

Dyadic Data Analysis

Richard Gonzalez
University of Michigan

Dale Griffin
University of British Columbia

February 16, 2011

Dyads: Homogeneity and Interdependence

1. Psychological rationale
2. Statistical framework
3. Examples and intuition

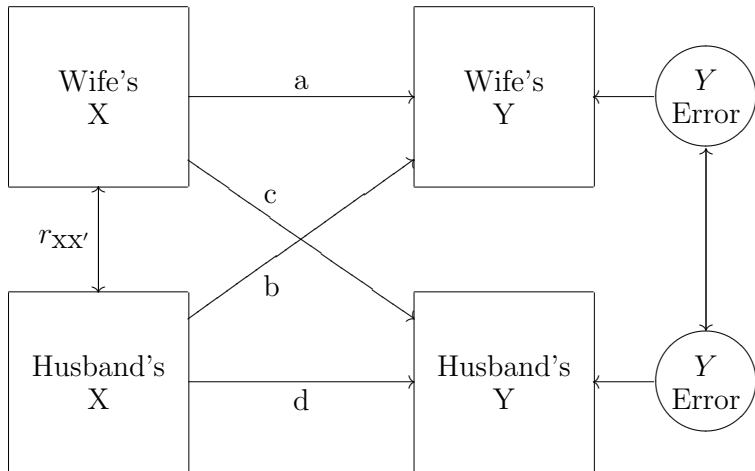
Dyads: Homogeneity and Interdependence

1. Psychological rationale
2. Statistical framework
3. Examples and intuition

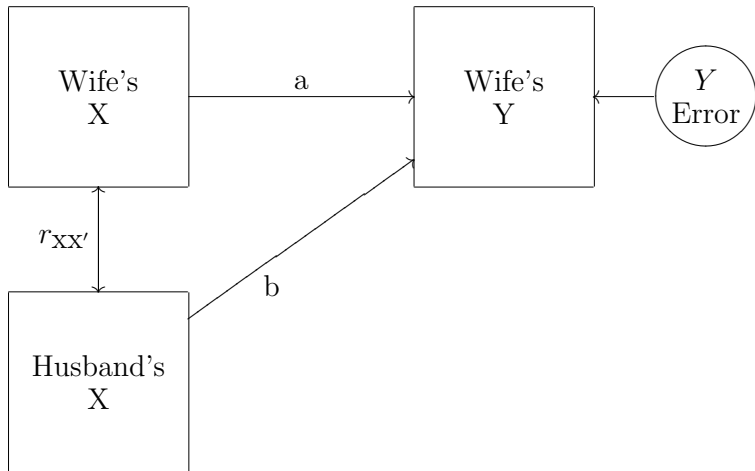
Beginning, middle and end, but not necessarily in that order

