R WORKSHOP – DAY 3

Randi L. Garcia

Smith College

July 17th, 19th, and 21st

DAY 3

- Multilevel Modeling
- Growth Curve Modeling
- Dyadic Data Analysis

MULTILEVEL MODELING

Multilevel Modeling AKA…

- Hierarchical Linear Model (HLM)
  - Because of the popular program crated by Raudenbush, Bryk et al.
- Random coefficient models
- Mixed-effects models
- Mixed linear models
  - This is what it's called in statistics
- Multilevel regression models

Nested Data

- Whenever you have nested data you (probably) need MLM
  - Students in classrooms
  - Therapists' ratings of their classrooms
  - People in romantic couples (dyads)
- Participants within a cluster are more similar to each other than two participants in different clusters
- Statistical techniques you have learned so far require independent observations
- Observations are not independent when we have nested data
  - If we know something about Jim's wife's satisfaction, then we already know little something about Jim's satisfaction

Consequences of Ignoring Nested Data

- When each participant does not give independent pieces of info, the *effective sample size* is reduced
- If sample size is bigger than it should be, standard errors are too small, so t-statistics too big, and finally, *p*-values are too small!
- SE's and p-values are biased, test are too liberal
- Coefficients are unchanged

$$N \uparrow \qquad SE = \left(\frac{s}{\sqrt{N}}\right) \downarrow \qquad t = \frac{b}{SE} \uparrow \qquad p\ value \downarrow$$

Fixed vs. Random Effects

- MLM includes specifying fixed and random effects in one linear model (hence: "mixed")

- The levels of a fixed factor are assumed to be a complete picture of that variable
  - If we are interested in gender and we have men and women in our sample—there is no other level of gender we could have included
- The levels of a random factor are a random sample of possible levels
  - If we have different dosages—there are other possible dosages we could have selected

Fixed vs. Random Effects

- For fixed effects, we look for mean differences between the conditions or levels
- For random effects, we estimate the variance in means across all conditions because the actual mean difference from one condition to another isn't really important
- In ANOVA, testing these two effects is mathematically similar, the interpretation changes

Multiple Regression Rewind

- Linear equation:

$$y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$$

- We have a *y* for every person, *i*
- Also we have *X1* and *X2* for each person *i*
- Each y_i is explained by a
  - fixed piece $[b_0 + b_1 X_{1i} + b_2 X_{2i}]$ and a
  - random piece $[e_i]$
- We assume the e_i's are normal with mean 0 and variance σ_e^2

What are two-level equations?

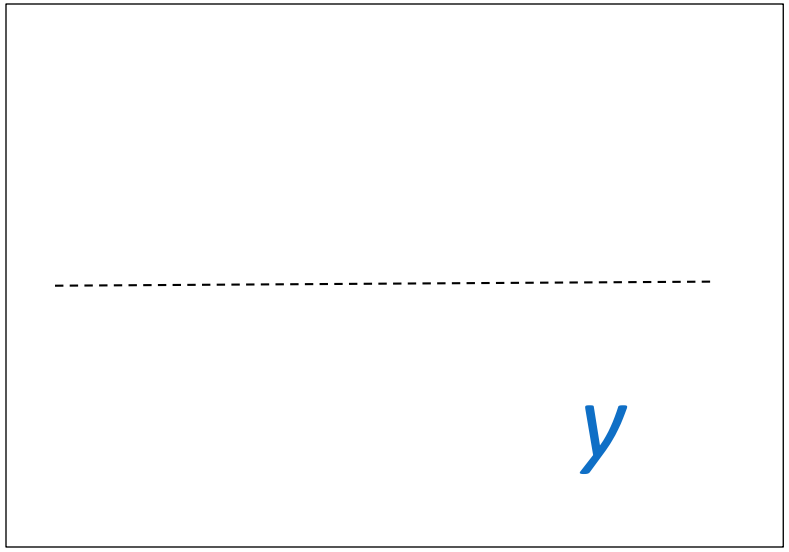
- In multilevel modeling we want to predict the outcome variable at the micro level, but also group differences in that outcome variable at the macro level.

- We need linear equations at *both* levels to capture these effects.

- Two-level versus combined equation for MLM
  - We can present our linear equations at both levels in **a set of separate equations**, OR
  - We can combine the set of equations into one simplified **combined equation**.

The Most Basic MLM

- **Example:** our outcome (*y*) is popularity of children in classes
- First, no predictors of *y*
- **Question:** Do some classrooms have pupils that are more popular than other classrooms?



The Most Basic MLM

Micro level $y_{ij} = b_{0j} + e_{ij}$

Macro level $b_{0j} = g_{00} + u_{0j}$

- $\boldsymbol{y_{ij}}$ is the score on *y* for person *i* in group *j*
- $\boldsymbol{g_{00}}$ is the grand or overall intercept
- $\boldsymbol{e_{ij}}$ is the residual for person *i* in group *j*
  - $\boldsymbol{e_{ij}}$**'s** are normal, mean of 0 and variance σ_w^2
- $\boldsymbol{u_{0j}}$ is the residual for group *j*
  - $\boldsymbol{u_{0j}}$**'s** are normal, with mean 0 and variance τ_{00}

Intraclass Correlation (ICC)

- The ICC is a measure of the proportion of variance in the outcome that is accounted for by group membership

$$ICC = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

- Group variance divided by total variance (group plus within variance)
- No predictors in the model, so σ_b^2 is just the group variance.

Intraclass Correlation (ICC)

- The ICC is a measure of the proportion of variance in the outcome that is accounted for by group membership

$$ICC = \frac{\boldsymbol{\tau_{00}}}{\boldsymbol{\tau_{00}} + \sigma_w^2}$$

- Group variance divided by total variance (group plus within variance)
- No predictors in the model, so $\boldsymbol{\tau_{00}}$ is just the group variance.

