as squares and latent variables as circles) is modeled as

$$\text{var}(Y) = r_{xx'} \text{var}(\text{shared}) + \text{var}(\text{unique}) \tag{3.1}$$

and the covariances among the observed variables follow standard covariance algebra rules through the r_i and r_d terms. The implementation of the model requires identification constraints, such as setting the dyad variance to one or the indicator path to one. The fitting of this model can be conducted in either a traditional SEM approach, where the observed covariance matrix is compared to the model-implied covariance matrix, or in a multilevel modeling (MLM) approach. The SEM and MLM approaches can give identical results, such as when data are balanced and the same estimation procedure is used (e.g., maximum likelihood estimation). The SEM representation is presented in Figure 3.2.

Models of dyadic processes provide interesting possibilities for psychologists interested in the study of social interaction. The models help quantify a somewhat fuzzy

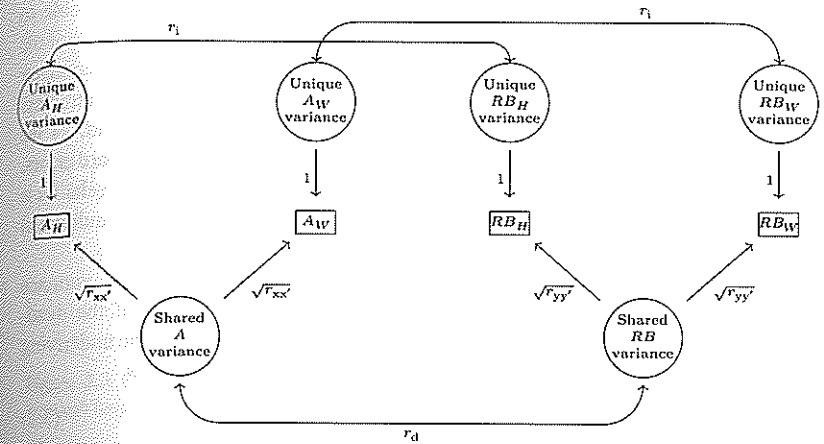


Figure 3.2 Latent variable approach to estimating the intraclass correlation between husband and wife on Anxiety (A) and Risky Behavior (RB), the dyad level correlation r_d , the individual level correlation r_i , and the intraclass correlations $r_{xx'}$ and $r_{yy'}$.

notion about what it means to be interdependent. The models assess interdependence, treating it as a parameter that can be used in more complicated models with additional predictors and outcome variables. The models can be extended to longitudinal data so that temporal processes in dyads can be examined.

In the remainder of the chapter we show how to estimate the parameters of these models in terms of observable covariances and correlations. Interesting relations between statistical parameters emerge, with the intraclass correlation playing a key role. To simplify our analysis we focus on the case of distinguishable dyads, where dyad members can be distinguished on theoretically relevant variables, such as gender in heterosexual couples and age in same-sex sibling pairs. We refer the reader interested in the issue of exchangeable dyads, such as homosexual dating couples and same-sex twins, to Griffin and Gonzalez (1995) and Kenny et al. (2006).

3.2 INTRAClass CORRELATION: UNIVARIATE MODELS

We begin with an essential ingredient to dyadic models: the intraclass correlation. Interdependence between interval scaled data in the context of linear models can be assessed through the intraclass correlation. The basic intuition for the intraclass

correlation is that it is the percentage of total variance that can be attributable to between dyad variance. The intraclass correlation appears in both the APIM and the latent variable model, and we will show that it plays an important role in parameter and interval (i.e., standard error) estimation. In the APIM (Figure 3.1) the intraclass correlation appears as the correlation between the predictor variables and, as we develop later in this chapter, appears in the formula for the β s and their standard errors. In the latent variable model (Figure 3.2) the square root of the intraclass correlation is equal to the respective indicator paths and also appears in the formula for the individual and dyad-level correlations as well as their standard errors.