Visualization Demonstration

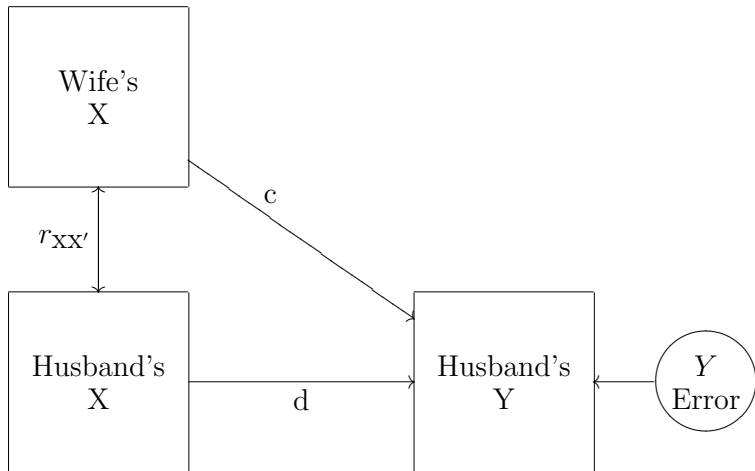
Visualization Demonstration



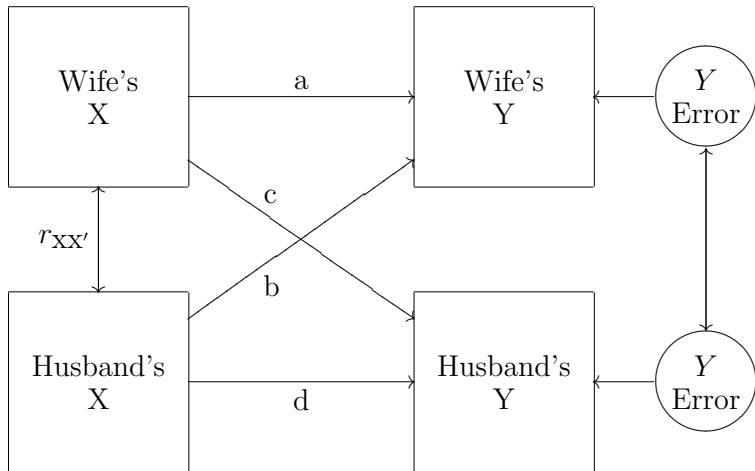
Visualization Demonstration



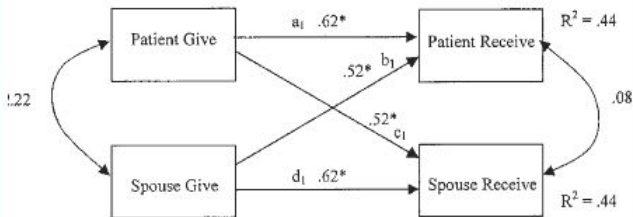
Visualization Demonstration



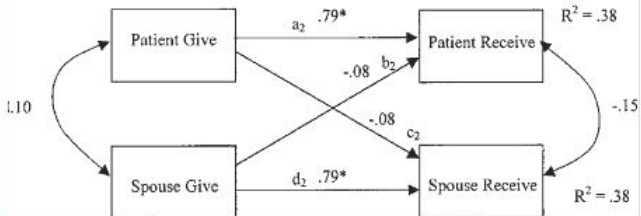
Visualization Demonstration



Couples in similar stages of exercise change



Couples in different stages of exercise change



Why Homogeneity and Interdependence?

How is an individual similar to or different from their spouse in thought, behavior or affect (e.g., shared norms)?

How do couple members influence each other?

Why Homogeneity and Interdependence?

How is an individual similar to or different from their spouse in thought, behavior or affect (e.g., shared norms)?

Similarity can be operationalized by shared variance and correlated error

How do couple members influence each other?

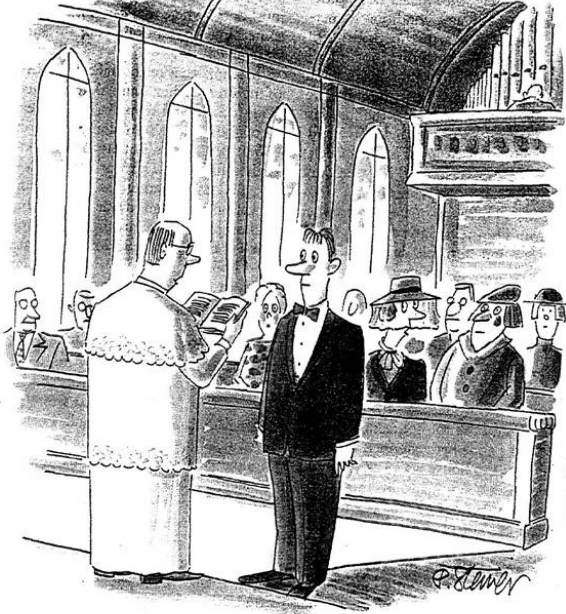
Influence can be operationalized by regression paths

Why Homogeneity and Interdependence?

How is an individual similar to or different from their spouse in thought, behavior or affect (e.g., shared norms)?

How do couple members influence each other?

Bottom line: We shouldn't analyze data from dyads as individuals



"They decided on separate weddings."

Interdependence

- We know about **temporal dependence** (time series, repeated measures, growth curves) and **multivariate structure** (factor and SEM models)

Interdependence

- We know about **temporal dependence** (time series, repeated measures, growth curves) and **multivariate structure** (factor and SEM models)
- We weaker intuitions about interdependence due to social interaction or pairing

Interdependence

- We know about **temporal dependence** (time series, repeated measures, growth curves) and **multivariate structure** (factor and SEM models)
- We weaker intuitions about interdependence due to social interaction or pairing
- We have good statistical models for each (e.g., HLM, SEM, latent growth curves), but lack a complete understanding of how these frameworks interrelate

Desiderata for a General Framework

single conceptual framework for relations due to time, grouping, and multiple variables.

Desiderata for a General Framework

single conceptual framework for relations due to time, grouping, and multiple variables.

should handle covariates and common procedures such as mediation and moderation.

Desiderata for a General Framework

single conceptual framework for relations due to time, grouping, and multiple variables.

should handle covariates and common procedures such as mediation and moderation.

flexible estimation and testing procedures (GLS, ML, REML, MCMC, bootstrap); deal with missing data and sample weights; deal with different distributions and models (e.g., generalized linear models, generalized additive models)

Desiderata for a General Framework

single conceptual framework for relations due to time, grouping, and multiple variables.

should handle covariates and common procedures such as mediation and moderation.

flexible estimation and testing procedures (GLS, ML, REML, MCMC, bootstrap); deal with missing data and sample weights; deal with different distributions and models (e.g., generalized linear models, generalized additive models)

easy to use with standard designs but flexible to deal with nonstandard design elements