Adding a Level 1 Predictor Fixed

Micro level

$$y_{ij} = b_{0j} + \boldsymbol{b_{1j}X_{1ij}} + e_{ij}$$

Macro level

$$b_{0j} = g_{00} + u_{0j}$$

$$\boldsymbol{b_{1j} = g_{10}}$$

- b_{1j} is the effect of X1
- We still have the group variance τ_{00}, which is the variance of the u_{0j}'s
- Combined equation:

$$y_{ij} = g_{00} + g_{10}X_{1ij} + u_{0j} + e_{ij}$$

Adding a Level 1 Predictor w/ Random Component

Micro level

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

Macro level

$$b_{0j} = g_{00} + u_{0j}$$

$$b_{1j} = g_{10} + \boldsymbol{u_{1j}}$$

- g_{10} is the overall (grand) effect of X1, but now this effect is allowed to vary across groups
- u_{1j} is the residual for each group's slope, they are normal with mean 0, variance τ_{11}
- It helps to write out the full equation

Adding a Level 1 Predictor w/ Random Component

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

$$b_{0j} = g_{00} + u_{0j}$$

$$b_{1j} = g_{10} + u_{1j}$$

- **Combined equation**:

$$y_{ij} = g_{00} + g_{10}X_{1ij} + u_{1j}X_{1ij} + u_{0j} + e_{ij}$$

Covariance Matrix of Random Effects

- We just added the variance of the level 1 slope across groups to the model
  - That is, are the effects of extroversion on popularity different across classrooms?
  - In some classrooms there is a strong effect of extroversion, and in others it's relatively weak.

- Now we can also ask, in classrooms with strong effects of extroversion on popularity, is popularity also higher?
  - The covariance between the group intercept and group slope, called τ_{01}

Covariance Matrix of Random Effects

- R estimates a whole matrix of random effects
  1. The group variance in the intercept, τ_{00}
  2. The group variance in the slope of extroversion→popularity, τ_{11}
  3. AND, the covariance between these two, τ_{01}

- Covariance matrix of random effects:

$$\begin{bmatrix} \tau_{00} & \\ \tau_{01} & \tau_{11} \end{bmatrix}$$

R MARKDOWN FILE

DYADIC DATA ANALYSIS

Definitions: Distinguishability

- Can all dyad members be distinguished from one another based on a meaningful factor?
- Distinguishable dyads
  - Gender in heterosexual couples
  - Patient and caregiver
  - Race in mixed race dyads

All or Nothing

- If most dyad members can be distinguished by a variable (e.g., gender), but a few cannot, then can we say that the dyad members are distinguishable?
- No, we cannot!

Indistinguishability

- There is no systematic or meaningful way to order the two scores

- Examples of indistinguishable dyads
  - Same-sex couples
  - Twins
  - Same-gender friends
  - Mix of same-sex and heterosexual couples
  - When all dyads are hetero except for even one couple!

It can be complicated…

- Distinguishability is a mix of theoretical and empirical considerations.

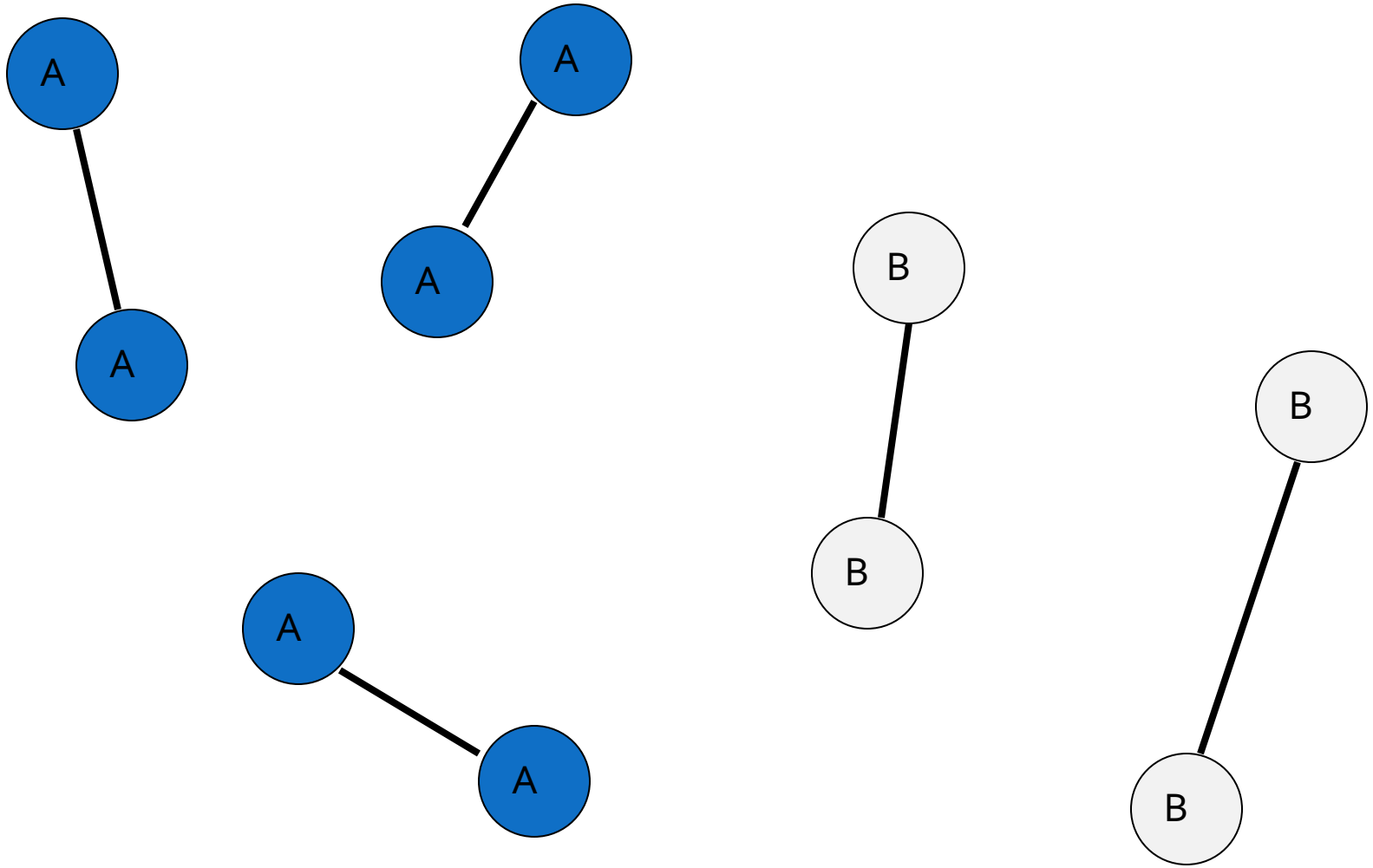
- For dyads to be considered distinguishable:
  1. It should be theoretically important to make such a distinction between members.
  2. Also it should be shown that empirically there are differences.