One standard formulation for the intraclass correlation takes the variance associated with a dyad-level variable and normalizes it by the sum of the dyad-level variance plus the variance of the individual-level error variance. The underlying structural model for an observation Y for the j th person in the i th dyad is

$$Y_{ij} = \mu + \beta_j + \alpha_i + \epsilon_{ij} \quad (3.2)$$

where the μ is the fixed effect constant, β_j is the fixed effect term for the j th dyad member (such as husband and wife), α_i is a random effect term for the dyad that is assumed to be normally distributed with mean 0 and variance σ_α^2 , and the usual error term ϵ with variance σ_ϵ^2 . There are slightly different estimation formulas for the intraclass correlation depending on whether one uses maximum likelihood or restricted maximum likelihood (the latter accounts for degrees of freedom as in an ANOVA) estimation but the basic logic is similar. The intraclass correlation becomes the ratio

$$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2} \quad (3.3)$$

This model (Equation 3.2) can be placed in the context of a multilevel model with two levels, such that the first level represents the raw data and the second level represents the part of the model that is between dyads:

$$Y_{ij} = \gamma_i + \beta_j + \epsilon_{ij} \quad (3.4)$$

$$\gamma_i = \mu + \alpha_i \quad (3.5)$$

where β is a fixed effect term that estimates the difference between the two distinguishable dyad members, γ is a random effect dyad term, and the ϵ is the usual error term. If one substitutes Equation (3.5) into Equation (3.4), then the result is the same as Equation (3.2).

A third way to conceptualize the intraclass correlation is as a structural equation model with two indicators, one latent factor and a specific set of restrictions. If one sets the variance of the latent factor to 1, the two indicator paths to the observed variables equal to each other, and the error variances equal to each other, then the indicator paths are equal to the square root of the intraclass correlation (see Figure 3.3). The two indicator paths are constrained to be equal because the interpretation of the

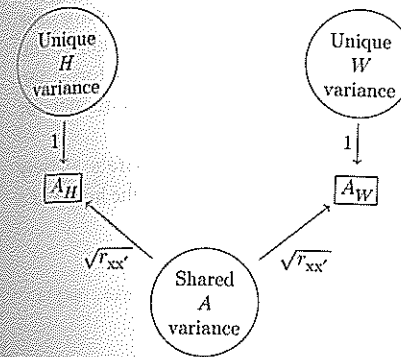


Figure 3.3 Structural equation modeling approach to estimate the intraclass correlation $r_{xx'}$ between husband and wife anxiety scores (A_H and A_W , respectively).

latent variable is one of “shared variance” where both individuals contribute equally. This is a different parameterization and interpretation than the usual latent factor, where indicators can have different paths and can relate differentially to the latent variable.

Thus, there are several ways of conceptualizing the logic of interdependence as indexed by the intraclass correlation, and they lead to the same result. One can model the intraclass correlation as a linear mixed model, as a multilevel model, or as a structural equations model. The results will be the same as long as the same estimation procedure is used (e.g., maximum likelihood) and the proper constraints on the parameters are imposed throughout. All three approaches impose the identical structure on the relevant covariance matrix.

3.2.1 Derivations Using Maple

To illustrate the approach and some basic Maple commands, we present a simple asymptotic Z test for two dependent covariances, such as testing whether the covariance between the wife’s anxiety and risky behavior is different than the covariance for the husband for those two variables. This analysis involves two variables (e.g., anxiety and risky behavior) from two dyad members for a total of four observations per dyad.

Browne (1974) made use of the asymptotic property that the covariance of two elements of a covariance matrix is given by

$$\text{cov}(c_{ij}, c_{kl}) = (c_{ik}c_{jl} + c_{il}c_{jk})/N$$

(see also Kendall & Stuart, 1966). This is useful when deriving standard errors for elements of a covariance matrix, and some of the subscripts i, j, k , and l can be the same value, such as in the case of a variance V_{jj} . This asymptotic result forms the building block for one approach to deriving analytic results for dyadic models of interdependence.