Nonindependence

Correlations due to

Nonindependence

Correlations due to

- temporal clustering

Nonindependence

Correlations due to

- temporal clustering
- variable clustering

Nonindependence

Correlations due to

- temporal clustering
- variable clustering
- interpersonal clustering

Independence

| 3 | Y | Dyad | Person | Time |
|----|-----|------|--------|------|
| 4 | y1 | 1 | 1 | 1 |
| 5 | y2 | 1 | 1 | 2 |
| 6 | y3 | 1 | 1 | 3 |
| 7 | y4 | 1 | 1 | 4 |
| 8 | y5 | 1 | 2 | 1 |
| 9 | y6 | 1 | 2 | 2 |
| 10 | y7 | 1 | 2 | 3 |
| 11 | y8 | 1 | 2 | 4 |
| 12 | y9 | 2 | 1 | 1 |
| 13 | y10 | 2 | 1 | 2 |
| 14 | y11 | 2 | 1 | 3 |
| 15 | y12 | 2 | 1 | 4 |
| 16 | y13 | 2 | 2 | 1 |

| Y | Dyad | Person | Time | | y1 | y2 | y3 | y4 | y5 | y6 | y7 | y8 | y9 | y10 | y11 | y12 | y13 |
|-----|------|--------|------|-----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|
| y1 | 1 | 1 | 1 | y1 | 1 | | | | | | | | | | | | |
| y2 | 1 | 1 | 2 | y2 | r | 1 | | | | | | | | | | | |
| y3 | 1 | 1 | 3 | y3 | r | r | 1 | | | | | | | | | | |
| y4 | 1 | 1 | 4 | y4 | r | r | r | 1 | | | | | | | | | |
| y5 | 1 | 2 | 1 | y5 | a | | | | 1 | | | | | | | | |
| y6 | 1 | 2 | 2 | y6 | | a | | | r | 1 | | | | | | | |
| y7 | 1 | 2 | 3 | y7 | | | a | | r | r | 1 | | | | | | |
| y8 | 1 | 2 | 4 | y8 | | | | a | r | r | r | 1 | | | | | |
| y9 | 2 | 1 | 1 | y9 | | | | | | | | | 1 | | | | |
| y10 | 2 | 1 | 2 | y10 | | | | | | | | | r | 1 | | | |
| y11 | 2 | 1 | 3 | y11 | | | | | | | | | r | r | 1 | | |
| y12 | 2 | 1 | 4 | y12 | | | | | | | | | r | r | r | 1 | |
| y13 | 2 | 2 | 1 | y13 | | | | | | | | | a | | | | 1 |

| Y | Dyad | Person | Time | | y1 | y2 | y3 | y4 | y5 | y6 | y7 | y8 | y9 | y10 | y11 | y12 | y13 |
|-----|------|--------|------|--|-----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|
| y1 | 1 | 1 | 1 | | y1 | 1 | | | | | | | | | | | |
| y2 | 1 | 1 | 2 | | y2 | r | 1 | | | | | | | | | | |
| y3 | 1 | 1 | 3 | | y3 | r | r | 1 | | | | | | | | | |
| y4 | 1 | 1 | 4 | | y4 | r | r | r | 1 | | | | | | | | |
| y5 | 1 | 2 | 1 | | y5 | a | | | | 1 | | | | | | | |
| y6 | 1 | 2 | 2 | | y6 | b | a | | | r | 1 | | | | | | |
| y7 | 1 | 2 | 3 | | y7 | | b | a | | r | r | 1 | | | | | |
| y8 | 1 | 2 | 4 | | y8 | | | b | a | r | r | r | 1 | | | | |
| y9 | 2 | 1 | 1 | | y9 | | | | | | | | | 1 | | | |
| y10 | 2 | 1 | 2 | | y10 | | | | | | | | r | 1 | | | |
| y11 | 2 | 1 | 3 | | y11 | | | | | | | | r | r | 1 | | |
| y12 | 2 | 1 | 4 | | y12 | | | | | | | | r | r | r | 1 | |
| y13 | 2 | 2 | 1 | | y13 | | | | | | | | a | | | | 1 |

Intraclass Correlation

The intraclass correlation will have the leading role in this play.

We'll denote it as

$$r_{xx'}$$

Intraclass Correlation & Waldo

$$r_{xx'}$$



Intraclass Correlation

ANOVA/HLM Language: Two level model approach

$$Y_{ij} = \beta_i + \epsilon_{ij}$$
$$\beta_i = \mu + \pi_i$$

Intraclass Correlation

ANOVA/HLM Language: Two level model approach

$$\begin{aligned}Y_{ij} &= \beta_i + \epsilon_{ij} \\ \beta_i &= \mu + \pi_i\end{aligned}$$

Intraclass correlation is given by

$$r_{xx'} = \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma_{\epsilon}^2}$$

Intraclass Correlation

ANOVA/HLM Language: Two level model approach

$$\begin{aligned}Y_{ij} &= \beta_i + \epsilon_{ij} \\ \beta_i &= \mu + \pi_i\end{aligned}$$

Intraclass correlation is given by

$$r_{xx'} = \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma_{\epsilon}^2}$$

A proportion interpretation.