- Sometimes there can be two variables that can be used to distinguish dyad members: Spouse vs. patient; husband vs. wife.

Types of Variables

- Between Dyads
  - Variable varies from dyad to dyad, BUT within each dyad all individuals have the same score
    - Example: Length of relationship

- Called a level 2, or macro variable in multilevel modeling



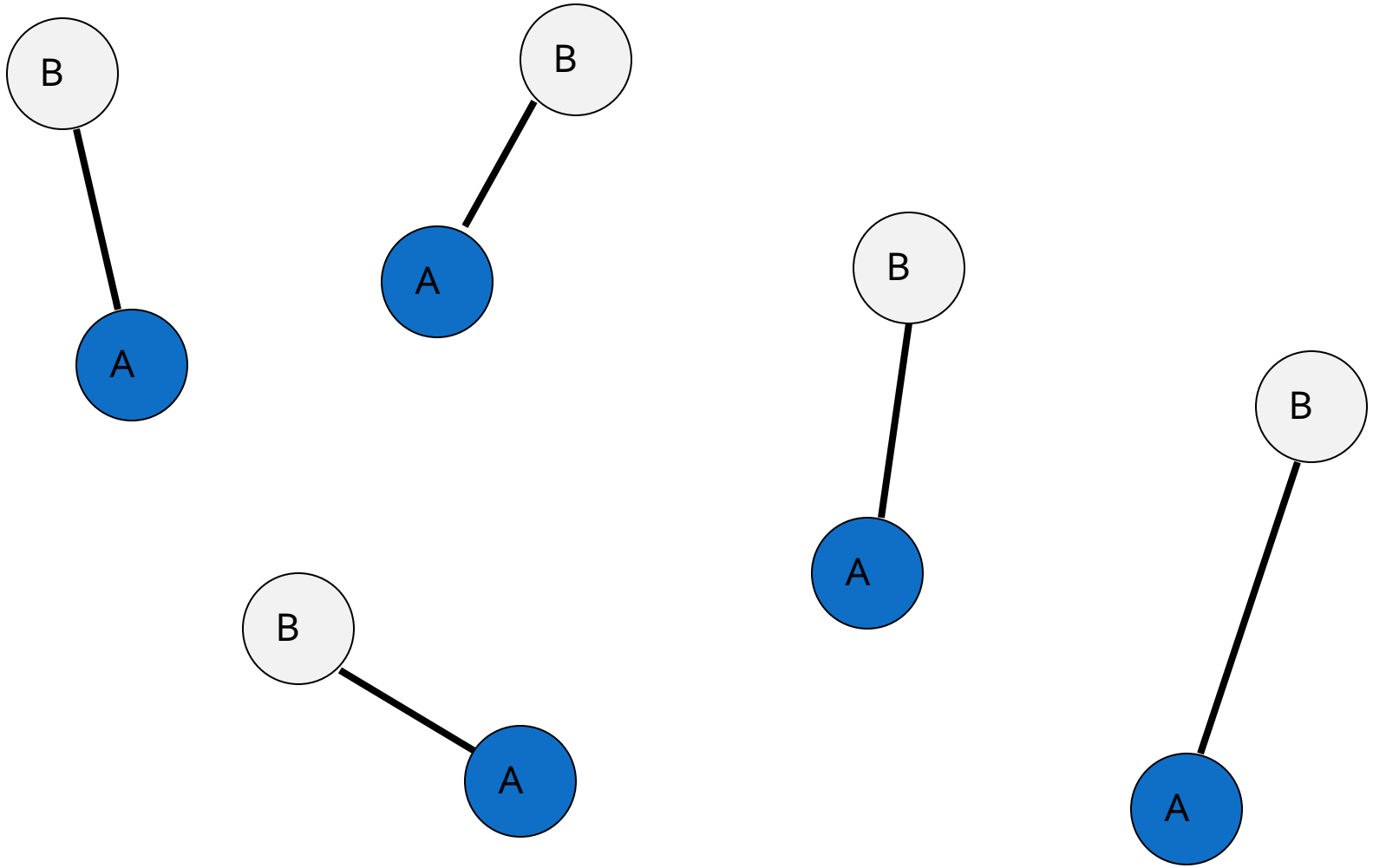
Between

Within Dyads

- Variable varies from person to person within a dyad, BUT there is no variation on the dyad average from dyad to dyad.
  - Percent time talking in a dyad
  - Reward allocation if each dyad is assigned the same total amount

- $X1 + X2$ equals the same value for each dyad

- Note: If in the data, there is a dichotomous within-dyads variable, then dyad members *can* be distinguished on that variable. But that doesn't mean it would be theoretically meaningful to do so.



Within

Mixed Variable

- Variable varies both between dyads and within dyads.

- In a given dyad, the two members may differ in their scores, and there is variation across dyads in the average score.
  - Age in married couples
  - Lots-o personality variables

- Most outcome variables are mixed variables.

It can be complicated…

Can you think of a variable that can be **between-dyads**, **within-dyads**, or **mixed** across different samples?