We begin by defining the observed covariance matrix for the wife's anxiety (variable 1), the wife's risky behavior (2), the husband's anxiety (3), and the husband's risky behavior (4). In Maple we type the following command

```
covo := matrix(4, 4, [[v1, c12, c13, c14], [c12, v2, c23, c24],
                    [c13, c23, v3, c34], [c14, c24, c34, v4]]);
```

which produces the matrix of symbolic terms

$$\begin{bmatrix} v1 & c12 & c13 & c14 \\ c12 & v2 & c23 & c24 \\ c13 & c23 & v3 & c34 \\ c14 & c24 & c34 & v4 \end{bmatrix}$$

The observed covariance matrix has 10 unique terms: there are 4 observed variances and 6 observed covariances. We apply the asymptotic result in Equation 3.6 repeatedly to create the 10×10 covariance matrix V of the 10 observed elements in the observed covariance matrix. In Maple one can execute the following commands to initialize the observed covariance matrix:

```
#lines beginning with pound sign are comments
#define procedure to compute asymptotic covariance

#initialize observed covariance matrix
covx := covo;
acov := proc(i,j,g,h);
    covx[i,g]*covx[j,h] + covx[i,h]*covx[j,g];
end;

k := 10;
ind := matrix(k,2);
counter := 0;
for c1 from 1 to rowdim(covx) do
    for c2 from 1 to c1 do
        counter := counter + 1;
        ind[counter,1] := c1;
        ind[counter,2] := c2;
    od;
od;

#initialize variance matrix
V := matrix(k,k);
```

```
apply asymptotic covariance formula to compute every term in V
for c1 from 1 to k do
    for c2 from 1 to k do
        V[c1,c2] := acov(ind[c1,1],ind[c1,2],ind[c2,1],ind[c2,2]);
    od;
od;
```

We illustrate the first 4 rows and columns of the output V

$$\begin{bmatrix} 2v1^2 & 2c12v1 & 2c12^2 & 2c13v1 \\ 2c12v1 & v2v1 + c12^2 & 2v2c12 & c23v1 + c13c12 \\ 2c12^2 & 2v2c12 & 2v2^2 & 2c23c12 \\ 2c13v1 & c23v1 + c13c12 & 2c23c12 & v3v1 + c13^2 \end{bmatrix}$$

Another approach to estimate matrix V is to compute the Hessian matrix directly from the maximum likelihood function, which can also be done in Maple. We opted to illustrate the asymptotic result mostly for simplicity in exposition. In our experience the Hessian approach can be relatively slow in Maple compared to the asymptotic approach we outline here, and additional manipulation is usually necessary to simplify the resulting expressions into forms that are more instructive.

Now that we have the covariances of the observed covariance matrix, the next step in the process of getting a Wald Z test for two dependent covariances is to apply the delta rule to matrix V for the desired function on the observed covariances. The function takes the difference between the covariance of the two variables for the wife C12 and the covariance for the two variables for the husband C34. We need the standard error of that difference in order to compute the Z test. To compute the standard error of a difference of two covariances, we take the Jacobian of the difference of the two covariances with respect to all 10 observed covariances (with the terms listed in the same order as the matrix V). The Maple command is

```
da := jacobian([c12-c34], [v1, c12, v2, c13, c23, v3, c14, c24, c34, v4]);
```

with output

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

The final step in computing the standard error for the difference C12-C34 involves applying the delta rule by pre- and post-multiplying the matrix V with the Jacobian we just computed (following standard rules of linear algebra to achieve the multivariate extension of multiplying V by the square of the first derivative).

```
new := multiply(da, V, transpose(da));
```

with output

$$\begin{bmatrix} v2v1 + c12^2 - 2c24c13 - 2c23c14 + v4v3 + c34^2 \end{bmatrix} \quad (3.6)$$

We now have all the ingredients to write the equation symbolically for the Z test between wife's $C12$ and husband's $C34$ (i.e., the test for equality between the wife's covariance between risk and anxiety and the husband's covariance between risk and anxiety). This Z test is needed to test one of the equality constraints necessary to justify the pooling across dyad members such as equivalent individual-level correlations r_i for each dyad member. The Z test for two dependent covariances (e.g., $C12$ and $C34$) is given by

$$Z = \frac{(c_{12} - c_{34})\sqrt{N}}{\sqrt{v_1 v_2 + v_3 v_4 + c_{12}^2 + c_{34}^2 - 2(c_{13}c_{24} + c_{14}c_{23})}}$$

The denominator is the square root of the variance of the difference between two covariances that we computed using the asymptotic result and the delta rule (Equation 3.6). This Z can be compared to the critical value of 1.96 for a two-tailed test at $\alpha = .05$.