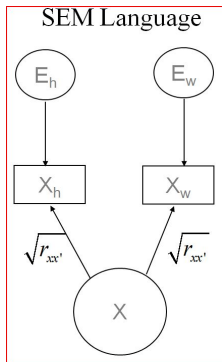
Intraclass Correlation: Another approach

ANOVA/HLM Language: Two level model

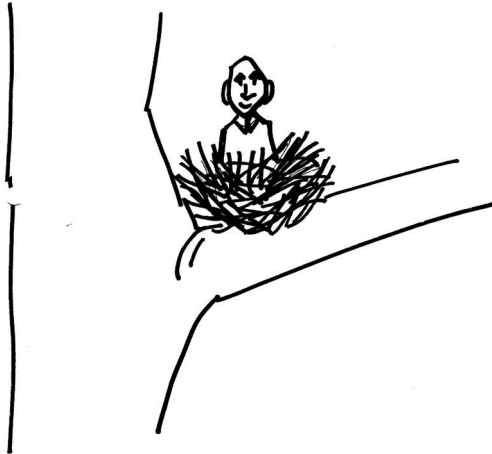
$$Y_{ij} = \beta_i + \epsilon_{ij}$$
$$\beta_i = \mu + \pi_i$$

Intraclass correlation is given by

$$r_{xx'} = \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma_{\epsilon}^2}$$



The Nested Individual



Symbolic representation for the pairwise setup

The first subscript represents the dyad and the second subscript represents the individual.

| Dyad # | Variable | |
|--------|----------|----------|
| | X | X' |
| 1 | X_{11} | X_{12} |
| | X_{12} | X_{11} |
| 2 | X_{21} | X_{22} |
| | X_{22} | X_{21} |
| 3 | X_{31} | X_{32} |
| | X_{32} | X_{31} |
| 4 | X_{41} | X_{42} |
| | X_{42} | X_{41} |

Concrete Illustration of the Pairwise Coding

| Dyad # | X | X' |
|--------|-------|-------|
| 1 | Amos | Bram |
| | Bram | Amos |
| 2 | Carl | Dan |
| | Dan | Carl |
| 3 | Ed | Frank |
| | Frank | Ed |
| | ETC | |

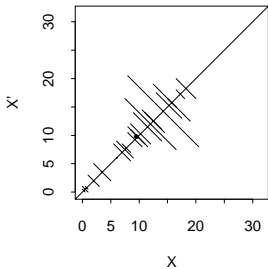
Concrete Illustration of the Pairwise Coding

| Dyad # | X | X' |
|--------|-------|-------|
| 1 | Amos | Bram |
| | Bram | Amos |
| 2 | Carl | Dan |
| | Dan | Carl |
| 3 | Ed | Frank |
| | Frank | Ed |

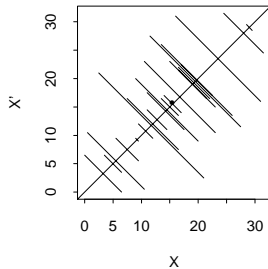
ETC

Pairwise Plots

strangers; $r_{xx'} = 0.72$



friends; $r_{xx'} = 0.4$



Pairwise Intraclass for the Exchangeable Case

ML intraclass correlation = Pearson correlation between variable X and X'

The significance test (against a null hypothesis of zero) is simply

$$Z = r_{xx'}\sqrt{n}$$

where Z is asymptotically normally distributed and n is the number of dyads.

Symbolic representation for the pairwise setup

The first subscript represents the dyad and the second subscript represents the individual. Categorization of individuals as 1 or 2 is based on the class variable C .

| Dyad # | C | Variable | |
|--------|-----|----------|----------|
| | | X | X' |
| 1 | 1 | X_{11} | X_{12} |
| | 2 | X_{12} | X_{11} |
| 2 | 1 | X_{21} | X_{22} |
| | 2 | X_{22} | X_{21} |
| 3 | 1 | X_{31} | X_{32} |
| | 2 | X_{32} | X_{31} |
| 4 | 1 | X_{41} | X_{42} |
| | 2 | X_{42} | X_{41} |

Pairwise Intraclass for the Distinguishable Case

Compute the partial correlation between variable X and the “reverse coded” version of X, partialling out the person code C.

The partial pairwise intraclass correlation is given by

$$r_{XX'.C} = \frac{r_{XX'} - r_{CX}r_{CX'}}{\sqrt{(1 - r_{CX}^2)(1 - r_{CX'}^2)}}$$

Intraclass Correlation

The structural model for the *exchangeable* case is

$$Y_{ij} = \mu + \pi_i + \epsilon_{ij}$$

where π is a random effect. The parameter π represents the “dyad effect.” Equivalent to a one-way random-effects ANOVA with “dyad” as the factor.

The structural model for the *distinguishable* case is

$$Y_{ijk} = \mu + \pi_i + \alpha_j + \epsilon_{ijk}$$

where π is a random effect and α is a fixed effect. The parameter π represents the “dyad effect” and the parameter α represents the effect on the “distinguishable” variable. Equivalent to a two-way ANOVA with “dyad” as a random-effects factor.

The standard definition of the intraclass correlation is

$$\rho_I = \frac{MSB - MSE}{MSB + (k - 1)MSE}$$

The terms MSB and MSE are from the ANOVA source table, and k represents the number of people in the “group” (i.e., in dyads $k = 2$).

The same formula is used whether a one-way ANOVA (exchangeable case) or a two-way ANOVA (distinguishable case) is used.

