Dyadic Data Analysis Using Multilevel Modeling: SRM Designs

David A. Kenny   Deborah A. Kashy

University of Connecticut   Michigan State University

SRM Designs

In these designs, each person is paired with more than one partner and each partner is also paired with multiple others. 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. 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. In addition to the round-robin design, the SRM can also be estimating 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.

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 and is denoted as σ_{ab}. At the dyadic level, the two members' relationship effects can be correlated and is denoted as σ_{gg}. 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,

σ_g^2: error variance in perceptions of intelligence

σ_{ab}: the covariance between how intelligent a person generally sees others with how intelligent other people generally see that person

σ_{gg}: the covariance between how one person uniquely sees another with how that other uniquely sees that person

MLM with Dummy Variables. The approach described here was initially proposed by Snijders and Kenny (1999), and involves creating $2n$ dummy variables, where n is the number of persons in the largest round robin. For example, assuming the largest group size is 5, we create the dummy variables *a1* through *a5* (e.g., the dummy variable *a3* is one if the actor is person 3 and zero otherwise), and *p1* through *p5* (e.g., the dummy variable *p3* is one if the partner is person 3 and zero otherwise). The model is specified via a series of constraints on the variance-covariance matrix of the dummy variables.

The multilevel model has three levels. Level 3 is group, level two is dyad, and level one is observation. Dyad refers to a pair of persons. Thus, persons 1 and 2 are in dyad 1, 1 and 3 are in dyad 2, 1 and 4 are in dyad 3, and so on. The variables *a1* through *a5* are random at level 3 with a zero mean, equal variances, and all covariances between them zero. Similarly, the variables *p1* through *p5* are random at level 3 with a zero mean, equal variances, and all covariances between them zero. Each "a" variable is correlated with its "p" variable; for instance, *a1* is correlated with *p1*. All other covariances between the "a" and "p" variables are set to zero.

As described by Snijders and Kenny (1999) these constraints can be imposed in MLwiN. To the best of our knowledge, these constraints cannot be imposed on the variance-covariance matrix within SPSS or HLM. Here we describe how to estimate this model using SAS. First, we create the dummy variables for a five-person round robin in which there is a variable for *actor* that ranges from 1 to 5 and a variable for *partner* that also ranges from 1 to 5:

a1 = 0; *a2* = 0; *a3* = 0; *a4* = 0; *a5* = 0;

IF *actor* = 1 THEN *a1* = 1;

….

IF *actor* = 5 THEN *a5* = 1;

p1=0; *p2*=0; *p3*=0; *p4*=0; *p5*=0;

IF *partner* = 1 THEN *p1* = 1;

….

IF *partner* = 5 THEN *p5*=1;

To impose the constraints on the variance-covariance matrix of the random effects, a dataset, which in the syntax below is called *g*, is created to set the *n* actor variances (parameter 1) equal, the *n* partner variances (parameter 2) equal, and the *n* actor-partner covariances (parameter 3) equal. The structure of this dataset for a five-person group is as follows.

DATA *g*;

INPUT PARM ROW COL VALUE;

DATALINES;

1 1 1 1

1 2 2 1

…

1 5 5 1

2 6 6 1

2 7 7 1

…

2 10 10 1

3 1 6 1

3 2 7 1

…

3 5 10 1

4 11 11 1

;

The structure of each record in the *g* dataset is very specific. The first value is the parameter

number (e.g., 1 refers to the actor variance, 2 refers to the partner variance, 3 refers to the actor-

partner covariance, and 4 refers to the group variance). The second value is the row of the 11 by

11 variance-covariance matrix that includes variables *a1-a5, p1-p5*, and *group* (the order of this

matrix is defined by the order of variables included in the RANDOM statement, see below)*,* and

the third value is the column of that same matrix. The final number specifies the value in the

variance-covariance matrix. We also discuss the meaning and structure of the *g* matrix in the

section on Overtime Data – Indistinguishable Dyads and the reader might wish to review that

section.