The computation we just developed shows that it is relatively straightforward to compute the standard error of a function of elements of a covariance matrix. The same logic extends to dyadic models (and SEM more generally) where we estimate parameters that are more complicated functions of elements from the observed covariance matrix and need to estimate the standard errors of those parameters. We illustrate with the computation of the intraclass correlation for the dyad (as depicted in Figure 3.3). The one key addition to the setup is that we first estimate the parameters directly from the likelihood function following Browne's (1974, 1982) papers.

We begin with the observed covariance matrix for two observed variables, say anxiety measured in each of the husband and wife. The covariance matrix is

$$\begin{bmatrix} v1 & c13 \\ c13 & v3 \end{bmatrix}$$

We implement the model in Figure 3.3 by parameterizing it with paths set to 1, though we could have parameterized it by setting the latent variance to one and estimating the constrained paths. We let $vlat$ represent the variance of the dyad-level latent variable and ve represent the variance of the residual, or individual-level latent variable. The model-based covariance matrix following the SEM approach to the intraclass correlation in Figure 3.3 is

$$\begin{bmatrix} vlat + ve & vlat \\ vlat & vlat + ve \end{bmatrix}$$

The following Maple commands set up the likelihood function (Browne, 1974, 1982), take derivatives and solve for each of the two parameters $vlat$ and ve in terms of the observed covariance matrix.

```
fitfuncml1 := normal(rationalize((-N/2) * (log(factor(det(covm)) +
trace(multiply(covo, inverse(covm))))));
g1 := grad(fitfuncml1, [vlat, ve]);
mle := solve(convert(g1, set), {ve, vlat});
```

We then run the same asymptotic variance procedure as above, with $k = 3$ since there are three unique elements in the observed covariance matrix, and apply the delta rule to derive the standard error.

```
templ :=
transpose(jacobian([rhs(mle[1]), rhs(mle[2])], [covo[1, 1], covo[1, 2], covo[2, 2]]));
se := scalarmul(multiply(transpose(templ), scalarmul(V, 1/N), templ), 1)
```

The resulting standard error for the variance of the dyad-level latent variance is

$$\sqrt{\frac{v3 v1 + c13^2}{N}}$$

If one plugs the three observed values from the covariance matrix into this expression, the result will be identical to the numerical result in SEM programs, as long as the covariances are computed the same way as one's program (i.e., either N or $N - 1$ in the denominator).

These examples illustrate the usefulness of a symbolic system to do some relatively complicated operations that may not be within the reach of many students in the social sciences. Standard errors are placed in the context of observed elements of the covariance matrix, so one can understand the components of the standard error to get a better sense of how the confidence interval and the test of significance behave as a function of the observed covariances. The understanding that can emerge by examining the definitional equations of both the parameter estimates and their standard errors can provide a deeper intuition about the underlying estimates and their tests. We now turn to a discussion of the latent variable model, showing maximum likelihood estimates and their standard errors.

3.3 INTRAClass CORRELATION: MULTIVARIATE MODELS

The model depicted in Figure 3.2 can be estimated (equivalently) using either SEM or MLM formulations. We begin with the SEM approach to the multivariate latent dyadic model as per Gonzalez and Griffin (1999). The model is a simple generalization of the single variable case (Figure 3.3). Each variable has the "V" shape graphical representation, with a single latent variable representing the dyadic variance, the variables for each dyad member having the unique error variance, and the indicators being the square root of the respective intraclass correlation. The multivariate structure

adds correlations across the separate “V” structures in the graph. There is a set of individual-level correlations r_i that represents the correlations between a single dyad member’s error variance across two variables, and a dyad-level correlation r_d that represents the correlation between the two dyad-level latent factors. This model performs a partitioning of individual-level and dyad-level variance separately for each variable and correlates across these partitioned terms to estimate r_i and r_d , respectively. In the case of distinguishable dyads one can also test the assumption of equality of the two individual-level correlations r_i .

