Introduction to Dyadic Data Analysis

Internal Workshop on Advanced Statistics

Sunthud Pornprasertmanit
Faculty of Psychology, Chulalongkorn University

Outline

- Introduction
- Problems with Regular Multiple Regression
- Simple Multilevel Analysis
- Negative Within-Dyad Correlation
- Distinguishability
- Data Structure
- Actor-Partner Interdependence Model
- Centering Parameterization
- Interaction Effect

Note on the Workshop

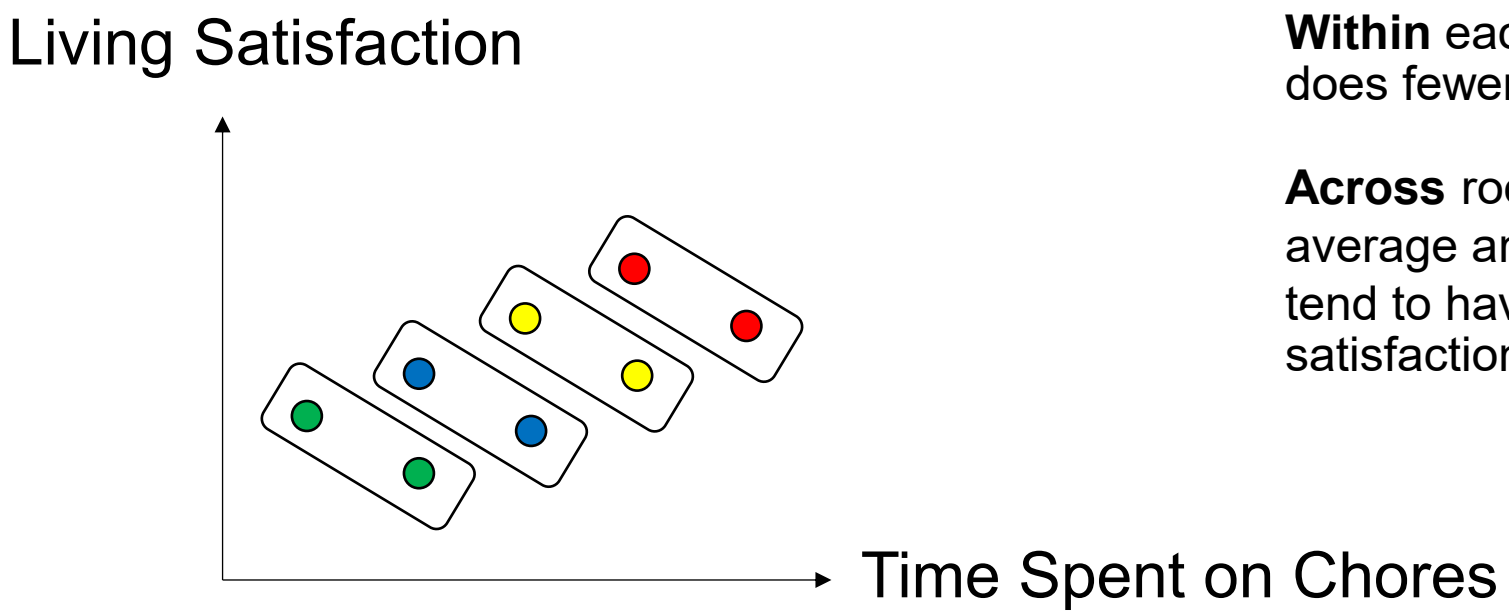
- Several examples use men and women or husbands and wives because these roles are easy to identify in distinguishable dyads. The data and effect patterns are simulated solely for illustration and are not intended to represent universal gender differences or stereotypes.
- During the lecture, please focus on understanding the main ideas. You do not need to understand every line of R code or every equation immediately. New statistical methods often take several rounds of review and practice to fully digest.

Introduction

- How does time spent on chores among roommates influence living satisfaction?
- Within the same room, the roommate who does fewer chores may be more satisfied than the roommate who does more chores.
- Thus, within-room association may be negative because of workload comparison.
- However, between rooms, higher average chore time may predict higher average living satisfaction, possibly because cleaner rooms are more pleasant.

Introduction

- Time Spent on Chores and Living Satisfaction

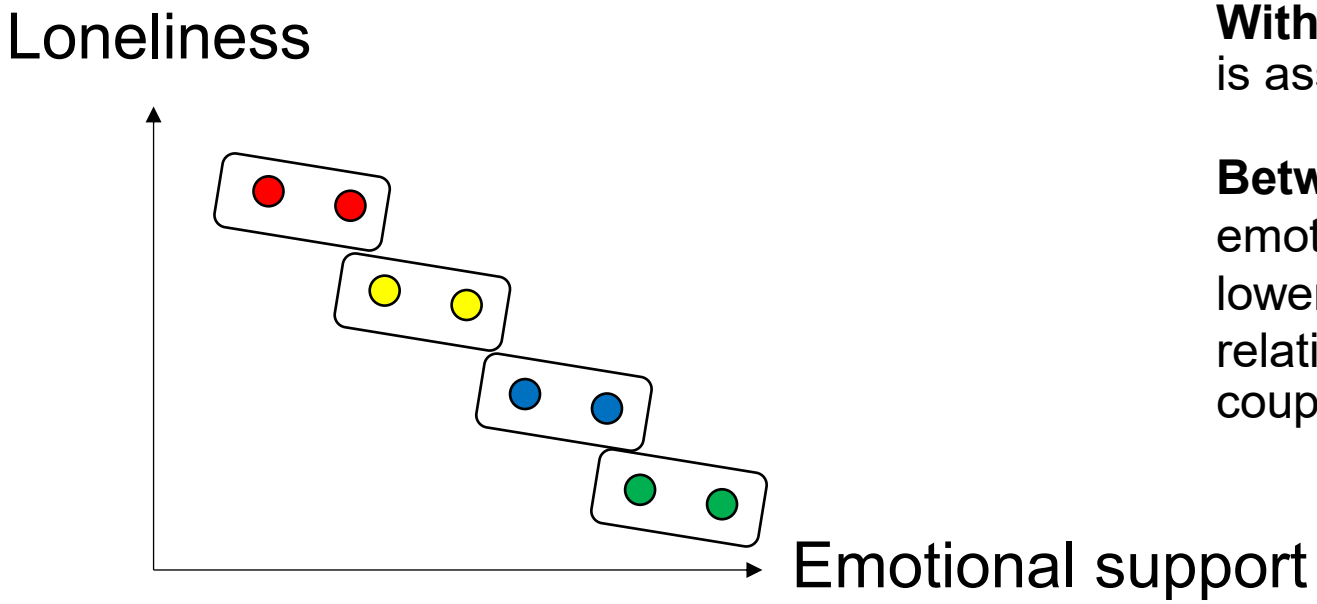


Within each room, the roommate who does fewer chores is more satisfied.

Across rooms, those with a higher average amount of time spent on chores tend to have higher average living satisfaction.

Introduction

- Emotional Support and Loneliness Among Couples

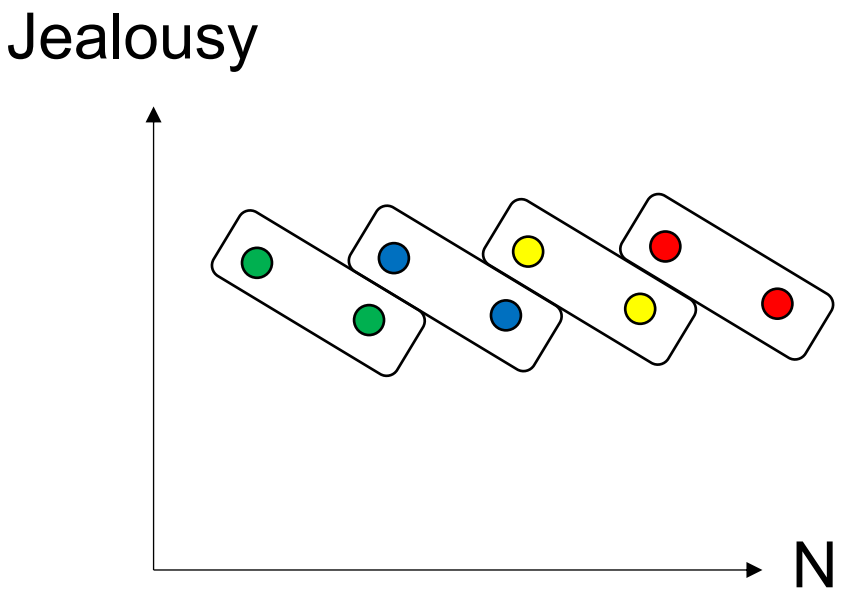


Within couples, more emotional support is associated with less loneliness.

Between couples, higher average emotional support is also associated with lower average loneliness. The relationship is stronger than the within-couple relationship.

Introduction

- Differential Compliments Received and Jealousy Within Twins



Within twin pairs, the twin who received more compliments felt less jealousy toward the other twin.

Between twin pairs, the average number of compliments received was not related to the average level of jealousy. The effect was mainly about comparison within the twin pair.