The SAS syntax that specifies the SRM for round-robin data using dummy variables is:

PROC MIXED COVTEST;

  CLASS *dyad group*;

  MODEL *y* = / S DDFM=SATTERTH;

  RANDOM *a1 a2 a3 a4 a5 p1 p2 p3 p4 p5* INTERCEPT

    / G SUB=GROUP TYPE=LIN(4) LDATA= *g*;

  REPEATED   /TYPE=CS SUB= *dyad*(*group*);

Note that "LDATA = *g*" statement in the RANDOM statement tells SAS that the name of the

dataset that includes the equality constraints is *g*. The "4" in LIN(4) refers to the four parameters

referred to in the *g* dataset. In the output, the parameters have the following meaning:

  LIN(1): actor variance or σ_a^2

LIN(2): partner variance or $\sigma_b{}^2$

LIN(3): actor-partner covariance or σ_{ab}

LIN(4): group variance or $\sigma_g{}^2$

SAS with dummy variables and MLwiN yield essentially the same estimates even though they use somewhat different estimation methods.

Round-Robin Designs with Distinguishable Group Members

In some cases members of the round-robin may be distinguishable by the role they play within the group, and in these situations there may be different actor, partner, and dyad effects for the different roles. The prototypical example of a round-robin with roles is a study of families in which each member of the family (e.g., mother or M, father or F, older sibling or O, and younger sibling or Y) reports on his or her relationship with every other family member. Given such data, the mother's actor effect would reflect the similarity in her ratings of her relationships with the father and the two children, and the mother's partner effect would reflect the degree to which the children and the father report having similar relationships with the mother. Thus, with a four-person family in which each person falls into a specific role, there can be four actor variances, four partner variances, four actor-partner covariances, 12 dyad-specific residual variances, and 6 dyadic reciprocities.

Kenny et al. (2006) discuss round-robin designs in which members are distinguishable extensively, but they emphasize a confirmatory factor analysis approach for analyzing such designs. The SRM with roles can be estimated using MLM with dummy codes. We again create a set of dummy codes that specify who the actor is and who the partner is for each dyadic score. In this case, however, the dummy codes specifically refer to the roles. For example, **aM** = 1 if the actor is the mother and 0 otherwise, **pF** = 1 if the partner is the father and 0 otherwise, and so

on. As in the distinguishable case, we need a *group* variable that identifies each family, and we also need a *dyad* variable that identifies which dyad generated the data. We also need to create a variable which we call *r* that has $n(n-1)$ levels representing the directionality of the dyadic scores. For example, *r* would be 1 for M's rating of F, *r* would be 2 for F's rating of M, and so on. If the study were to include 20 four-person families, *group* would range from 1 to 20, *dyad* from 1 to 6, and *r* from 1 to 12

We then impose fewer constraints in the variance-covariance matrix of the random effects as compared to the indistinguishable case because the variances and covariances can differ as a function of role. We again name the dataset that specifies the constraints on the variance-covariance matrix as *g*. For a study of four-person families this dataset would be:

DATA *g*;

INPUT PARM ROW COL VALUE;

DATALINES;

1 1 1 1

2 2 2 1

3 3 3 1

4 4 4 1

5 5 5 1

6 6 6 1

7 7 7 1

8 8 8 1

9 1 5 1

10 2 6 1

11 3 7 1

12 4 8 1

13 9 9 1

;

The SAS syntax that specifies the SRM with roles is:

PROC MIXED COVTEST;

    CLASS *dyad group r*;

    MODEL *y* = *r* / S DDFM=SATTERTH;

    RANDOM *aM aF aO aY pM pF pO pY* INTERCEPT /G SUB=*group*

        TYPE=LIN(13) LDATA = *g*;

    REPEATED   /TYPE=CSH SUB=*dyad*(*group*) GRP=*dyad* ;

Notice that in the RANDOM statement there are now 13 parameters estimated as specified in the

g matrix. Note also that by using a CSH error structure and the GRP = dyad option, the

REPEATED statement results in estimates of 12 error variances and 6 error covariances (the

dyadic reciprocities.

We could compare the fit of the models that treat dyads as distinguishable and as

indistinguishable. We again would use ML estimation and the chi square for groups of size 4

would have 36 degrees of freedom (9 constraints in the G matrix, 16 in the error variance

covariance matrix, and 11 in the means).