The Gonzalez and Griffin (1999) paper presents maximum likelihood estimates of the two intraclass correlations (one for each variable), r_i and r_d as well as their standard errors and Wald Z tests. The paper also presents SEM syntax. In terms of the present chapter, the derivation of the parameter estimates and their standard error proceeds just like in the previous section, where the maximum likelihood estimate for the intraclass correlation and its standard error was presented. The maximum likelihood estimates for the four key parameters (r_{xx} , r_{yy} , r_i , and r_d) are estimated by taking derivatives of the likelihood function, setting each derivative equal to zero, and solving for the derivatives. Formulas for the asymptotic standard errors follow the procedure discussed in the previous section of using the asymptotic result for a covariance of two covariance terms (Equation 3.6) and the delta method.

Here we highlight a few key aspects of the parameter estimate for r_d and its standard error. As shown in Gonzalez and Griffin (1999), r_d is related to the intraclass correlations of both variables. So, in our running example, the intraclass correlation for the anxiety variable and the intraclass correlation for the risky behavior variable appear as reliability corrections in the formula for r_d (with small shared variance implying low reliability of measuring the dyadic covariance, and hence an “inflation factor” that increases r_d). This closed form estimate for r_d provides an explanation for why in some datasets the estimate of r_d is out of bounds (e.g., greater than 1). If the product of the intraclass correlations is relatively close to zero, then r_d can be overly inflated. One should question the applicability of the latent variable model when either or both intraclass correlations are close to zero—there may not be sufficient shared variance (interdependence) on which to develop a dyadic model. Another interesting point to highlight is the role that the intraclass correlations play in the standard error of r_d . Using the method described above we showed (Gonzalez & Griffin, 1999; Griffin & Gonzalez, 1995) that the product of the two intraclass correlations also enters the standard error of r_d . All other things being equal, the standard error of r_d will be smaller the greater the product of the two intraclass correlations.

3.3.1 Connections to Multilevel Model

This latent variable model can be equivalently estimated in the context of a multilevel model. We develop the parallel language for multilevel models because it is instructive

to show the equivalence of the two approaches, and we encourage students to develop the skills to switch easily between the two approaches. Unfortunately, the description of the multivariate model is not straightforward because multilevel models are typically expressed for a single variable with a single residual term ϵ at level 1. But we need separate error terms for each of the two variables, and in the case of distinguishable dyads, possibly separate error terms for each dyad member. In order to fit the multilevel model to the multivariate example in Figure 3.2 it is necessary to express the multilevel model in a way that allows for different residual variances for each variable and person. We present one method to model multiple variables in the multilevel model using a switching regression technique.

The first level model is

$$Y = \beta_a D1 + \beta_r D2 \quad (3.7)$$

One dummy code $D1$ assigns a 1 to all anxiety scores and a 0 to all risky behavior scores. The other dummy code $D2$ assigns a 1 to all risky behavior scores and a 0 to all anxiety scores. The first level equation does not include an error variance nor an intercept—those terms appear in the second level separately for each variable. Thus, this switching regression merely partitions the “outcome variable Y ” into either anxiety scores or risky behavior scores (in our running example).

The second level of regression equations are:

$$\beta_a = \text{intercept}_a + v_{ag} + u_{ai} \quad (3.8)$$

$$\beta_r = \text{intercept}_r + v_{rg} + u_{ri} \quad (3.9)$$

where v and u are random effects that code group and individual terms (respectively), each equation has its own fixed effect intercept term, the subscripts a , r , g , and i refer to anxiety, risky behavior, group and individual, respectively. The random effect v assesses group-level variance and the random effect u assesses individual-level variance.

Now comes the important part of this particular formulation. We formulate a covariance structure on each of the random effects v (group-level) and u (individual-level). Let the two group-level v 's be bivariate normally distributed with covariance matrix

$$\Omega_v = \begin{bmatrix} \sigma_{va}^2 & \\ \sigma_{vr} & \sigma_{vr}^2 \end{bmatrix} \quad (3.10)$$

where a and r denote anxiety and risky behavior, respectively. This means that the random effect v associated with anxiety has variance σ_{va}^2 , the random effect v associated with risky behavior has variance σ_{vr}^2 , and the two v s have covariance σ_{vr} . Similarly, an analogous covariance is formulated on the two individual-level u s

$$\Omega_u = \begin{bmatrix} \sigma_{ua}^2 & \\ \sigma_{ur} & \sigma_{ur}^2 \end{bmatrix} \quad (3.11)$$

This covariance matrix Ω_u gives the variances and covariance between anxiety and risky behavior at the individual level. In this formulation we require equality of all individual-level correlations (i.e., referring back to Figure 3.2, this particular multilevel model implementation forces the two individual-level correlations for husband and wife to be identical).