The intraclass correlation compares the variability between dyads v. the variability within dyads.

But the ANOVA approach is difficult to work with...

1. tedious to generalize to many variables
2. not easy to develop intuition for the relevant mean square terms and to connect the parameters to meaningful psychological statements.
3. not easy to develop tests of significance

The ANOVA approach can be generalized through “hierarchical linear models” (HLM).

But the ANOVA approach is difficult to work with...

1. tedious to generalize to many variables
2. not easy to develop intuition for the relevant mean square terms and to connect the parameters to meaningful psychological statements.
3. not easy to develop tests of significance

The ANOVA approach can be generalized through “hierarchical linear models” (HLM).

The **pairwise approach** is a special case of HLM when all groups have the same size (as in dyads), i.e., in the case of dyads the pairwise approach is identical to HLM.

In the special case of dyads the pairwise intraclass correlation is

$$\rho_p = \frac{SSB - SSE}{SSB + SSE}$$

In the special case of dyads the pairwise intraclass correlation is

$$\rho_p = \frac{SSB - SSE}{SSB + SSE}$$

Pairwise ICC = ML

ANOVA ICC = REML

CODE

SPSS

```
MIXED dv BY person  
  /fixed person  
  /print solution testcov  
  /repeated = person | SUBJECT(dyad) covtype(CS).
```

Covariance Parameters

Estimates of Covariance Parameters^a

| Parameter | Estimate | Std. Error | Wald Z | Sig. | 95% Confidence Interval | |
|--------------------------------------|---------------|------------|------------|-------|-------------------------|-------------|
| | | | | | Lower Bound | Upper Bound |
| Repeated Measures CS diagonal offset | 14.291667 | 5.834548 | 2.449 | .014 | 6.420684 | 31.811522 |
| | CS covariance | 401.511364 | 174.276433 | 2.304 | .021 | 59.935832 |

a. Dependent Variable: score.

```
/METHOD = ML OR REML
```

Dyadic Correlation Between Two Variables

Example: Trust and Satisfaction from each couple member

- how much do you trust your partner
- how satisfied are you with your marriage

What is the relation between trust and satisfaction?

How would you approach this analysis problem?

1. correlate the trust scores with the satisfaction scores ignoring group membership
2. correlate mean trust score (within couple) with mean satisfaction score

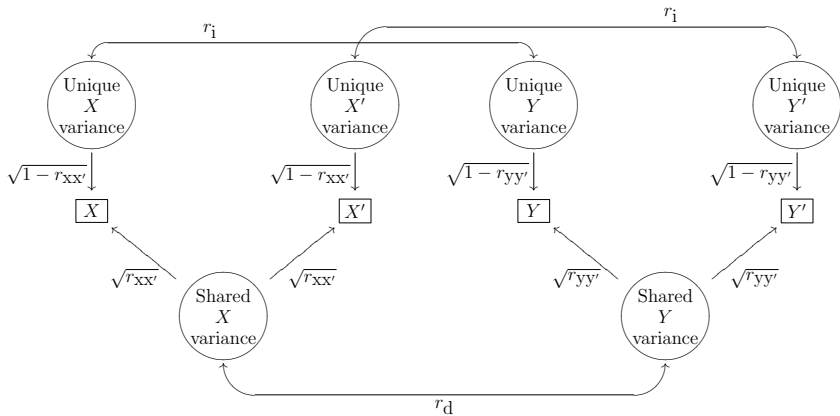
How would you approach this analysis problem?

1. correlate the trust scores with the satisfaction scores ignoring group membership
2. correlate mean trust score (within couple) with mean satisfaction score

There are problems with these two correlations!

The first confounds dyad-level effects and the second confounds individual-level effects.

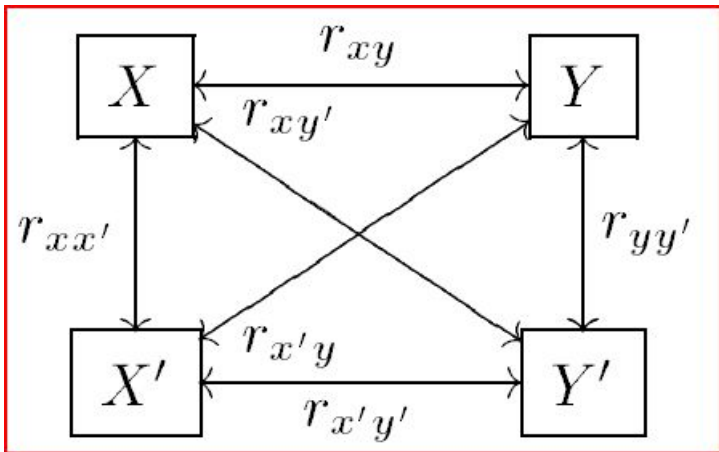
Thus, these two correlations are indeterminate as to the “psychological” mechanisms they represent.