DATA STRUCTURES

Illustration of Data Structures: Individual

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

Illustration of Data Structures: Individual

AAAAA
AAAAA
AAAAA
AAAAA
BBBBB
BBBBB
BBBBB
BBBBB

Illustration of Data Structures: Dyad

Dyad	X_1	Y_1	Z_1	X_2	Y_2	Z_2[a]
1	5	9	3	2	8	3
2	6	3	7	4	6	7
3	3	6	5	9	7	5

Illustration of Data Structures: Dyad

AAAAABBBBB
AAAAABBBBB
AAAAABBBBB
AAAAABBBBB

Illustration of Data Structures: Pairwise

Dyad	*Person*	X_1	Y_1	Z_1	X_2	Y_2	Z_2[a]
1	1	5	9	3	2	8	3
1	2	2	8	3	5	9	3
2	1	6	3	7	4	6	7
2	2	4	6	7	6	3	7
3	1	3	6	5	9	7	5
3	2	9	7	5	3	6	5

[a]This variable is redundant with Z_1 and need not be included.

Illustration of Data Structures: Pairwise

AAAAABBBBB
AAAAABBBBB
AAAAABBBBB
AAAAABBBBB
BBBBBAAAAA
BBBBBAAAAA
BBBBBAAAAA
BBBBBAAAAA

R MARKDOWN FILE

NONINDEPENDENCE IN DYADS

Negative Nonindependence

- Nonindependence is often defined as the proportion of variance explained by the dyad (or group).
- BUT, nonindependence can be negative…variance cannot!

- This is super important
- **THE MOST IMPORTANT THING ABOUT DYADS!**

How Might Negative Correlations Arise?

Examples

- **Division of labor:** Dyad members assign one member to do one task and the other member to do another. For instance, the amount of housework done in the household may be negatively correlated.

- **Power:** If one member is dominant, the other member is submissive. For example, self-objectification is negatively correlated in dyadic interactions.

Effect of Nonindependence

- Consequences of ignoring clustering classic MLM
  - Effect Estimates Unbiased
- For dyads especially
  - Standard Errors Biased
    - Sometimes too large
    - Sometimes too small
    - Sometimes hardly biased

Direction of Bias Depends on

1. Direction of Nonindependence
   - Positive
   - Negative
2. Is the predictor a between or within dyads variable? (or somewhere in between: mixed)

Effect of Ignoring Nonindependence on Significance Tests

	Positive	Negative
Between		
Within		

What Not To Do!

- Ignore it and treat individual as unit
- Discard the data from one dyad member and analyze only one members' data
- Collect data from only one dyad member to avoid the problem
- Treat the data as if they were from two samples (e.g., doing an analysis for husbands and a separate one for wives)
  - Presumes differences between genders (or whatever the distinguishing variable is)
  - Loss of power

What To Do

- Consider both individual and dyad in one analysis!
  1. Multilevel Modeling
  2. Structural Equation Modeling

Traditional Model: Random Intercepts

Micro level

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

Macro level

$$b_{0j} = g_{00} + g_{01}Z_{1j} + u_{0j}$$

$$b_{1j} = g_{10}$$

- i from 1 to 2, because there are only 2 people in each "group".
- X_{1ij} is a mixed or within variable, and Z_{1j} is a between variable.
- Note b_{0j} is the common intercept for dyad j which captures the nonindependence.
- Works well with positive nonindependence, but not negative.

Alternative Model: Correlated Errors

Micro level

$$y_{1j} = b_0 + b_{1j}X_{11j} + e_{1j}$$

$$y_{2j} = b_0 + b_{1j}X_{12j} + e_{2j}$$

ρ called "*rho*"

Macro level

$$b_{1j} = g_{10}$$

- ρ is the correlation between e_{1j} and e_{2j}, the 2 members' residuals (errors).
- Note b_0 is now the grand intercept
- Works well with positive nonindependence AND negative.

ACTOR-PARTNER INTERDEPENDENCE MODEL (APIM)

Actor-Partner Interdependence Model (APIM)

- A model that simultaneously estimates the effect of a person's own variable (actor effect) and the effect of same variable but from the partner (partner effect) on an outcome variable
- The actor and partner variables are the same variable from different persons.
- All individuals are treated as actors and partners.

Data Requirements

- Two variables, X and Y, and X causes or predicts Y
- Both X and Y are mixed variables—both members of the dyad have scores on X and Y.

- Example
  - Dyads, one a patient with a serious disease and other being the patient's spouse. We are interested in the effects of depression on relationship quality

Actor Effect

- Definition: The effect of a person's X variable on that person's Y variable
  - the effect of patients' depression on patients' quality of life
  - the effect of spouses' depression on spouses' quality of life

- Both members of the dyad have an actor effect.