These two matrices provide all the information necessary to compute the terms r_d and r_i in the latent group model as well as each of the two intraclass correlations. Using the terms in those two covariance matrices we have

intraclass correlation for anxiety:

$$\frac{\sigma_{va}^2}{\sigma_{va}^2 + \sigma_{ia}^2}$$

intraclass correlation for risky behavior:

$$\frac{\sigma_{vr}^2}{\sigma_{vr}^2 + \sigma_{ir}^2}$$

individual level correlation between anxiety and risky behavior:

$$\frac{\sigma_{uar}}{\sqrt{\sigma_{ia}^2 \sigma_{ir}^2}}$$

dyad level correlation between anxiety and risky behavior:

$$\frac{\sigma_{vda}}{\sqrt{\sigma_{va}^2 \sigma_{vr}^2}}$$

This multilevel model formulation leads to identical parameters (i.e., intraclass correlations, r_i and r_d correlations) to the SEM approach as long as one uses similar estimation methods, such as maximum likelihood, across both multilevel and SEM approaches. One can use Maple to estimate these parameters directly from the likelihood function to get closed form solutions. Of course, the closed form tractability of these estimates and their standard error follows because we have equal sized clusters (dyads). It becomes difficult to derive closed form solutions when the clusters have different sizes, such as the case when individuals are nested in families and the sample includes families of different sizes. In that case numerical methods are needed, as implemented in popular multilevel modeling programs.

3.3.2 Actor-partner Model

The actor-partner model (Figure 3.1) provides another opportunity to highlight how the exercise of deriving closed form solutions for parameter estimates and their standard errors provides deeper understanding of how the model works. As with the latent

variable model, the intraclass correlations play an important role. Before presenting the regression parameter and its standard error, we discuss the pairwise approach to the actor-partner model that we described in Gonzalez and Griffin (1999) and Griffin and Gonzalez (1995). The pairwise approach organizes the data matrix in "long form" in such a way that each variable is entered twice (e.g., in one column the data are ordered as husband and wife anxiety, in another column the same data are reordered as wife and husband anxiety). We call one variable X or Y and its reordered form X' or Y' . We showed that the maximum likelihood intraclass correlation for variable X is equivalent to the Pearson correlation between X and X' (e.g., Griffin & Gonzalez, 1995). The pairwise organization to the data makes some derivations a little easier, and they follow the procedure described earlier for estimating parameters and their standard errors.

In the case of the actor-partner interaction model described in Figure 3.1 let's define the predictors as X and the dependent variable as Y . The raw score actor effect (i.e., the actor path from X to Y in Figure 3.1) is given by

$$\frac{s_y(r_{xy} - r_{xy'}r_{xx'})}{s_x(1 - r_{xx'})} \tag{3.12}$$

where s_y and s_x are the standard deviations of the criterion variable X and the predictor variable Y , respectively. It is interesting that the intraclass correlation for the predictor variable X appears both in the numerator and the denominator (in squared form) for the regression parameter. One can see that the regression parameter increases with increasing values of the intraclass, and when the intraclass correlation is 1 (perfect agreement between dyad members) the regression parameter is undefined, as expected. The interested reader can refer to our papers for interpretations of the correlations r_{xy} and $r_{xy'}$. The regression coefficient for the partner effect has the same form with the role of r_{xy} and $r_{xy'}$ interchanged.

The standard error of the regression weights is also instructive. Under the null hypothesis that the population $\beta = 0$, the estimate of the variance for the actor regression slope reduces to

$$V(\beta_{actor}) = \frac{s_y^2(r_{xy'}^2 r_{xx'}^2 - r_{xx'} r_{xy'} + 1 - r_{xy'}^2)}{2Ns_x^2(1 - r_{xx'}^2)} \tag{3.13}$$

where N is the number of dyads. The standard error for the partner regression effect is analogous, except that r_{xy} appears in Equation 3.13 in place of $r_{xy'}$. Clearly, the intraclass correlation of the predictor variable X and the dependent variable Y play an important role in the standard error of the slopes in the APIM.