Introduction

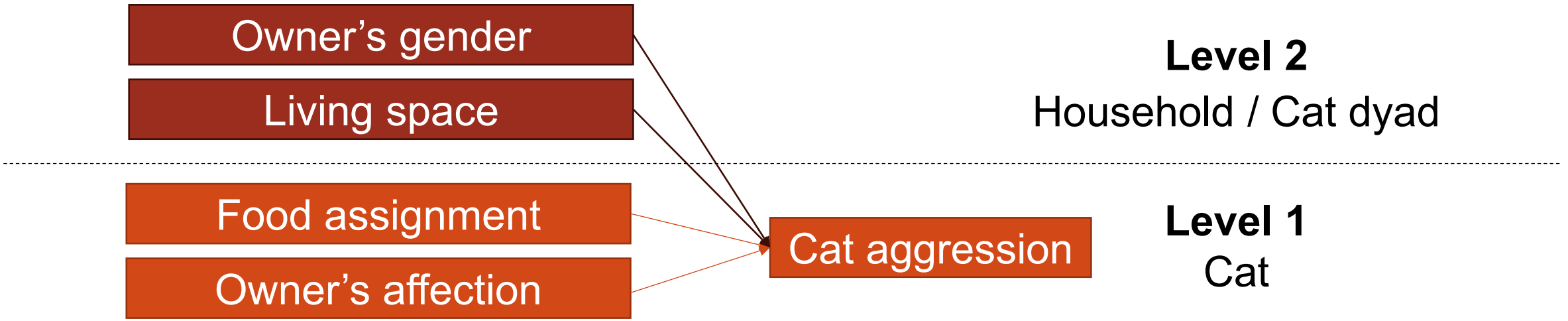
- Dyadic data analysis is used to analyze data in which observations can be matched into pairs.
- The data have a nested structure: two persons or entities are nested within each pair.
- Many naturally occurring relationships can be studied as dyads:
  - Couples
  - Best friends
  - Supervisors and subordinates
  - Siblings
  - Directly competing brands
  - Penalty kickers and goalkeepers in soccer

Introduction

- Dyadic data analysis can handle variables at two levels:
  - **Individual level**: Time spent on chores, living satisfaction, tenant gender, tenant age
  - **Pair level**: Length of time living together, room type (studio, one-bedroom, or two-bedroom), room size in square meters
  - The outcome is usually measured at the individual level, such as living satisfaction.
  - However, pair-level outcomes can also be analyzed, such as length of time living together or frequency of parties

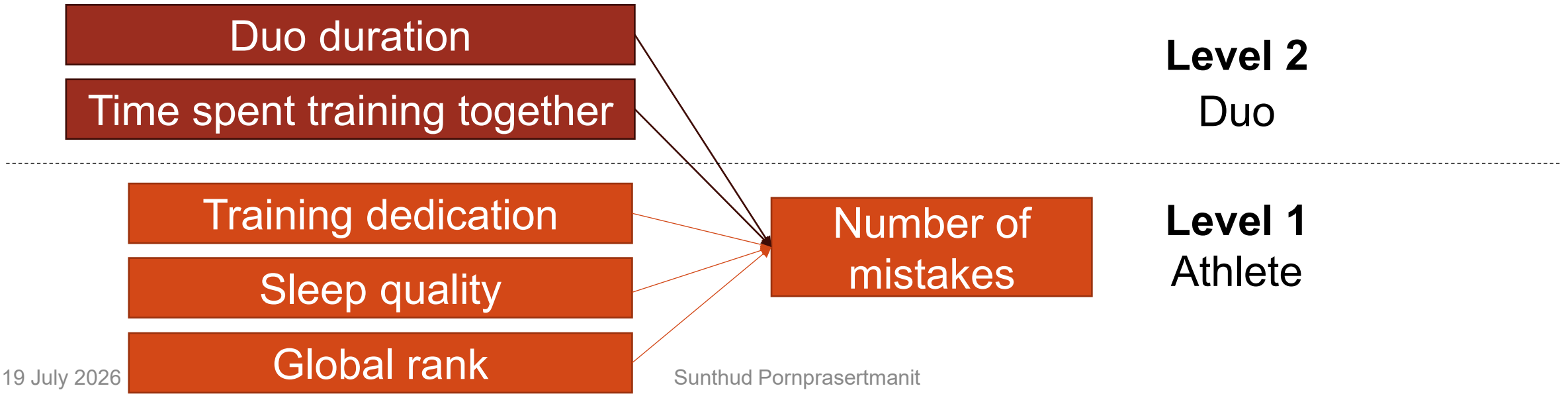
Introduction

- Cat-dyad study predicting cats' aggressive behavior
  - Cat-level predictors: food assignment, owner's affection toward each cat
  - Dyad-level predictors: owner's gender, living space



Introduction

- Badminton-duo study predicting the number of mistakes
  - Individual-level predictors: training dedication, sleep quality, individual global rank
  - Duo-level predictors: duo duration, time spent training together



Introduction

- Notice that the effect of an individual-level predictor on an individual-level outcome may differ across levels:
  - Roommate example
    - Individual level: The roommate who does fewer chores is more satisfied.
    - Room level: Rooms with a higher average amount of time spent on chores tend to have higher average living satisfaction.
  - Twin-pair example
    - Individual level: The twin who received more compliments felt less jealousy toward the other twin.
    - Dyad level: The average number of compliments received was not related to the average level of jealousy.

Introduction

- Dyad-level predictors can play three roles:
  - Predicting the average outcome of the dyad
    - Example: Older twin pairs may feel less jealousy on average.
  - Moderating individual-level effects
    - Example: A family's compliment policy may change how strongly relative compliments predict jealousy within the twin pair.
  - Moderating dyad-level effects
    - Example: A family's compliment policy may change how strongly average compliment frequency predicts average jealousy.

Problems with Regular Multiple Regression

- In the usual nested data structure, individuals are represented in rows and dyads are identified by a dyad ID column.
- Individual-level and dyad-level predictors are distinguished by whether their values differ within each dyad.

ID	Dyad ID	Role	Work	Satisfaction	Marital Status
1	1	H	1	50	1
2	1	W	0	60	1
3	2	H	0	55	1
4	2	W	1	65	1
5	3	H	1	45	0
6	3	W	1	45	0
7	4	H	0	65	0
8	4	W	0	70	0

Problems with Regular Multiple Regression

- Without dyadic data analysis, we might use one of two approaches:
  - **Disaggregated analysis**: Analyze individuals directly and ignore their dyad membership.
  - **Aggregated analysis**: Compute dyad means for individual-level variables and analyze the data at the dyad level.

ID	Dyad ID	Role	Work	Satisfaction	Marital Status
1	1	H	1	50	1
2	1	W	0	60	1
3	2	H	0	55	1
4	2	W	1	65	1
5	3	H	1	45	0
6	3	W	1	45	0
7	4	H	0	65	0
8	4	W	0	70	0

Problems with Regular Multiple Regression

- Disaggregated analysis
  - Analyze satisfaction using work status, marital status, and possibly role as predictors in a regular multiple regression.
  - This ignores the dependence between members of the same dyad.
  - As a result, standard errors may be incorrect and Type I error rates may be distorted.

ID	Dyad ID	Role	Work	Satisfaction	Marital Status
1	1	H	1	50	1
2	1	W	0	60	1
3	2	H	0	55	1
4	2	W	1	65	1
5	3	H	1	45	0
6	3	W	1	45	0
7	4	H	0	65	0
8	4	W	0	70	0

Problems with Regular Multiple Regression

- Aggregated analysis

ID	Dyad ID	Role	Work	Satisfaction	Marital Status
1	1	H	1	50	1
2	1	W	0	60	1
3	2	H	0	55	1
4	2	W	1	65	1
5	3	H	1	45	0
6	3	W	1	45	0
7	4	H	0	65	0
8	4	W	0	70	0



Dyad ID	Average Work Status	Average Satisfaction	Marital Status
1	0.5	55	1
2	0.5	60	1
3	1	45	0
4	0	67.5	0

- Compute the mean of each individual-level variable within each dyad.
- Each row then represents one dyad.
- Use the aggregated data to run a regular multiple regression.

Problems with Regular Multiple Regression

- Aggregated analysis
  - The independence assumption is satisfied because the dyads are independent.
  - However, aggregation removes all within-dyad information.
    - Average work status compares couples in which both, one, or neither partner works.
    - It cannot compare the working and nonworking partners within the same couple.

Simple Multilevel Analysis

- Let Y_{ij} denote the score from individual i in dyad j. The score can be decomposed as

$$Y_{ij} = \mu_j + r_{ij} = a + e_j + r_{ij}$$

Individual score = grand mean + dyad-level deviation + individual-level deviation

- a is the overall intercept. In a model without predictors, it represents the grand mean across all individuals and dyads.
- μ_j is the expected mean for dyad j, and e_j is the deviation of that dyad's mean from the grand mean.
- r_{ij} is the deviation of individual i's score from the expected mean of dyad j.

Simple Multilevel Analysis

- Score decomposition in an example dataset

ID	Dyad ID	Y_{ij} (Observed Score)	μ_j (Dyad mean)	a (Grand mean)	e_j (Between-dyad deviation)	r_{ij} (Within-dyad deviation)
1	1	50	55	56.875	-1.875	-5
2	1	60	55	56.875	-1.875	5
3	2	55	60	56.875	3.125	-5
4	2	65	60	56.875	3.125	5
5	3	45	45	56.875	-11.875	0
6	3	45	45	56.875	-11.875	0
7	4	65	67.5	56.875	10.625	-2.5
8	4	70	67.5	56.875	10.625	2.5

Simple Multilevel Analysis

- In multilevel analysis, we are generally not interested in the deviations of specific dyads or individuals. Instead, we quantify how much variability occurs at each level:

$$Y_{ij} = \mu_j + e_{ij} = a + e_j + r_{ij}$$

- $Var(e_j) = Var(\mu_j)$ represents between-dyad variance: how much the dyad means differ from one another.
- $Var(r_{ij})$ represents within-dyad variance: how much individuals differ from their own dyad mean.
- Assuming the dyad-level and individual-level deviations are independent, the total variance is

$$Var(Y_{ij}) = Var(e_j) + Var(r_{ij})$$

Total Variance = Between-dyad variance + Within-dyad variance

Simple Multilevel Analysis

- The Intraclass correlation (ICC) quantifies the degree of similarity between individuals within the same dyad.

$$ICC = \frac{Var(e_j)}{Var(Y_{ij})} = \frac{Var(e_j)}{Var(e_j) + Var(r_{ij})}$$

- That is, ICC is the proportion of the total variance attribute to differences between dyads.
- Conceptually, place one randomly selected member of each dyad in one column and the other member in a second column. The correlation between the two columns is the ICC.