Partner Effect

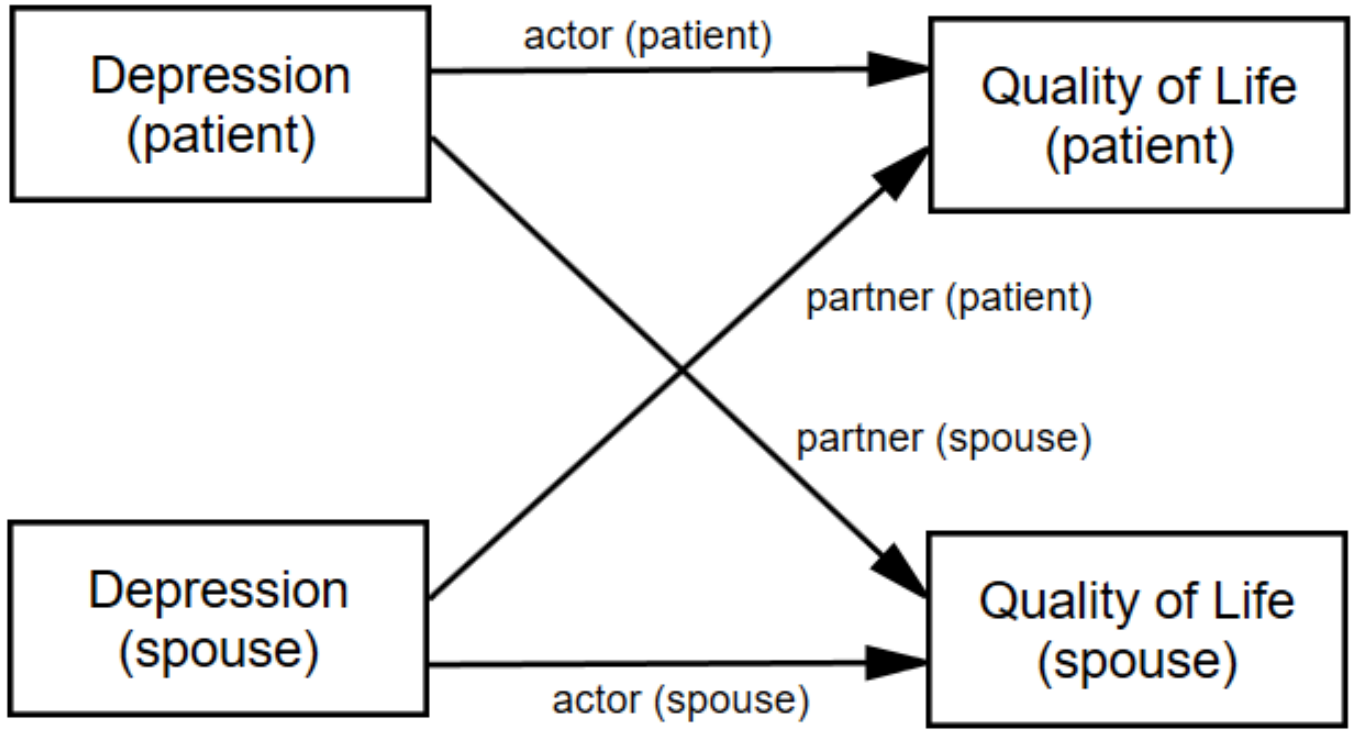
- Definition: The effect of a person's partner's X variable on the person's Y variable
  - the effect of patients' depression on spouses' quality of life
  - the effect of spouses' depression on patients' quality of life

- Both members of the dyad have a partner effect.

Distinguishability and the APIM

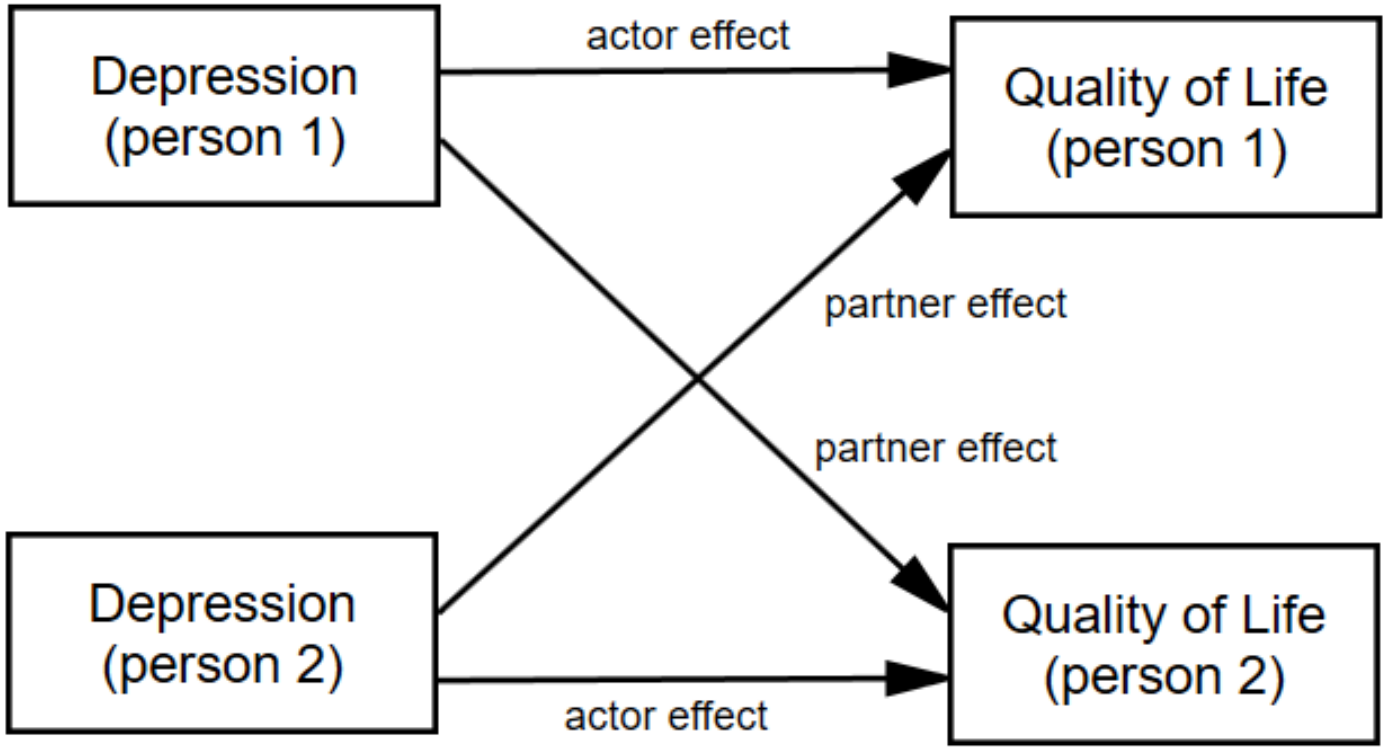
- Distinguishable dyads
  - Two actor effects
    - An actor effect for patients and an actor effect for spouses
  - Two partner effects
    - A partner effect from spouses to patients and a partner effect from patients to spouses