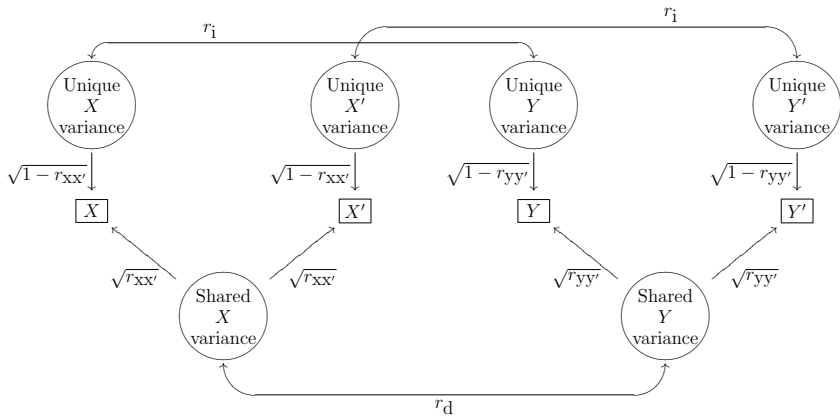


Symbolic representation for the pairwise setup for two variables

| Dyad # | C | Variable | | | |
|--------|-----|----------|----------|----------|----------|
| | | X | X' | Y | Y' |
| 1 | 1 | X_{11} | X_{12} | Y_{11} | Y_{12} |
| | 2 | X_{12} | X_{11} | Y_{12} | Y_{11} |
| 2 | 1 | X_{21} | X_{22} | Y_{21} | Y_{22} |
| | 2 | X_{22} | X_{21} | Y_{22} | Y_{21} |
| 3 | 1 | X_{31} | X_{32} | Y_{31} | Y_{32} |
| | 2 | X_{32} | X_{31} | Y_{32} | Y_{31} |
| 4 | 1 | X_{41} | X_{42} | Y_{41} | Y_{42} |
| | 2 | X_{42} | X_{41} | Y_{42} | Y_{41} |

Graphical Representation of the Correlations





| Dyad | Person | Var | | y1 | y2 | y3 | y4 | y5 | y6 | y7 | y8 | y9 | y10 | y11 | y12 |
|------|--------|-----|-----|----|----|----|----|----|----|----|----|----|-----|-----|-----|
| 1 | 1 | 1 | y1 | 1 | | | | | | | | | | | |
| 1 | 2 | 1 | y2 | r1 | 1 | | | | | | | | | | |
| 1 | 1 | 2 | y3 | ri | rd | 1 | | | | | | | | | |
| 1 | 2 | 2 | y4 | rd | ri | r2 | 1 | | | | | | | | |
| 2 | 1 | 1 | y5 | | | | | 1 | | | | | | | |
| 2 | 1 | 1 | y6 | | | | | r1 | 1 | | | | | | |
| 2 | 1 | 2 | y7 | | | | | ri | rd | 1 | | | | | |
| 2 | 2 | 2 | y8 | | | | | rd | ri | r2 | 1 | | | | |
| 3 | 1 | 1 | y9 | | | | | | | | | | 1 | | |
| 3 | 2 | 1 | y10 | | | | | | | | | r1 | 1 | | |
| 3 | 1 | 2 | y11 | | | | | | | | | ri | rd | 1 | |
| 3 | 2 | 2 | y12 | | | | | | | | | rd | ri | r2 | |

According to the model, the two **observed** correlations decompose as

$$r_{xy} = \sqrt{r_{xx'}} r_d \sqrt{r_{yy'}} + \sqrt{1 - r_{xx'}} r_i \sqrt{1 - r_{yy'}}$$

and

$$r_{xy'} = \sqrt{r_{xx'}} r_d \sqrt{r_{yy'}}.$$

According to the model, the two **observed** correlations decompose as

$$r_{xy} = \sqrt{r_{xx'}} r_d \sqrt{r_{yy'}} + \sqrt{1 - r_{xx'}} r_i \sqrt{1 - r_{yy'}}$$

and

$$r_{xy'} = \sqrt{r_{xx'}} r_d \sqrt{r_{yy'}}.$$

With these decompositions, simple algebra solves for r_i and r_d .

$$r_i = \frac{r_{xy} - r_{xy'}}{\sqrt{1 - r_{xx'}} \sqrt{1 - r_{yy'}}}$$

and

$$r_d = \frac{r_{xy'}}{\sqrt{r_{xx'}} \sqrt{r_{yy'}}}.$$

Example: Individual and Dyad Level Relationship

Frequency of verbalization and frequency of gaze.

| | r_i | r_d |
|-----------|-------|-------|
| Strangers | -.33 | .68 |
| Friends | .14 | .30 |

CODE

SAS

```
proc calis cov edf=N-1 se method=mls residual pcorr;  
lineqs  
v1 = 1 F1 + E1,  
v2 = 1 F1 + E2,  
v3 = 1 F2 + E3,  
v4 = 1 F2 + E4;  
  
STD  
F1-F2 = v1 v2,  
E1-E4 = x1 x1 x2 x2;  
  
COV  
F1 F2 = rd,  
E1 E3 = ri,  
E2 E4 = ri;  
run;
```

Cov matrix as input; state N.