Simple Multilevel Analysis

Random Labeling Within Each Twin Pair

ID	Twin ID	IQ
1	1	100
2	1	105
3	2	97
4	2	92
5	3	103
6	3	108
7	4	98
8	4	102

Twin ID	IQ1	IQ2
1	105	100
2	92	97
3	108	103
4	98	102

$r = .746$

With a large number of dyads, the correlation between IQ1 and IQ2 approaches the ICC.

Simple Multilevel Analysis

- We will use the `nlme` package in R because it supports several features useful for data analysis.
- The example dataset includes 216 identical-twin pairs and 239 fraternal-twin pairs. We will estimate the ICC for IQ.

Fit the multilevel model

Read the CSV data file

Load the `nlme` package

```
> dat <- read.table("twiniq.csv", sep=",", header=TRUE)
> library(nlme)
> out <- lme(iq ~ 1, random=~1|pairid, data=dat)
```

Specify the dataset containing `iq` and `pairid`. Without this argument, R will search for these variables outside the dataset.

Save the fitted model as `out` for later use

Use `lme()` to fit a multilevel model

Model IQ using an intercept only. The intercept, a, represents the grand mean.

Allow the intercept to vary across twin pairs. This captures differences in mean IQ between pairs.

Use 1 because 1 is implicitly in the (null) model: $Y_{ij} = a(1) + e_j + r_{ij}$

Display a summary of the fitted model

```
> dat <- read.table("twiniq.csv", sep=",", header=TRUE)
> library(nlme)
> out <- lme(iq ~ 1, random=~1|pairid, data=dat)
> summary(out)
Linear mixed-effects model fit by REML
  Data: dat
       AIC      BIC    logLik
  7196.625 7211.062 -3595.313

Random effects:
 Formula: ~1 | pairid
        (Intercept) Residual
StdDev:    11.65243 8.541174

Fixed effects:  iq ~ 1
               Value Std.Error  DF t-value p-value
(Intercept) 99.73187 0.6152906 455 162.089       0

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-5.589005640 -0.485073265 -0.009543101  0.515242306  2.393682058

Number of Observations: 910
Number of Groups: 455
```

$SD(e_j) = 11.65$ $Var(e_j) = 135.779$

$SD(r_{ij}) = 8.54$ $Var(r_{ij}) = 72.952$

$$ICC = \frac{Var(e_j)}{Var(e_j) + Var(r_{ij})} = \frac{135.779}{135.779 + 72.952} = .65$$

Approximately 65% of the total variance in IQ is attributable to differences between twin pairs. The estimate is pooled across identical and fraternal twin pairs.

The estimated grand mean of IQ is 99.73, *SE* = 0.62.

910 individuals from 455 twin pairs

Negative Within-Dyad Correlation

$$ICC = \frac{Var(e_j)}{Var(Y_{ij})} = \frac{Var(e_j)}{Var(e_j) + Var(r_{ij})}$$

- Under the random-intercept variance-components model, the estimated ICC is constrained to range from 0 to 1.
- However, members of a dyad may be negatively associated: when one member has a high value, the other tends to have a low value.
  - For example, if one roommate spends more time on chores, the other roommate may spend less time on chores.

Negative Within-Dyad Correlation

- Another way to analyze dyadic data is to allow the two members' residual to be correlated.

$$Y_{ij} = a + r_{ij}$$

- For each dyad j, the residuals of the two members are correlated:

$$Cor(r_{1j}, r_{2j}) = \rho$$

- The correlation ρ is assumed to be the same across dyads.
- This model is similar to regular regression, except that the two observations within each dyad are allowed to be related.
- Unlike the random-intercept ICC, ρ can be positive, zero, or negative.

Negative Within-Dyad Correlation

- The within-dyad correlation (i.e., ICC) describes how the scores of two members of a dyad are related:
  - Positive correlation: When one member has a higher value, the other member also tends to have a higher value.
    - Examples: spouses' heights; siblings' IQ scores
  - Zero correlation: One member's value provides no information about the other member's value.
    - Example: the values shown by two dice rolled independently
  - Negative correlation: When one member has a higher value, the other member tends to have a lower value.
    - Example: the amounts of speaking time used by a client and a counselor during a session of fixed length

Negative Within-Dyad Correlation

- To estimate the residual correlation between members of each dyad, we can use the `gls()` function in the `nlme` package.

Specify the residual correlation structure. Other correlation structures can also be specified, such as autocorrelation for longitudinal data

```
> out <- gls(iq ~ 1, correlation = corCompSymm(form = ~1|pairid), data=dat)
```

Allow observations within the same pair to have a common residual correlation.

Technically, `corCompSymm()` assumes that all observations within the same group have the same correlation. For dyadic data, each group has only two members, so the model estimates one within-dyad correlation.

```
> out <- gls(iq ~ 1, correlation = corCompSymm(form = ~1|pairid), data=dat)
> summary(out)
Generalized least squares fit by REML
  Model: iq ~ 1
  Data: dat
       AIC      BIC    logLik
  7196.625 7211.062 -3595.313

Correlation Structure: Compound symmetry
 Formula: ~1 | pairid
 Parameter estimate(s):
     Rho
0.650499

Coefficients:
               Value Std.Error t-value p-value
(Intercept) 99.73187 0.6152906 162.089       0

Standardized residuals:
       Min         Q1        Med         Q3        Max
-6.9722597 -0.6043852 -0.0506570  0.7107193  2.6487680

Residual standard error: 14.44752
Degrees of freedom: 910 total; 909 residual
```

Within-dyad correlation = .65 → Equal to the ICC from the random-intercept model

The estimated grand mean of IQ is 99.73, *SE* = 0.62.

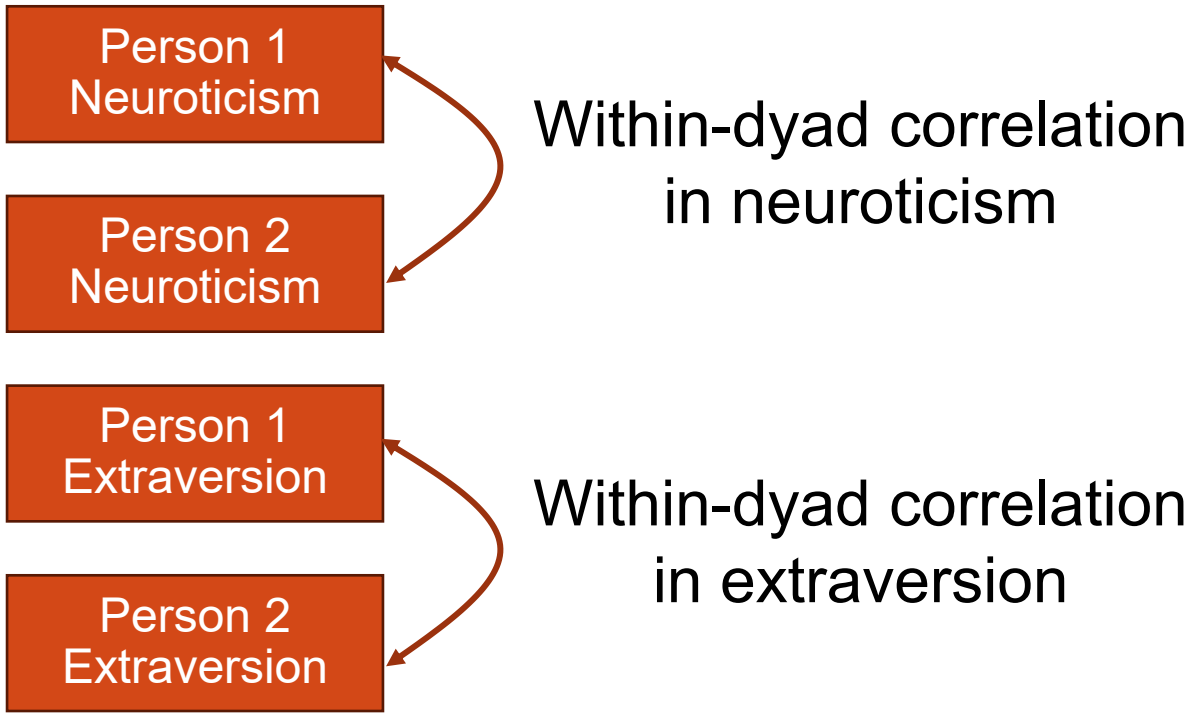
$SD(r_{ij}) = 14.45$

The GLS model estimates the total residual variance and its within-dyad correlation rather than separating the variance into between- and within-dyad components.

Negative Within-Dyad Correlation

Personality data from 500 best-friend dyads include measures of neuroticism and extraversion. Estimate the within-dyad correlation for each trait.