Distinguishable Dyads



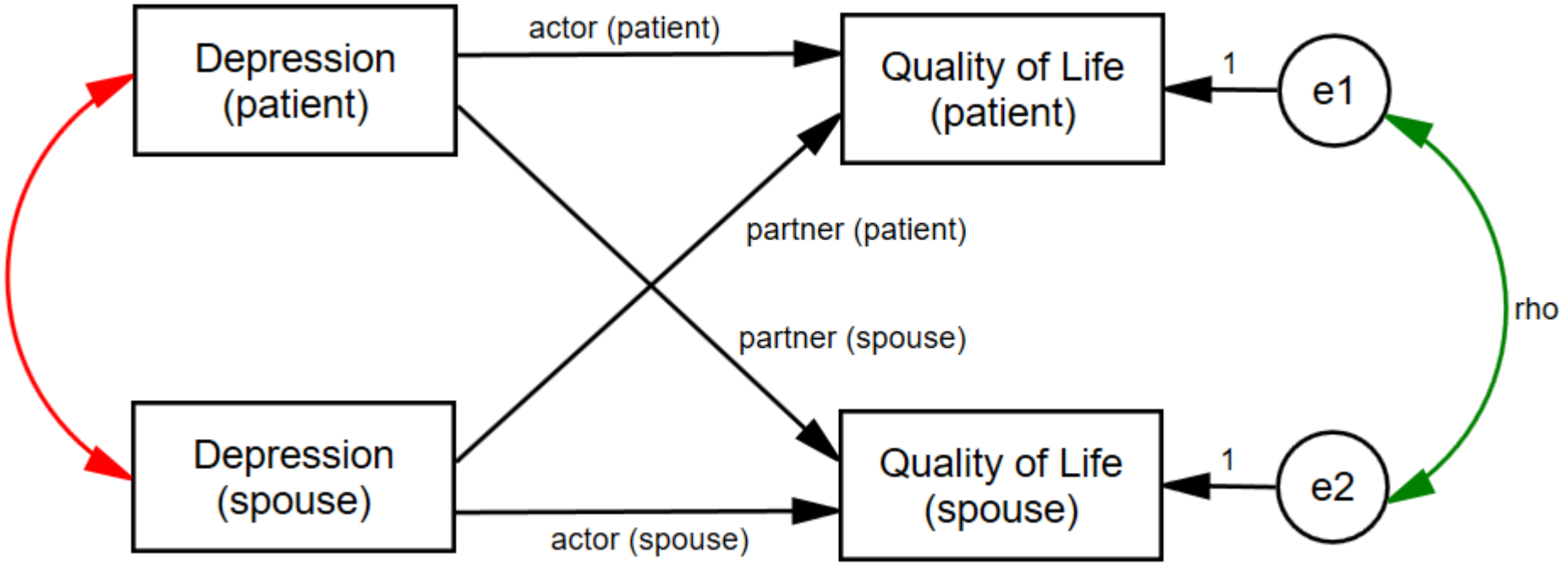
- Errors not pictured (but important)
- *The partner effect is fundamentally dyadic.* A common convention is to refer to it by the outcome variable. Researcher should be clear!

Indistinguishable Dyads



- The two actor effects are set to be equal and the two partner effects are set to be equal.

Nonindependence in the APIM



- Green curved line: Nonindependence in Y
- Red curved line: X as a mixed variable (r cannot be 1 or -1)
- Note that the combination of actor and partner effects explain some of the nonindependence in the dyad.

R MARKDOWN FILE

GENERALIZED LINEAR MIXED MODELS

Generalized Linear Models

- In general we wrap the response variables in a link function (log, logit, probit, identity, etc.).
- For example
  - A logistic regression is a generalized linear model making use of a logit link function.
  - A log-linear of Poisson regression is a generalized linear model making use of a log link function.
  - A regression model is a generalized linear model making use of an “identity” link function—the response is multiplied by 1.

Logistic Regression Review

- DV is dichotomous
  - probability of belonging to group 1: P_1
  - probability of belonging to group 0: $P_0 = 1 - P_1$.
  - There are only two choices!

Odds and Odds Ratios

		committed Committed to Hospital		
		0 No	1 Yes	Total
minority Minority Classification	0 No	138	120	258
	1 Yes	54	42	96
Total		192	162	354

- **Probability** of being committed = $\frac{162}{354} = .458$
- **Odds** of being committed = $\frac{.458}{1-.458} = .845$
  - Odds of being committed for minorities = $\frac{.438}{1-.438} = .778$
  - Odds of being committed for non-minorities = $\frac{.465}{1-.465} = .870$
- Odds ratio for non-minorities vs. minorities = $\frac{.870}{.778} = 1.118$

"Non-minorities are **1.118** times more likely to be committed than minorities."

Logistic Regression Equation

$$\ln\left(\frac{\widehat{P_1}}{1-\widehat{P_1}}\right) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n$$

- Where $\widehat{P_1}$ is the predicted probability of being in group coded as 1
- $\frac{\widehat{P_1}}{1-\widehat{P_1}}$ is the odds of being in group 1
- $\ln\left(\frac{\widehat{P_1}}{1-\widehat{P_1}}\right)$ is the "logit" function

Logistic Regression Equation

$$\ln\left(\frac{\widehat{P_1}}{1-\widehat{P_1}}\right) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n$$

- The b's are interpreted as the increase in log-odds of being in the target group for 1-unit increase in X.
- Exp(b) is the increase in odds for 1 unit increase in X—this works out to the odds ratio between X = a and X = a+1.

Log-Linear (Poisson) Regression Equation

- Used when the response variable is a count (e.g., number of cigarettes smoked per day).

$$\ln(Y) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n$$

- Where $\boldsymbol{Y}$ is the response vairable
- $\boldsymbol{ln(Y)}$ is the "log" link function
- $\boldsymbol{b_1}$ is interpreted as the increase in log-Y for every increase in $\boldsymbol{X_1}$
- **Exp($\boldsymbol{b_1}$)** is interpreted in the usual way—as in the general linear model.

Generalized Mixed Linear Models

- Generalized linear models
  - In general we wrap the response in a link function (log, logit, probit, identity, etc.).

- Generalized Mixed Linear Models
  - Do the same, include a link function that is appropriate for your response, but then include random effects in the model.
  - "Mixed" refers to the mixture of fixed and random effects in the model.