CODE

Mplus

```
Title: SEM model;
Data: File = G:\FTS\files from Rich\SEM data L3.dat;
variable: names = ID x1 x2 y1 y2;
USEV = x1 x2 y1 y2;
Analysis: type = meanstructure;
model:
    x by x1@1 x2@1;
    y by y1@1 y2@1;
    x1 x2 (1);
    y1 y2 (2);
    [x1 x2] (4);
    [y1 y2] (5);
    x1 with y1 (3);
    x2 with y2 (3);
    x with y;

output: sampstat standardized;
```


Correlation Between Dyad Means

$$r_m = \frac{r_{xy} + r_{xy'}}{\sqrt{1 + r_{xx'}} \sqrt{1 + r_{yy'}}$$

r_m can be positive under different combinations of r_{xy} and $r_{xy'}$.

r_m reflects a combination of individual and dyad level processes, and should not be routinely interpreted as reflecting only dyad level processes.

Latent Variable Model: HLM Lingo

Three-level model: one level for the variable, one level for individual effect, and one level for group effect.

$$\begin{aligned}Y_{ijk} &= \beta_0 X_0 + \beta_1 X_1 \\ \beta_0 &= \mu_0 + \pi_0 + \epsilon_0 \\ \beta_1 &= \mu_1 + \pi_1 + \epsilon_1\end{aligned}$$

$$\pi \sim N\left(0, \begin{bmatrix} V_{\pi_0} & C_{\pi_0\pi_1} \\ C_{\pi_0\pi_1} & V_{\pi_1} \end{bmatrix}\right) \quad \epsilon \sim N\left(0, \begin{bmatrix} V_{\epsilon_0} & C_{\epsilon_0\epsilon_1} \\ C_{\epsilon_0\epsilon_1} & V_{\epsilon_1} \end{bmatrix}\right)$$

Alternative Model: Interdependence

The degree to which one individual influences another (e.g., Lewin).

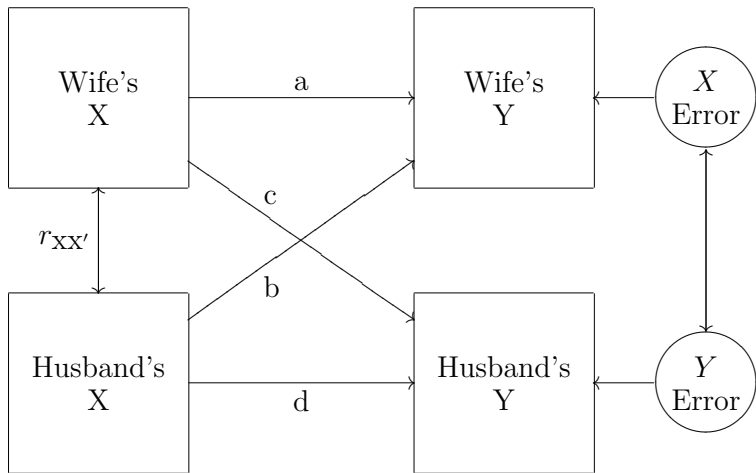
This influence need to occur face-to-face:

We have a good time together, even when we're not together. Yogi Berra

Kelley & Thibaut: Early APIM

Interaction separated into three types of control or influence

1. actor effect (reflexive)
2. partner effect (fate)
3. mutual effect (behavior)



APIM is a Pairwise Model

$$Y = \beta_0 + \beta_1 X + \beta_2 X' + \beta_3 X X'$$

such that

- predictor X represents the actor's influence on the actor's Y ,
- predictor X' represents the partner's influence on the actor's Y ,
- the product XX' represents the mutual influence of both people on the actor's Y .

Example: Generalized Pairwise Model

Regress frequency of smiles/laughter on frequency of verbalization

- Strangers: partner's verbalization frequency on actor's laughter—the more the other talks, the more the actor smiles

no other effects

Example: Generalized Pairwise Model

Regress frequency of smiles/laughter on frequency of verbalization

- Strangers: partner's verbalization frequency on actor's laughter—the more the other talks, the more the actor smiles

no other effects

- Friends: actor's verbalization frequency on the actor's laughter—the more I talk, the more I smile)

no other effects

Actor-partner model: The ICC again

$$Y = \beta_0 + \beta_1 X + \beta_2 X'$$

such that

- actor's X predicts actor's Y and
- partners X (denoted X') predicts actor's Y

Actor-partner model: The ICC again

$$Y = \beta_0 + \beta_1 X + \beta_2 X'$$

such that

- actor's X predicts actor's Y and
- partners X (denoted X') predicts actor's Y

The actor regression coefficient can be expressed in terms of pairwise correlations

$$\beta_1 = \frac{s_y(r_{xy} - r_{xy'}r_{xx'})}{s_x(1 - r_{xx'}^2)}$$

Similarly, the partner regression coefficient is

$$\beta_2 = \frac{s_y(r_{xy'} - r_{xy}r_{xx'})}{s_x(1 - r_{xx'}^2)}$$

Variance of β related to the ICC

The variance of the actor β is

$$V(\beta_{\text{actor}}) = \frac{s_y^2(r_{xy'}^2 r_{xx'}^2 - r_{xx'} r_{yy'} + 1 - r_{xy'}^2)}{2N s_x^2 (1 - r_{xx'}^2)}$$