Dyad ID

Individual ID

```
> dat2 <- read.table("bestfriendpersonality.csv", sep=",", header=TRUE)
> head(dat2)
  coupleid subjectid neuro extro
1        1         1    49    40
2        1         2    62    58
3        2         3    37    54
4        2         4    53    45
5        3         5    47    56
6        3         6    56    62
> library(psych)
> describe(dat2)
           vars    n   mean     sd median trimmed    mad min  max range skew kurtosis   se
coupleid      1 1000 250.50 144.41  250.5  250.50 185.32   1  500   499 0.00    -1.20 4.57
subjectid     2 1000 500.50 288.82  500.5  500.50 370.65   1 1000   999 0.00    -1.20 9.13
neuro         3 1000  49.61  10.02   49.0   49.43  10.38  16   80    64 0.15    -0.15 0.32
extro         4 1000  49.58   9.68   49.0   49.48  10.38  20   75    55 0.03    -0.09 0.31
```

Use `describe()` from the `psych` package to inspect descriptive statistics

Target variables:

- `neuro`: neuroticism → *M* = 49.61, *SD* = 10.02
- `extro`: extraversion → *M* = 49.58, *SD* = 9.68

```
> out2a <- lme(extro ~ 1, random = ~ 1|coupleid, data=dat2)
> summary(out2a)
Linear mixed-effects model fit by REML
  Data: dat2
       AIC      BIC    logLik
  7325.848 7340.569 -3659.924

Random effects:
 Formula: ~1 | coupleid
        (Intercept) Residual
StdDev:    5.552372 7.928678

Fixed effects:  extro ~ 1
             Value Std.Error  DF  t-value p-value
(Intercept) 49.576 0.3528762 500 140.4912       0

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.19222997 -0.58203818 -0.02106288  0.56760793  2.60949368

Number of Observations: 1000
Number of Groups: 500
```

$$\rho = \frac{(5.55)^2}{(5.55)^2 + (7.93)^2} = .33$$

```
> out2b <- lme(neuro ~ 1, random = ~ 1|coupleid, data=dat2)
> summary(out2b)
Linear mixed-effects model fit by REML
  Data: dat2
       AIC      BIC    logLik
  7453.24 7467.96 -3723.62

Random effects:
 Formula: ~1 | coupleid
         (Intercept) Residual
StdDev: 0.0005250498 10.02369

Fixed effects:  neuro ~ 1
             Value Std.Error  DF  t-value p-value
(Intercept) 49.61 0.3169769 500 156.5098       0

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.35305639 -0.65943775 -0.06085583  0.63748975  3.03181742

Number of Observations: 1000
Number of Groups: 500
```

$$\rho = \frac{0}{0 + (10.02)^2} = 0$$

The model converged, but the estimated between-dyad variance is at the boundary of zero! This may indicate either:

- little or no within-dyad association, or
- a negative within-dyad correlation that the random-intercept model cannot represent.

```
> out2c <- gls(extro ~ 1, data = dat2,
+              correlation = corCompSymm(form = ~1|coupleid))
> summary(out2c)
Generalized least squares fit by REML
  Model: extro ~ 1
  Data: dat2
       AIC      BIC    logLik
  7325.848 7340.569 -3659.924

Correlation Structure: Compound symmetry
 Formula: ~1 | coupleid
 Parameter estimate(s):
      Rho
0.3290423

Coefficients:
             Value Std.Error  t-value p-value
(Intercept) 49.576 0.3528763 140.4912       0

Standardized residuals:
        Min          Q1         Med          Q3         Max
-3.05552852 -0.67937367 -0.05950718  0.66367038  2.62658092

Residual standard error: 9.679504
Degrees of freedom: 1000 total; 999 residual
```

$\rho = .33$

Best friends tend to be similar in their levels of extraversion.

```
> out2d <- gls(neuro ~ 1, data = dat2,
+              correlation = corCompSymm(form = ~1|coupleid))
> summary(out2d)
Generalized least squares fit by REML
  Model: neuro ~ 1
  Data: dat2
       AIC      BIC    logLik
  7332.788 7347.508 -3663.394

Correlation Structure: Compound symmetry
 Formula: ~1 | coupleid
 Parameter estimate(s):
      Rho
-0.462961

Coefficients:
             Value Std.Error  t-value p-value
(Intercept) 49.61 0.2322362 213.6187       0

Standardized residuals:
        Min          Q1         Med          Q3         Max
-3.35383324 -0.65959053 -0.06086993  0.63763744  3.03251985

Residual standard error: 10.02137
Degrees of freedom: 1000 total; 999 residual
```

$\rho = -.46$

Within-dyad correlation was indeed negative.

Within best-friend dyads, when one person has higher neuroticism, the other tends to have lower neuroticism.

Negative Within-Dyad Correlation

- For simple dyadic regression, `gls()` is more flexible than a standard random-intercept model because it allows the within-dyad residual correlation to be positive, zero, or negative.
- A positive within-dyad correlation in an intercept-only model does not guarantee that the residual correlation will remain positive after predictors are added.
- Therefore, I recommend using `gls()` unless additional multilevel-model features are needed.
- When using a standard random-intercept model, make sure that the assumption of a nonnegative within-dyad residual correlation is reasonable.

Distinguishability

- In the previous examples, the two members of each dyad did not have clear defined roles. These are called indistinguishable dyads, such as twins, roommates, coworkers, or close friends.
- In other situations, dyad members have clearly defined and consistently identifiable roles. These are called distinguishable dyads, such as: husband and wife, older and younger siblings, staff member and customer, supervisor and subordinate.

Distinguishability

- Both `gls()` and multilevel modeling can be use for indistinguishable and distinguishable dyads.
- However, for distinguishable dyads, path analysis or structural equation modeling is often easier because the two roles can be represented explicitly in separate variables and equations.

Distinguishability

- A role-identification variable must assign one member of each dyad to one role and the other member to the other role.
- The same role definition must be used consistently across all dyads.
  - For example, in supervisor-subordinate dyads, every dyad should contain one supervisor and one subordinate.
  - Dyads consisting of two subordinates, two supervisors, or member with different roles structures should not be included in the same analysis.
- Friends are usually indistinguishable because neither member has a clearly defined role.
  - However, they may be treated as distinguishable if each dyad contains one clearly identified leader and one clearly identified follower.

Distinguishability

- Dyad members may be distinguishable in theory, but the two roles may show little empirical difference.
- A test of distinguishability evaluates whether role-specific parameters differ.
- When meaningful differences are not supported, treating the dyads as indistinguishable may provide a simpler model.
- However, the decision should consider both statistical evidence and theory. A nonsignificant test does not prove that the roles are identical.

Distinguishability

- Clearly meaningful roles should usually be retained, even when the test of distinguishability is nonsignificant.
  - Example: husband and wife, supervisor and subordinate.
- Pooling roles may be reasonable when the role distinction has limited substantive importance and the test of distinguishability is nonsignificant.
  - Examples: red versus blue corner in boxing, first or second mover in chess, taller versus shorter friend, or older and younger adult sibling
- When roles are pooled, corresponding effects are constrained to be equal.
  - Example: the effect from the red boxer to the blue boxer is assumed to equal the effect from the blue boxer to the red boxer.

Data Structure

- In the dyadic data analysis literature, variables can be classified into three types:
  - Between-dyad variables: Values differ across dyads but are the same for both members within each dyad.
    - Examples: length of marriage, twins' age
  - Within-dyads variables: Values differ between the two members within each dyad, but the dyad mean is constant across dyads.
    - Role-identification variables are common examples, such as husband vs. wife or older vs. younger sibling.
  - Mixed variables: Values differ both within dyads and across dyads.
    - Examples: marital satisfaction, personality traits, work hours

Data Structure

- Dyadic data can be arranged in different formats depending on the statistical model:
  - Long format: Each row represents one individual, with a dyad ID indicating which two individuals belong together.
  - Wide format: Each row represents one dyad, with separate columns for the two members' variables.
    - This format is primarily used for distinguishable dyads.
  - Pairwise format: Each row represents one individual and includes additional columns containing the partner's scores.

Long Format

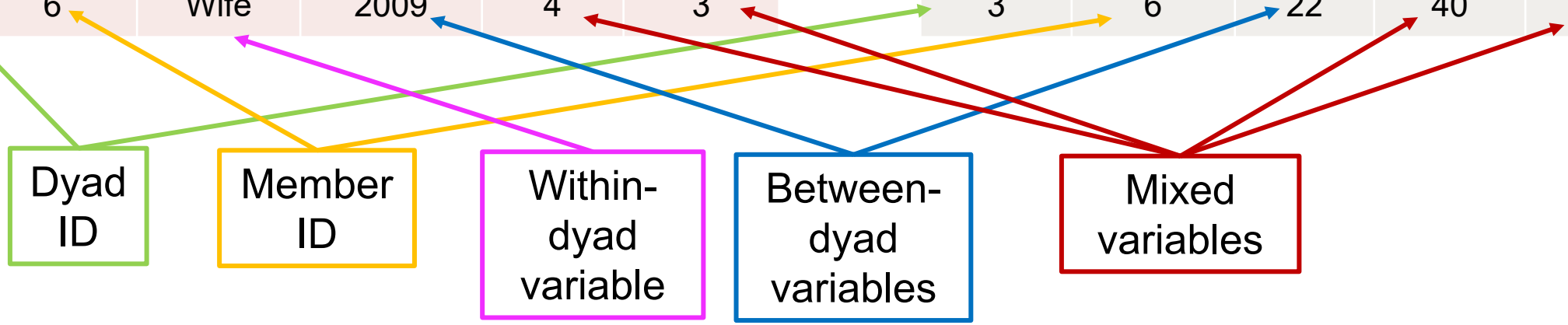
Rows represent individuals.

Distinguishable Dyads (Spouses)

Dyad ID	Member ID	Role	Year Married	Satisfaction	Feeling Loved
1	1	Husband	2019	8	7
1	2	Wife	2019	9	7
2	3	Husband	2015	6	5
2	4	Wife	2015	7	7
3	5	Husband	2009	5	6
3	6	Wife	2009	4	3

Indistinguishable Dyads (Twins)

Dyad ID	Member ID	Age	Extraversion	Anxiety
1	1	19	45	55
1	2	19	42	64
2	3	16	56	46
2	4	16	58	54
3	5	22	46	46
3	6	22	40	37



Wide Format

Use `reshape()` to convert between long and wide formats.

Dyad ID	Member ID	Role	Year Married	Satisfaction	Feeling Loved
1	1	Husband	2019	8	7
1	2	Wife	2019	9	7
2	3	Husband	2015	6	5
2	4	Wife	2015	7	7
3	5	Husband	2009	5	6
3	6	Wife	2009	4	3

Between-dyad variable require only one column.