If the APIM is constrained to have equal actor paths and equal partner paths (e.g., husband to wife is constrained to equal wife to husband), then the APIM yields the identical goodness-of-fit to the latent variable model under the constraint that the two individual-level correlations r_i (one for each dyad member) are constrained to be

equal. In other words, they are identical models under a set of constraints. The APIM can be generalized in different ways from the latent variable model, and in this sense the two models can assess different dyadic processes.

3.3.3 Summary of Multilevel Models

The take-home message of this type of modeling is that the observed variances and covariances between variables and between individuals in a dyad can be partitioned into individual-level and dyad-level terms or into an actor-partner regression framework. The approaches can be formulated either as structural equation models or as multilevel models. Once the partition is performed, substantive research questions can be tested at each level. For example, one can test for predictors of the dyad and individual-level terms, and can examine the outcome variables associated with the decomposed variances. Of course, these models can be extended to cases with more than two variables. For example, in the latent variable model one would have a separate "V-structure" for each variable to model each variable's interdependence. One could then correlate the individual-level variances and the dyadic-level variances across variables, or could construct more complicated models with directional paths (e.g., the dyadic variance from one variable predicting the dyadic variance of another variable, mediational models at each of the individual-level and dyadic-levels, etc.). The actor-partner model could also be extended to multiple variable and even longitudinal designs; one needs to keep track of all the actor paths and the partner paths across the different variables and times as one would do in standard cross-lagged designs.

Further, the two modeling approaches highlight different dyadic processes. The latent variable model highlights shared and individual variance processes (e.g., conformity and social norms), whereas the actor-partner model highlights social influence. It is in this way that the analysis of dyadic data goes beyond "correcting" data for violations of independence by directly modeling the interdependence that characterizes a multivariate dyadic data set. The interdependence, as captured by the intraclass correlation and related dyadic measures, plays an important role throughout these models.

3.4 CONCLUSIONS

This chapter had a modest goal. We wanted to illustrate a few analytic derivations of the parameter estimates and their standard errors for dyadic analysis. We demonstrated that increasingly easy-to-use computer programs that perform symbolic computations can make it efficient and educational to derive analytic results in a relatively painless way. The potential for this type of training and understanding is very high. In our

simple examples we showed how the intraclass correlation, a standard measure of interdependence used in dyadic analysis, plays a major role not only in parameter estimates but also their standard errors. It is not clear that such an understanding would emerge from only studying simulations, even if they systematically varied the relevant parameters. There is value in analytic derivation in terms of the understanding that can emerge.

The contributions of Michael Browne to covariance analysis and SEM are much deeper than the simple examples that we demonstrated in this chapter. But if we have succeeded in exciting the reader about the role for analytic exercises in developing deeper understanding of models of covariance structures, and the use of computer software in achieving that goal, then we have played a small role in highlighting the deeper contribution that Michael Browne has made to quantitative psychology.

REFERENCES

- Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistics Journal*, 8, 1-24.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge: Cambridge University Press.
- Gonzalez, R., & Griffin, D. (1999). The correlational analysis of dyad-level data in the distinguishable case. *Personal Relationships*, 6, 449-469.
- Gonzalez, R., & Griffin, D. (2001a). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods*, 6, 258-269.
- Gonzalez, R., & Griffin, D. (2001b). A statistical framework for modeling homogeneity and interdependence in groups. In M. Clark & G. Fletcher (Eds.), *Handbook of social psychology, Vol 2: Interpersonal processes* (pp. 505-534). Mass: Blackwell.
- Griffin, D., & Gonzalez, R. (1995). The correlational analysis of dyad-level data: Models for the exchangeable case. *Psychological Bulletin*, 118, 430-439.
- Kendall, M., & Stuart, A. (1966). *The advanced theory of statistics* (2nd ed.). London: Griffin.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.
- Kenny, D. A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology*, 48, 339-348.
- Rose, C., & Smith, M. (2002). *Mathematical statistics with Mathematica*. New York: Springer-Verlag.