- We'll fit these models with the `lme4` package in R, specifically, the `glmer()` function.

Generalized Estimating Equations (GEE)

- Nonindependence treated as a "nuisance" to be removed; no statistical tests of nonindependence
- Can be extended to:
  - Binomial outcome
  - Multinomial outcome (Categories: home/work/leisure)
  - Count data (Poisson, negative binomial)
  - Can also be used for continuous outcomes (normal distribution)

- Fit these models with the `gee` package in R, specifically, the `gee()` function.

R MARKDOWN FILE

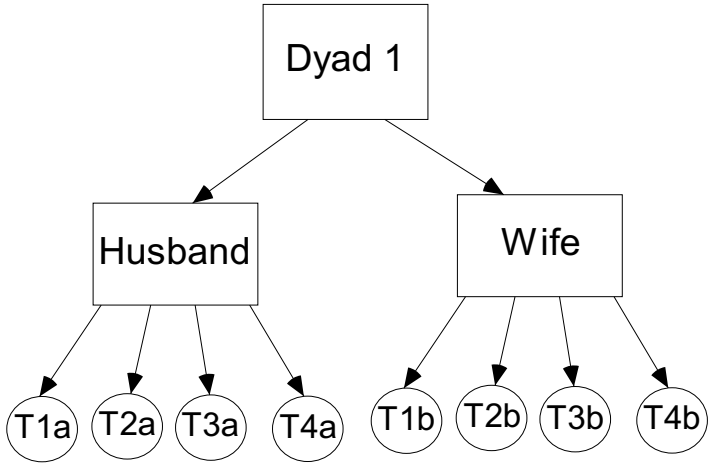
GROWTH CURVE MODELING

Examples of Over-Time Dyadic Data

- Daily diary reports of relationship experiences from both members of heterosexual dating partners over 14 days
- Repeated measures experiment where dyads interact with each other multiple times and make ratings after each interaction
- Daily reports of closeness from both members of college roommate dyads

Basic Data Structure

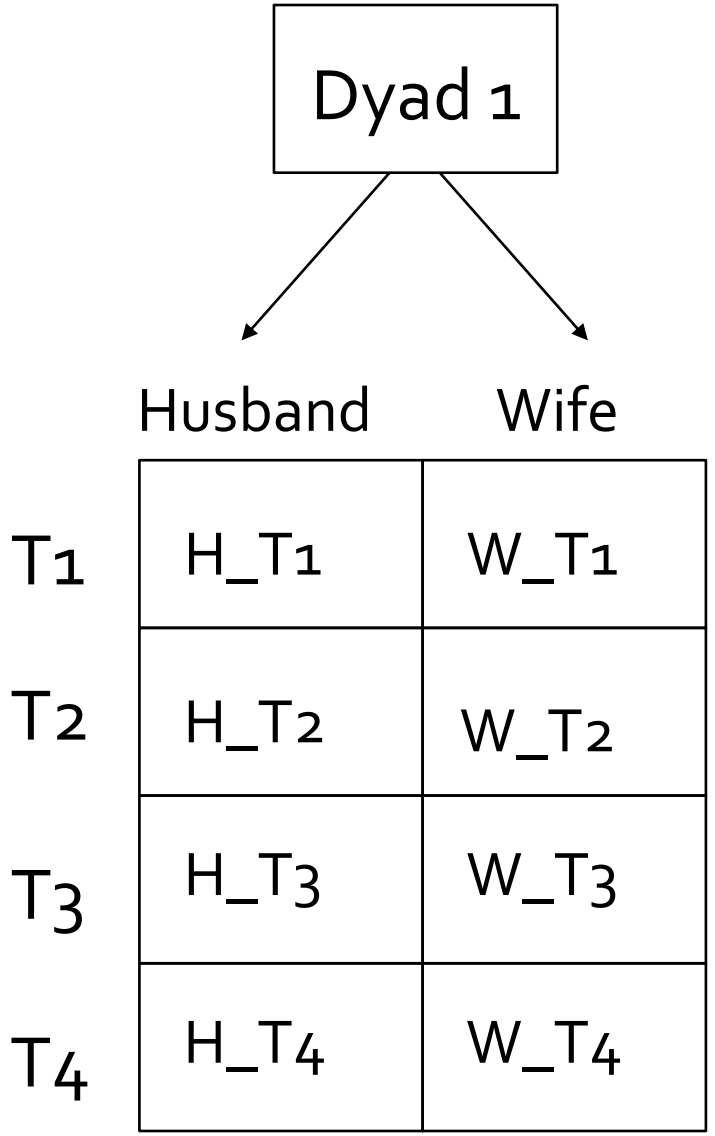
- The three-level nested myth: Time is nested within person and person is nested within dyad



- Three-level nested only if the four time points differ such that T1a ≠ T1b, T2a ≠ T2b, etc.