Mixed variables require separate columns for each role.

Dyad ID	Year Married	Husband Satisfaction	Wife Satisfaction	Husband Feeling Loved	Wife Feeling Loved
1	2019	8	9	7	7
1	2015	6	7	5	7
2	2009	5	4	6	3

Each row represents one dyad.

Pairwise Format

Rows represent individuals.
Additional columns contain the partner's scores.

Distinguishable Dyads (Spouses)

Dyad ID	Member ID	Role	Year Married	Self Satisfaction	Self Feeling Loved	Partner Satisfaction	Partner Feeling Loved
1	1	Husband	2019	8	7	9	7
1	2	Wife	2019	9	7	8	7
2	3	Husband	2015	6	5	7	7
2	4	Wife	2015	7	7	6	5
3	5	Husband	2009	5	6	4	3
3	6	Wife	2009	4	3	5	6

Pairwise Format

Rows represent individuals.
Additional columns contain the partner's scores.

Indistinguishable Dyads (Twins)

Dyad ID	Member ID	Age	Self Extraversion	Self Anxiety	Partner Extraversion	Partner Anxiety
1	1	19	45	55	42	64
1	2	19	42	64	45	55
2	3	16	56	46	58	54
2	4	16	58	54	56	46
3	5	22	46	46	40	37
3	6	22	40	37	46	46

Data Structure

- For distinguishable dyads in wide format, the within-dyad correlation can be computed directly as the correlation between two role-specific columns.
- In long format, the same correlation can be estimated using `gls()` by:
  - including role as a predictor,
  - allowing residuals within each dyad to be correlated, and
  - allowing the two roles to have different residual variances.
- With these specifications, the estimated within-dyad correlation should match the ordinary correlation computed from the wide-format data.

```
> dat3 <- read.table("spousedistress.csv", sep=",", header=TRUE)
> head(dat3)
  coupleid menneuro mendistress womenneuro womendistress lengthrela
1        1 57.96475    46.85858   46.86245      41.20560   2.756603
2        2 59.67379    51.74511   45.94048      59.98045   4.868493
3        3 55.40466    41.23618   35.45253      44.24412   2.329553
4        4 62.87687    48.03789   51.41035      48.63686   3.428704
5        5 37.22799    48.83937   64.84604      43.73324   2.295287
6        6 53.68804    48.67451   39.46492      54.74921   3.663715

> library(psych)
> describe(dat3)
             vars   n   mean    sd median trimmed   mad   min    max  range  skew kurtosis   se
coupleid        1 250 125.50 72.31 125.50  125.50 92.66  1.00 250.00 249.00  0.00    -1.21 4.57
menneuro        2 250  50.83 10.27  51.75   51.16  9.88 17.31  76.30  58.99 -0.35     0.02 0.65
mendistress     3 250  49.78  9.23  50.39   50.07  9.40 25.78  71.54  45.76 -0.26    -0.30 0.58
womenneuro      4 250  50.00 10.24  50.29   49.94 10.76 20.67  77.09  56.42 -0.02    -0.30 0.65
womendistress   5 250  50.40  8.76  51.13   50.41  8.85 27.55  73.31  45.76 -0.03    -0.23 0.55
lengthrela      6 250   2.59  1.67   2.30    2.37  1.25  0.10  10.58  10.48  1.54     3.33 0.11

> with(dat3, cor(menneuro, womenneuro))
[1] -0.06424998
> with(dat3, cor(mendistress, womendistress))
[1] 0.7704816
```

Compute the within-dyad correlation between two partners' scores.

Men and women had similar mean neuroticism scores, *M* = 50.83 and 50.00, with *SD* = 10.27 and 10.24, respectively. Their neuroticism were unrelated within couples, *r* = -.06. Mean distress scores were 49.78 for men and 50.40 for women, with *SD* = 9.23 and 8.76, respectively. Distress was strongly positively correlated within couples, *r* = .77.

Use `reshape()` to convert data between long and wide formats.

```
> dat3long <- reshape(dat3,
+                     varying = list(
+                       neuro = c("menneuro", "womenneuro"),
+                       distress = c("mendistress", "womendistress")
+                     ),
+                     v.names = c("neuro", "distress"),
+                     timevar = "gender",
+                     times = c("men", "women"),
+                     idvar = "coupleid",
+                     direction = "long")
```

Specify the sets of role-specific variables to combine.

Name the variables in long-format dataset.

Name the role-identification variable

Specify the role labels in the same order as specified in `varying`

Specify the dyad ID variable.

Change the data to long format.

You do need to memorize this syntax. Focus on understanding what each argument does.

```
> head(dat3long)
      coupleid lengthrela gender    neuro distress
1.men        1   2.756603    men 57.96475 46.85858
2.men        2   4.868493    men 59.67379 51.74511
3.men        3   2.329553    men 55.40466 41.23618
4.men        4   3.428704    men 62.87687 48.03789
5.men        5   2.295287    men 37.22799 48.83937
6.men        6   3.663715    men 53.68804 48.67451
> tail(dat3long)
          coupleid lengthrela gender    neuro distress
245.women      245  1.6962358  women 55.06318 50.97174
246.women      246  0.6168626  women 61.93904 57.37814
247.women      247  1.0067264  women 61.81908 59.08371
248.women      248  5.8865185  women 37.25310 43.28311
249.women      249  2.1395992  women 49.97646 42.25911
250.women      250  1.6421977  women 55.33418 50.00199
```

`lengthrela` is a between-dyad variable, so its value is repeated for both members of each couple.

`neuro` and `distress` are mixed variables. Their values are aligned with the appropriate gender and couple ID.

The argument names `timevar` and `times` come from the common usage of `reshape()` for longitudinal data.

To estimate a within-dyad correlation for distinguishable dyads using `gls()`:

1. Include role as a predictor to allow different role means.
2. Allow the two roles to have different residual variances.
3. Allow the residuals of members within the same dyad to be correlated.

Estimate the role difference in neuroticism

Estimate residual correlation within dyads

```
> out31 <- gls(neuro ~ gender, data = dat3long,
+              correlation = corCompSymm(form = ~1|coupleid),
+             weights = varIdent(form = ~1|gender))
```

Specify the residual variance structure

Allow men and women to have different residual variances

In `gls()`, indistinguishable and distinguishable dyads are analyzed in similar ways. For distinguishable dyads, analysts simply include role variable as a predictor and allow residual variances to differ across roles.

```
> out31 <- gls(neuro ~ gender, data = dat3long,
+              correlation = corCompSymm(form = ~1|coupleid),
+             weights = varIdent(form = ~1|gender))
> summary(out31)
Generalized least squares fit by REML
  Model: neuro ~ gender
  Data: dat3long
       AIC      BIC    logLik
  3751.805 3772.858 -1870.903

Correlation Structure: Compound symmetry
 Formula: ~1 | coupleid
 Parameter estimate(s):
        Rho
-0.0642459
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | gender
 Parameter estimates:
     men    women
1.000000 0.997213

Coefficients:
               Value Std.Error  t-value p-value
(Intercept) 50.83044 0.6495379 78.25631  0.0000
genderwomen -0.82938 0.9463140 -0.87643  0.3812

 Correlation:
            (Intr)
genderwomen -0.73

Standardized residuals:
        Min          Q1         Med          Q3         Max
-3.26379369 -0.73965051  0.06067321  0.64713827  2.64521743

Residual standard error: 10.2701
Degrees of freedom: 500 total; 498 residual
```

Matches the correlations computed from the wide-format data

Estimated mean for the reference role, men

Women's mean minus men's mean

SD(men) = 10.27 × 1 = 10.27
SD(women) = 10.27 × 0.9972 = 10.24

```
> out32 <- gls(distress ~ gender, data = dat3long,
+              correlation = corCompSymm(form = ~1|coupleid),
+              weights = varIdent(form = ~1|gender))
> summary(out32)
Generalized least squares fit by REML
  Model: distress ~ gender
  Data: dat3long
       AIC      BIC    logLik
  3397.433 3418.486 -1693.716

Correlation Structure: Compound symmetry
 Formula: ~1 | coupleid
 Parameter estimate(s):
       Rho
0.7704815
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | gender
 Parameter estimates:
      men     women
1.0000000 0.9488146

Coefficients:
               Value Std.Error  t-value p-value
(Intercept) 49.77872 0.5837289 85.27712  0.0000
genderwomen  0.62094 0.3863919  1.60701  0.1087

 Correlation:
            (Intr)
genderwomen -0.406

Standardized residuals:
        Min          Q1         Med          Q3         Max
-2.60943290 -0.70916924  0.06832958  0.66718724  2.61579276

Residual standard error: 9.229565
Degrees of freedom: 500 total; 498 residual
```

SD(men) = 9.23 × 1 = 9.23
SD(women) = 9.23 × 0.9488 = 8.76

Convert the data from long format to wide format using `reshape()`.

```
> dat3wide <- reshape(dat3long,
+                     idvar = "coupleid",
+                     timevar = "gender",
+                     v.names = c("neuro", "distress"),
+                     direction = "wide")
> head(dat3wide)
      coupleid lengthrela neuro.men distress.men neuro.women distress.women
1.men        1   2.756603  57.96475     46.85858    46.86245       41.20560
2.men        2   4.868493  59.67379     51.74511    45.94048       59.98045
3.men        3   2.329553  55.40466     41.23618    35.45253       44.24412
4.men        4   3.428704  62.87687     48.03789    51.41035       48.63686
5.men        5   2.295287  37.22799     48.83937    64.84604       43.73324
6.men        6   3.663715  53.68804     48.67451    39.46492       54.74921
```

Specify the dyad ID variable.

Specify the role-identification variable.