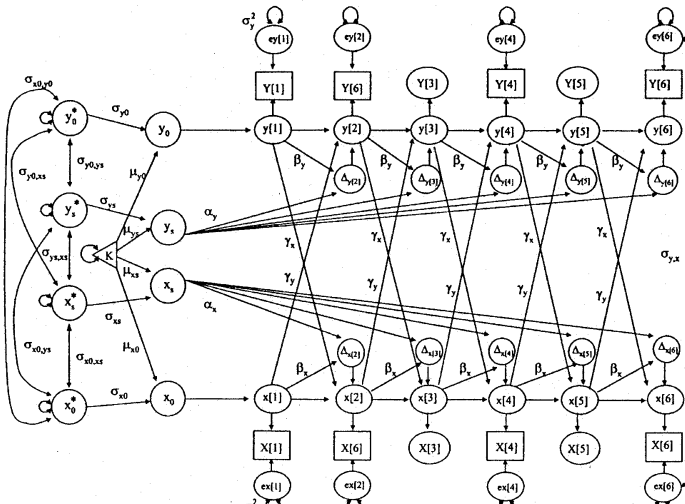
Longitudinal Models

Get complicated.

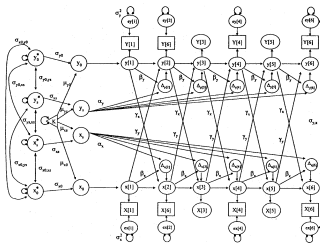
Different ways of representing change in a single person, now there are two individuals.

If data have multiple variables, then there are temporal, multivariate and interpersonal relations.

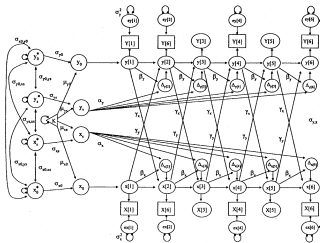
One Person: McArdle's Bivariate Latent Difference Model

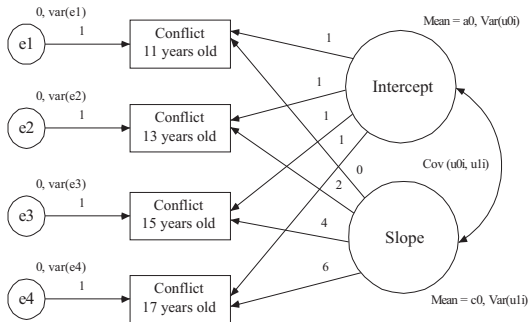


Person 1

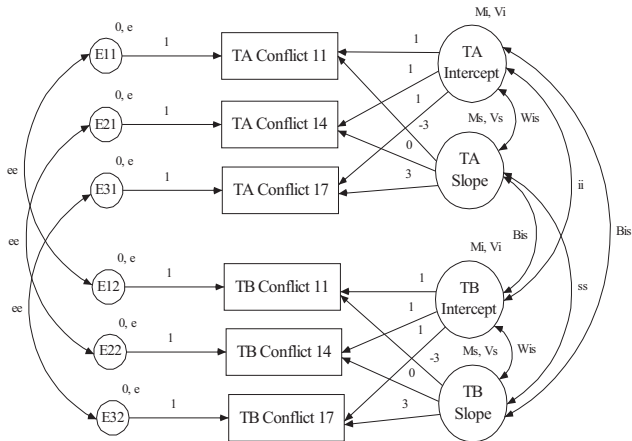


Person 2





Shared Variance Intercept/Slope



Which Programs to Use?

Multilevel models provide one unified approach to dyadic and longitudinal models.

Advantages: Arbitrary nested models with multiple levels of analysis

Disadvantages: Models can be complicated to implement and interpret, constraints can't easily be set

Which Programs to Use?

SEM provides another unified approach to dyadic and longitudinal models.

Advantages: Multiple variables easy to handle

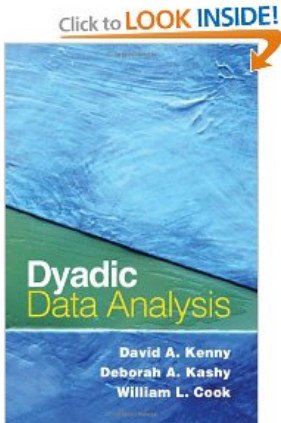
Disadvantages: Difficult to implement unequal size groups;
longitudinal designs can get complicated

Prescriptions

- Violation of independence is not a problem
- Follow your research question
- Not completely a matter of choosing a statistical technique; more important to think about design and research question

A book devoted to dyadic data analysis:

Kenny, D., Kashy, D., & Cook, W. (2006). *Dyadic Data Analysis*. Guilford Press.



Interdependence Mantra

Study

Model

Celebrate

Interdependence

Thanks

For more detail send me an email (gonzo@umich.edu) or check out the link

<http://www.umich.edu/~gonzo>

(in a few days).