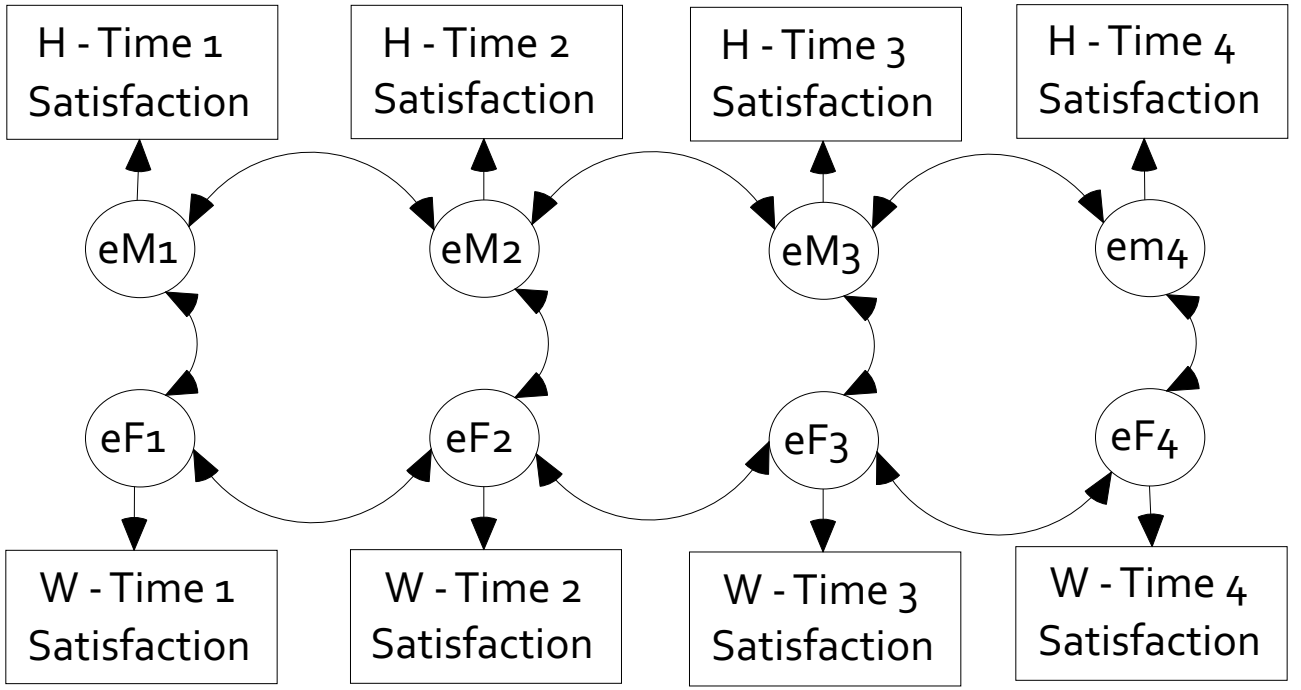
Basic Data Structure

- In most cases the two dyad members are measured at the same time points, so Time is *crossed* with person.



Basic Data Structure

- This two-level crossed structure results in an error structure in which the residuals may be correlated both
  - A) across dyad members
  - B) across time



Types of Over-Time Models

- Repeated Measures Model
  - Interest only in the effects of "time" across persons and dyads.
- Growth Curve Model
  - Are there linear changes over time in the outcome variable?
- Stability and Influence Model
  - Stability: Does Person A's score at time 1 predict Person A's score at time 2?
  - Influence: Does Person A's score at time 1 predict Person B's score at time 2?
- Standard APIM
  - Different variables as the predictors and at the outcome
  - Does Variable 1 predict Variable 2?

Types of Variables

- Time Invariant
  - Do not change over time
  - Measured at one time point only (typically the beginning of the study)
  - E.g., gender, attachment style, race
- Time Varying
  - Measured at each time
  - E.g., daily mood, twice-weekly reports of friendship
  - Outcome variable must be time varying

How Many Time Points?

- Depends on type of analysis
  - The more complicated the model, the more time points needed
- Minimum
  - Repeated measures: Two
  - Other models: Three
- More is better.
- Ultimately depends on the model, the research setting, and research questions.

Example: Daily reports of conflict, support, and relationship satisfaction

- Kashy data set
- 103 heterosexual dating couples
- Assessed once daily for 14 days
- Completed daily reports of relationship satisfaction and amount of conflict that day
  - Satisfaction and Conflict are time-varying
- Pretest data for attachment avoidance
  - Measured for both people
  - Time invariant

Person Period Pairwise Dataset

- Each Person by Time combination has its own record
  - Person has its own variable (e.g., Person = 1, 2)
  - Occasion has its own variable (e.g., Day = 1 to 14)

- Required for Multilevel Modeling

- We'll look at it when we get to R

Modeling Two Growth Curves

$$Y_{Wti} = c_{Wi} + b_{Wi}T_{ti} + e_{Wti}$$

$$Y_{Mti} = c_{Mi} + b_{Mi}T_{ti} + e_{Mti}$$

Intercepts

c_{Wi} = Predicted value of women's satisfaction at study midpoint for dyad i

c_{Mi} = Predicted value of men's satisfaction at study midpoint for dyad i

Slopes

b_{Wi} = Average change in women's satisfaction over time for dyad i

b_{Mi} = Average change in men's satisfaction over time for dyad I

Errors at each time point

Women = e_{Wti}

Men = e_{Mti}

Correlation of the Residuals

- If the man reports more satisfaction for a particular day than would be expected given the overall effect of time, does the woman also report more satisfaction for that day?

Random Effects: Variances

- There are six variances
  - two intercepts
    - Do men (and women) differ from each other in their "time zero" predicted score?
  - two slopes for time
    - Do the slopes for men (and women) differ?
  - two error (distance from the line) variances
    - Error variances (deviations from the slope) for men and women

Random Effects: Within Person Correlations

- Man intercept-slope correlation
  - If a man is highly satisfied at the study midpoint, is his change in satisfaction steeper?
- Woman intercept-slope correlation
  - If a woman is highly satisfied at the study midpoint, is her change in satisfaction steeper?

Four Between-Person Correlations

- Correlation of the intercepts between partners
  - Overall, do women who have higher levels of satisfaction at the study midpoint tend to have male partners who are also higher in satisfaction at the study midpoint?
  - That is: Is there a correspondence between level of satisfaction?

- Correlation of the slopes
  - Do women whose satisfaction changes over time tend to have male partners whose satisfaction also changes over time?
  - That is: Is there a correspondence between linear change in satisfaction?

- Two slope-intercept correlations
  - Do women with higher levels of satisfaction have male partners who increase or decrease?
  - Do men with higher levels of satisfaction have female partners who increase or decrease

Estimates of Random Effects

- The random option specifies the variances (given as standard deviations) and covariances (given as correlations) between the intercepts and slopes

Rho

		Man Intercept	Woman Intercept	Man Time Slope	Woman Time Slope
Man Intercept	*sd*				
Woman Intercept	*sd*	*r*			
Man Slope	*sd*	*r*	*r*		
Woman Slope	*sd*	*r*	*r*	*r*	
Residual	*sd*				

R MARKDOWN FILE