Specify the mixed variables to separate by role.

Change the data to wide format.

Only `neuro` and `distress`, which are specified in `v.names`, are separated into role-specific columns. Because `lengthrela` is a between-dyad variable, it remains in a single column.

Test of Distinguishability

- Let *A* and *B* represent the two roles in a dyad and let *X* and *Y* be the variables of interest. For dyad members to be treated as indistinguishable, the following equality constraints mush hold:
  - $\mu_{X_A} = \mu_{X_B}$ and $\mu_{Y_A} = \mu_{Y_B}$ → Equal means
  - $\sigma_{X_A} = \sigma_{X_B}$ and $\sigma_{Y_A} = \sigma_{Y_B}$ → Equal variances
  - $\mathrm{Cov}(X_A, Y_A) = \mathrm{Cov}(X_B, Y_B)$ → Equal within-person associations
  - $\mathrm{Cov}(X_A, Y_B) = \mathrm{Cov}(X_B, Y_A)$ → Equal cross-partner associations
- Example: Are first-born and second-born twins distinguishable based on IQ and emotional intelligence.

Test of Distinguishability

- This test is most useful when the role distinction is unclear and it is uncertain whether separating the two roles is substantively meaningful.
  - When the roles are clearly meaningful, they should generally be retained without testing whether they can be pooled.
  - Treating clearly different roles as indistinguishable may be difficult to justify—for example, treating husbands and wives as interchangeable.
- I do not find it especially useful to conduct a separate omnibus test of distinguishability by imposing many equality constraints in SEM or by conducting multiple univariate tests.

Test of Distinguishability

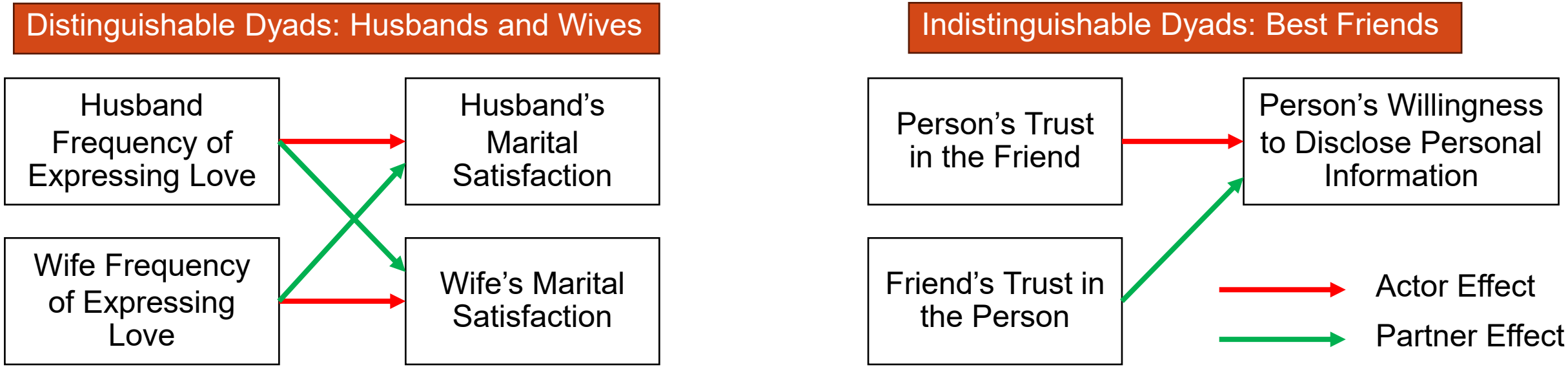
- Rather, fit the final substantive model in two ways
  - Treating the dyads as distinguishable
  - Treating the dyads as indistinguishable
- The two nested models can then be compared. If the distinguishable model does not fit significantly better, and statistical power is adequate, the simpler indistinguishable-dyad model may be used.
- We will return to this model-comparison approach later.

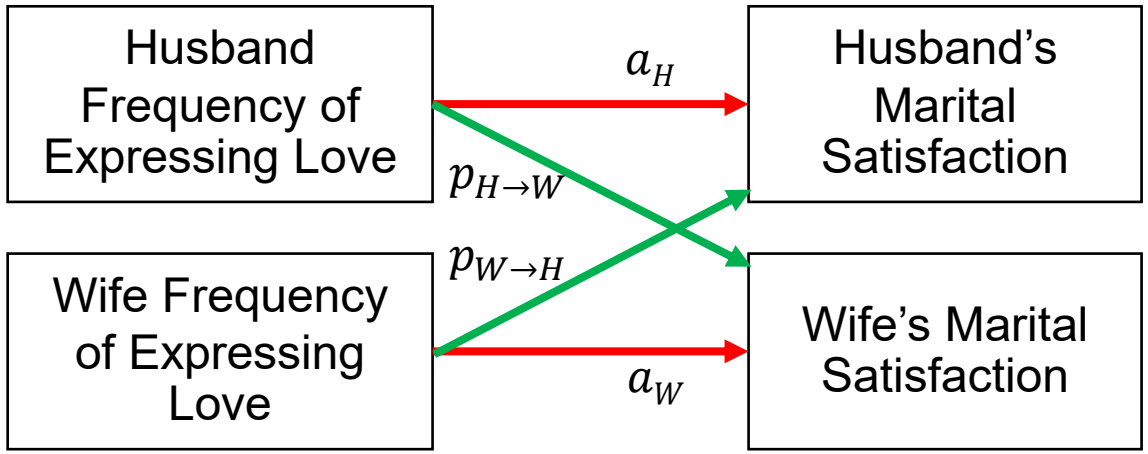
Two Equivalent Parameterizations

- A dyadic regression model can be expressed using two equivalent reparameterizations:
  - Actor-partner parameterization: separate the actor effect from the partner effect.
  - Centering parameterization: separates the effect of the dyad mean from the effect of an individual's deviation from that mean.
- From dyads with two members, these parameterizations describe the same model under equivalent specifications, but their coefficients have different interpretations.

Actor-Partner Interdependence Model

- The actor-partner interdependence model (APIM) estimates how a person's own predictor and the partner's predictor are associated with that person's outcome.
  - Actor effect: the effect of a person's own predictor on their own outcome
  - Partner effect: the effect of the partner's predictor on the person's outcome



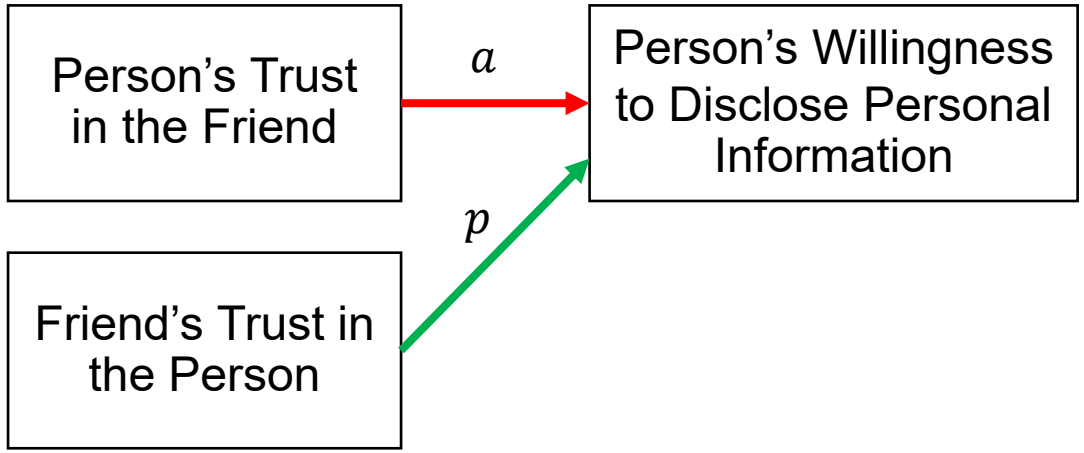


Actor Effect

- a_H: The effect of the husband's frequency of expressive love on his own marital satisfaction.
- a_W: The effect of wife's frequency of expressive love on her own marital satisfaction.
- Possible mechanism: self-fulfilling processes
- Actor effects may differ across roles. For example, expressing love may be strongly associated with marital satisfaction for wives than for husbands.

Partner Effect

- $p_{H\rightarrow W}$: The effect of husband frequency of expressive love on wife's marital satisfaction.
- $p_{W\rightarrow H}$: The effect of wife frequency of expressive love on husband's marital satisfaction.
- Possible mechanism: feeling appreciated by one's partner
- Partner effects may differ across roles. For example, a husband's expression of love may have a stronger association with his wife's satisfaction if such expressions are relatively infrequent and therefore especially meaningful.



Actor Effect

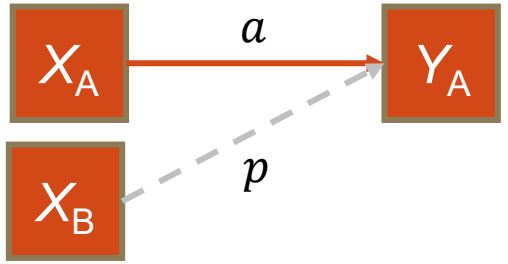
- a: Does a person's own trust in the friend predict that person's willingness to disclose personal information?
- Possible mechanism: Trust makes the person feel safe enough to share personal information.

Partner Effect

- p: Does the friend's trust in the person predict that person's willingness to disclose personal information?
- Possible mechanism: The friend's trust may increase the friend's own disclosure. Observing that disclosure may then make the person more willing to disclose in return.

- In indistinguishable dyads, members do not have distinct roles. Therefore, the actor effect is constrained to be equal across members, and the partner effect is constrained to be equal across member.
- In the APIM, the actor and partner predictors are entered simultaneously. Each effect therefore represents its unique association with the outcome while holding the other predictor constant.