Analysis of Block Designs

The SRM variances and covariances can also be estimated using block designs. In a

block design, the group is divided into two subgroups (e.g., persons A, B, and C might be one

subgroup and persons D, E, F, and G might be the other subgroup) and each group member rates

or interacts with members of the other subgroup. There are three different variants of the block design: half block, symmetric block, and asymmetric block. The half-block design is unique among the SRM designs in that in this design each person is either an actor or a partner, but not both. For example, persons A, B, and C might rate persons D, E, F, and G, but not vice versa. In other words, the half-block design is not reciprocal. The asymmetric block design occurs when there is a variable that distinguishes between members of the two subgroups. For example one subgroup could be managers (persons A, B, and C) and the other subgroup could be employees (persons D, E, F, and G). In the asymmetric block design the managers would rate the employees and the employees would rate the managers. Because of this asymmetry, the SRM components are estimated separately as a function of the distinguishing variable. Finally, in the symmetric block design there is no distinguishing variable between the two subgroups, and so the SRM estimates are pooled across the subgroups to derive the variance estimates. The symmetric block is the block design that is most similar to the round-robin design.

Because block designs involve two subgroups that rate or interact with members of the other subgroup we need to create unique identifiers for members of the two subgroups. We denote these identification variables as *a* and *b*. We also need to create two dummy variables that represent the directionality of the ratings. The first dummy variable, *ab*, would be 1 for data in which the A participants are actors and B participants are partners, and *ab* would be 0 otherwise.   The second dummy variable, *ba* would be 1 when the B participants are actors and the A participants are partners, and 0 otherwise.

Half block. In this design, just one of the subgroups rates or interacts with members of the other group subgroup. We presume that the A's are the actors and the B's are the partners.

Because the data are one-sided, there are no actor-partner or dyadic covariances. This turns out

to be a relatively simple MLM. The syntax for SAS is as follows:

PROC MIXED COVTEST;

CLASS *a b group*;

MODEL *y* = / S DDFM=SATTERTH;

RANDOM INTERCEPT / TYPE=VC SUB= *a*(*group*);

RANDOM INTERCEPT / TYPE=VC SUB= *b*(*group*);

RANDOM INTERCEPT / TYPE=VC SUB=*group*;

The syntax for SPSS is as follows:

MIXED

y BY *a b group*

/PRINT = SOLUTION TESTCOV

 /RANDOM INTERCEPT | SUBJECT(*group*) COVTYPE(VC)

 /RANDOM INTERCEPT | SUBJECT(*a*) COVTYPE(VC)

 /RANDOM INTERCEPT | SUBJECT(*b*) COVTYPE(VC).

Note that for SPSS the values of *a* and *b* must be unique codes for persons in each group.

Asymmetric block. In the asymmetric block design the two half-blocks differ as a

function of the distinguishing factor, and so there may be differences between the *ab* data and the

ba data. To analyze an asymmetric block SRM using SAS, we need to create a categorical or

classification variable that distinguishes between the two blocks of data. This new variable,

block in the syntax below, would be identical to *ab*, and should be included in the CLASS

statement. The SAS syntax for the asymmetric block SRM design is:

PROC MIXED COVTEST;

CLASS *a b group block*;

MODEL *y* = *ab ba*   /NOINT S;

RANDOM *ab ba*/ TYPE=CSH SUB=*group*;

RANDOM *ab ba* / TYPE=CSH SUB= *a*(*group*) ;

RANDOM *ab ba* / TYPE=CSH SUB=*b*(*group*) ;

REPEATED *block* / TYPE=CSH SUBJECT=*a*b*(*group*);

The syntax for SPSS is

MIXED

y BY *a b group* WITH *ab ba*

/FIXED = *ab ba* | NOINT

/PRINT = SOLUTION TESTCOV

/RANDOM *ab ba* | COVTYPE(CSH) SUBJECT(*group*)

/RANDOM *ab ba* | COVTYPE(CSH) SUBJECT(*a*)

/RANDOM *ab ba* | COVTYPE(CSH) SUBJECT(*b*)

/REPEATED *ab ba* | COVTYPE(CSH) SUBJECT(*a*b*group*).

Again, for SPSS the values of *a* and *b* must be unique codes for persons in each group.

Symmetric block. For this design, both subgroups rate or interact with members of the other subgroup and there are presumed to be no differences between members of the two subgroups. We recommend treating the symmetric block design as a round-robin design with missing data. In this case, the syntax we presented for the indistinguishable round-robin design would be used. Note that the fit of the symmetric and asymmetric block designs could be compared to determine if the asymmetry makes an empirical difference.