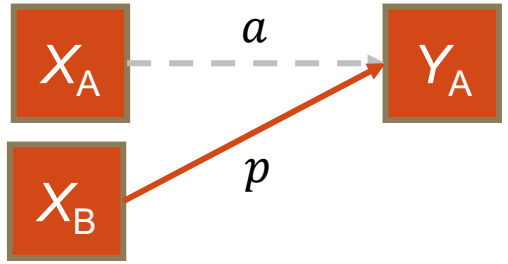
Actor-oriented



$a \neq 0$
$p = 0$

Example: A person's own evaluation of the benefits of a relationship predicts their own intention to remain in the relationship, whereas the partner's evaluation does not.

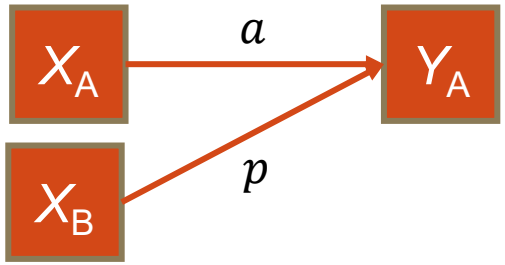
Partner-oriented



$a = 0$
$p \neq 0$

Example: A partner's emotional support predicts a person's stress, whereas the person's own provision of support does not predict their own stress.

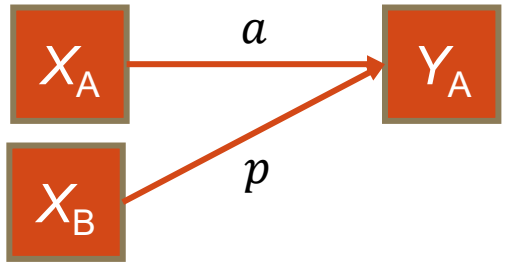
Couple-oriented



$a = p$
$a \neq 0$
$p \neq 0$

Example: Both players' effort contribute equally to each player's satisfaction with their performance in a badminton doubles match.

Social Comparison / Contrast



$a + p = 0$
$a \neq 0$
$p \neq 0$

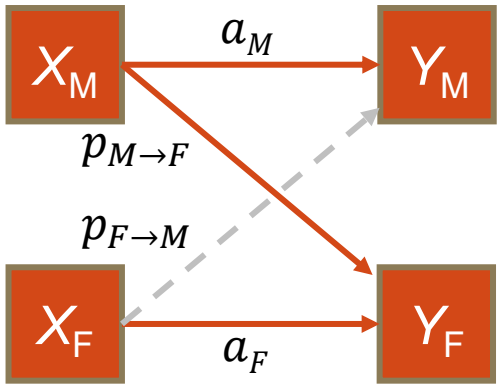
Example: In a singles badminton match, a player's own performance positively predicts satisfaction with the result, whereas the opponent's performance has an equally strong negative effect.

Male patterns	Female patterns: Actor-only	Partner-only	Couple-oriented	Contrast
Actor-only	Double actor-only pattern	Male as the sole predictor	No *partner effect* from male partner	
Partner-only	Female as the sole predictor	Double partner-only pattern	No *actor effect* from male partner	
Couple-oriented	No *partner effect* from female partner	No *actor effect* from female partner	Double couple-oriented pattern	mixed dyadic pattern
Contrast			mixed dyadic pattern	Double contrast pattern

From Table 1, Fitzpatrick et al. (2016)

- The APIM patterns may differ across roles in distinguishable dyads. Because each role can show one of four pattern, there are $4 \times 4 = 16$ possible combinations.
- The table summarizes these combinations, but if is often easier to understand a specific combination by drawing its path diagram.

For example,
Male actor-only +
Female couple-oriented



Male outcome:
Actor-only
$a_M \neq 0$
$p_{F \to M} = 0$

Female outcome:
Couple-oriented
$a_F = p_{M \to F}$
$a_F \neq 0$
$p_{M \to F} \neq 0$

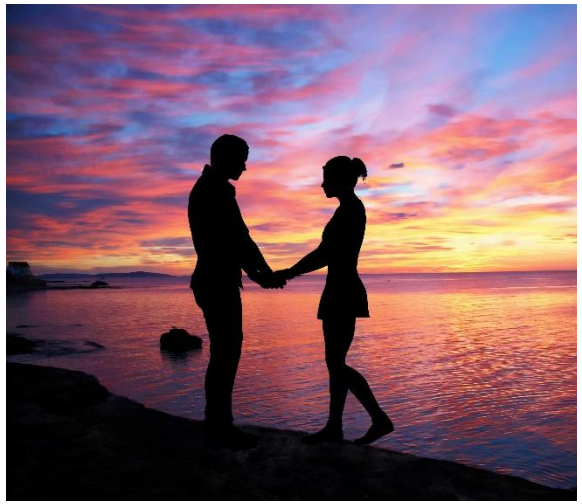
For example, *X* = Income, *Y* = Perceived Financial Success.

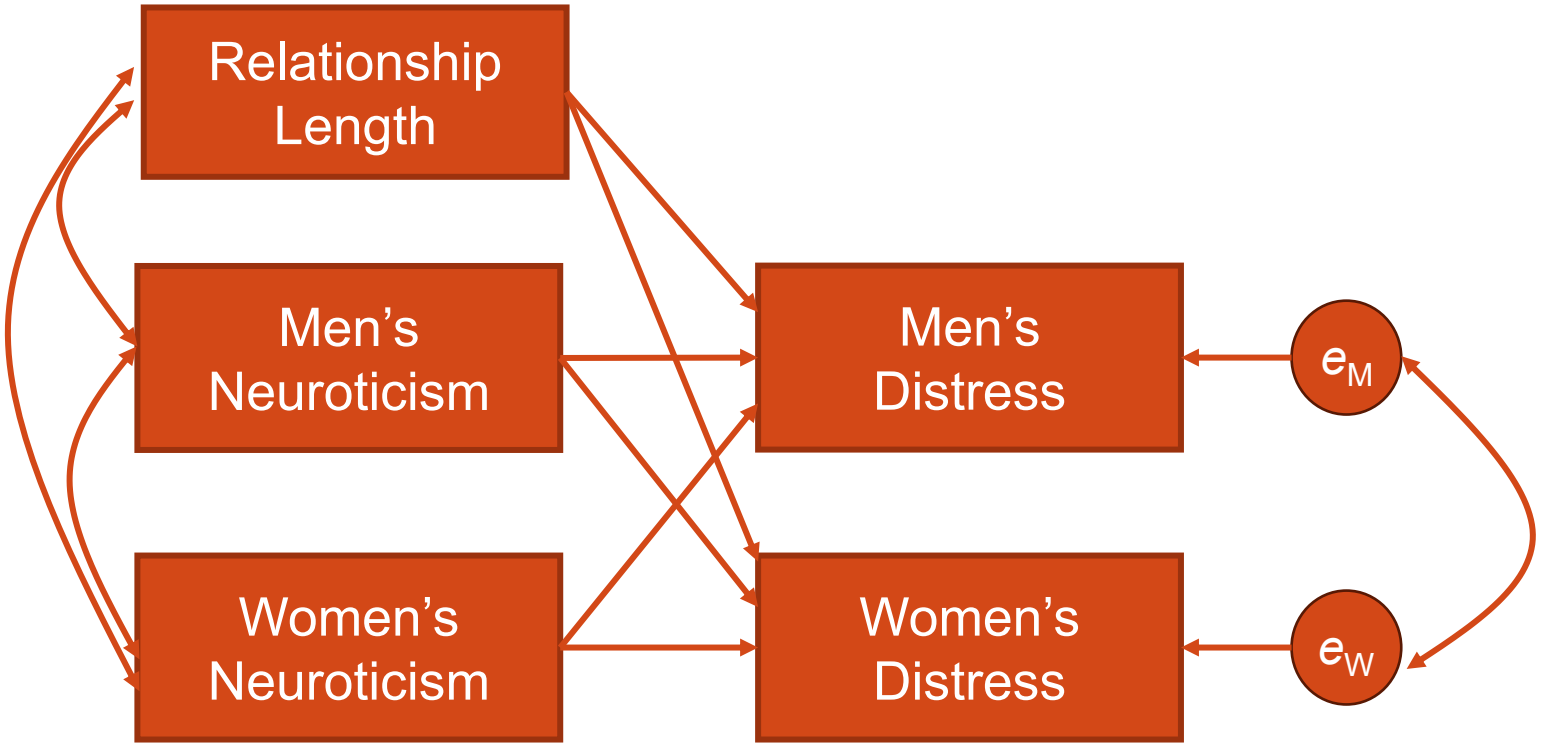
Actor-Partner Interdependence Model

- Estimating the APIM:
  - Indistinguishable dyads: Use `gls()` with pairwise data
  - Distinguishable dyads: Estimate separate role-specific equations using `gls()` or structural equation modeling.
    - For structural equation modeling, we will use `lavaan` because it easily accommodate correlated residuals, equality constraints, and model comparisons.

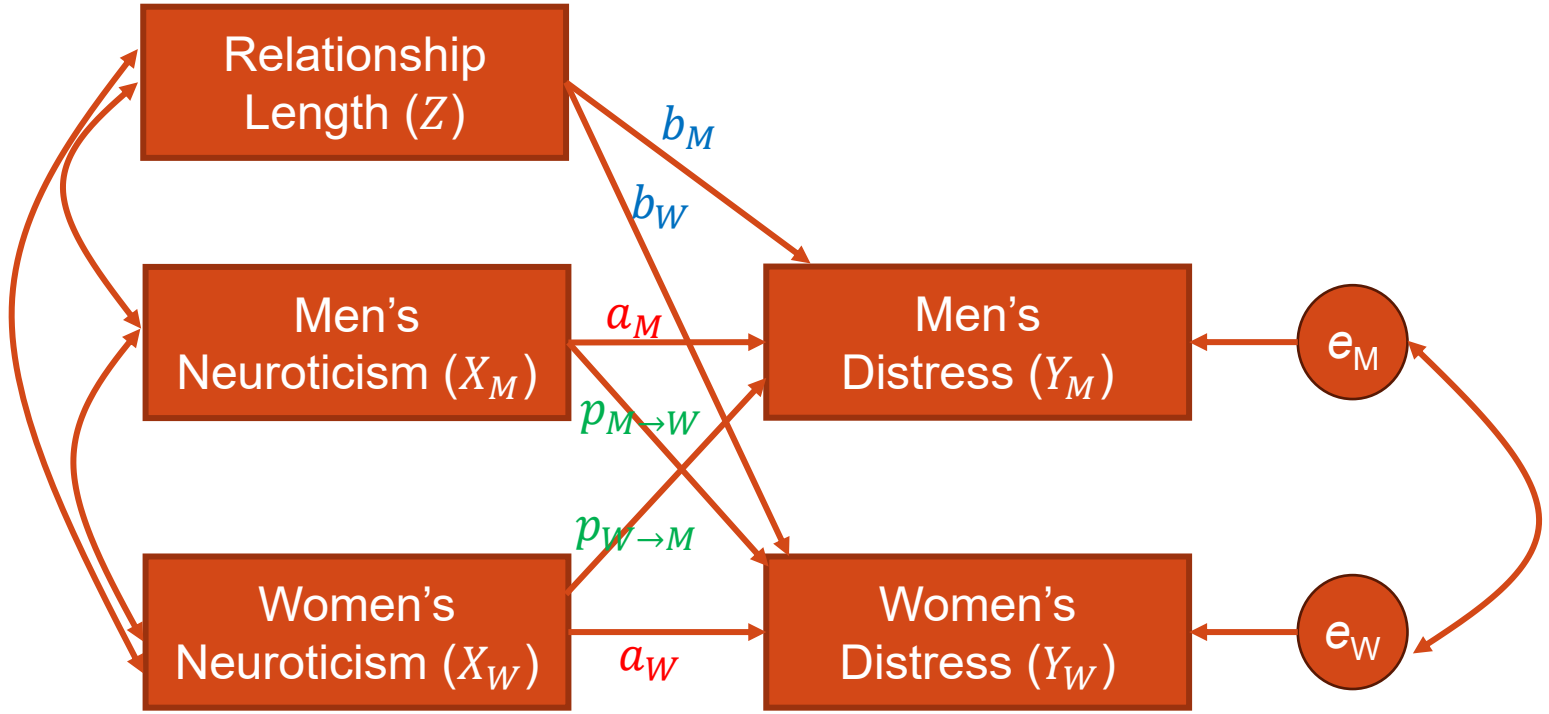
Actor-Partner Interdependence Model

Data from 250 couples. Examine the actor and partner effects of neuroticism on psychological distress, controlling for relationship length.





The exogenous predictors are allowed to correlate, as in regular multiple regression.



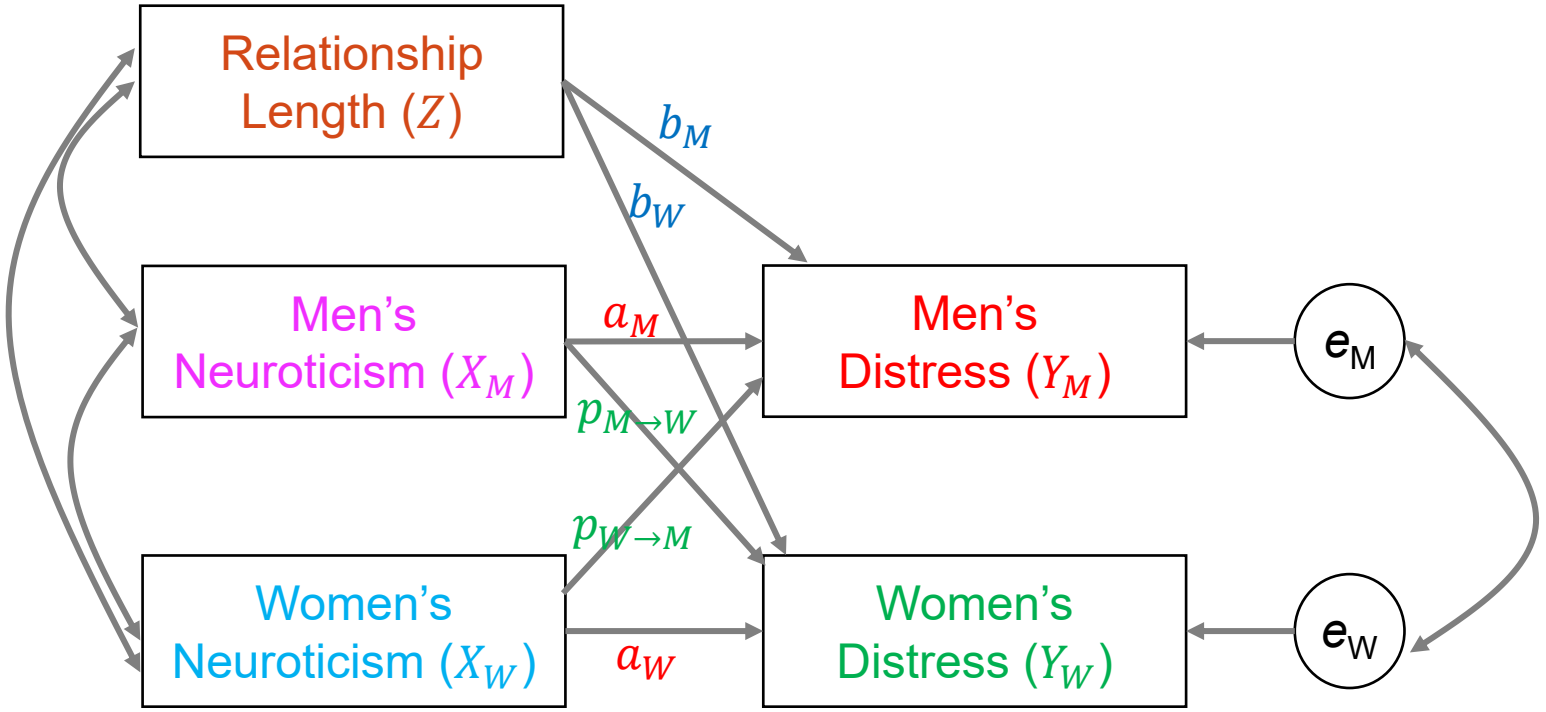
Remaining association between partners' distress scores

This APIM consists of two simultaneous regression equations with correlated residuals:

$$Y_M = i_M + a_M X_M + p_{W\rightarrow M} X_W + b_M Z + e_M$$

$$Y_W = i_W + a_W X_W + p_{M\rightarrow W} X_M + b_W Z + e_W$$

$$\begin{bmatrix} e_M \\ e_W \end{bmatrix} \overset{\text{i.i.d.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_M^2 & \sigma_{MW} \\ \sigma_{MW} & \sigma_W^2 \end{bmatrix} \right)$$



`lavaan` syntax

```
mendistress ~ menneuro + womenneuro + lengthrela
womendistress ~ menneuro + womenneuro + lengthrela
```

Specify two regression equations

```
mendistress ~~ womendistress
```

When both variables are outcomes,
`~~` specifies their residual covariances.

```
> dat3 <- read.table("spousedistress.csv", sep=",", header=TRUE)
> library(lavaan)
> script3 <- '
+ mendistress ~ menneuro + womenneuro + lengthrela
+ womendistress ~ menneuro + womenneuro + lengthrela
+ mendistress ~~ womendistress
+ '
> apim3 <- sem(script3, data=dat3)

> summary(apim3, standardize=TRUE, rsquare=TRUE)
lavaan 0.6-21 ended normally after 20 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         9

  Number of observations                           250

Model Test User Model:

  Test statistic                                 0.000
  Degrees of freedom                                 0

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured
```

Load the `lavaan` package

Fit the specified path model

`standardize`: request standardized estimates

`rsquare`: request the proportion of variances explained

"ended normally" indicates that the model converged

250 couples

This model is just-identified.
These zero values are expected.

Unstandardized estimates | *SE* | *z* | *p* | Standardized estimates

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  mendistress ~
    menneuro          0.357    0.035   10.196    0.000    0.357    0.398
    womenneuro        0.451    0.036   12.456    0.000    0.451    0.500
    lengthrela       -1.850    0.224   -8.260    0.000   -1.850   -0.336
  womendistress ~
    menneuro          0.085    0.042    2.014    0.044    0.085    0.099
    womenneuro        0.291    0.043    6.697    0.000    0.291    0.341
    lengthrela       -2.300    0.269   -8.544    0.000   -2.300   -0.440

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
 .mendistress ~~
   .womendistress    23.757    2.791    8.512    0.000   23.757    0.639

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .mendistress      30.949    2.768   11.180    0.000   30.949    0.365
   .womendistress    44.694    3.998   11.180    0.000   44.694    0.585

R-Square:
                   Estimate
    mendistress       0.635
    womendistress     0.415
```

The predictors explained approximately 64% of the variance in men's distress and 42% of the variance in women's distress.

Actor effect:
- Men's distress: $b = 0.36, p < .001$
- Women's distress: $b = 0.29, p < .001$
- Relatively similar across roles

Partner effect:
- Men's distress: $b = 0.45, p < .001$
- Women's distress: $b = 0.09, p = .044$
- Substantially different across roles

Relationship Length:
- Men's distress: $b = -1.85, p < .001$
- Women's distress: $b = -2.30, p < .001$
- Differ somewhat across roles

After accounting for both partners' neuroticism and relationship length, the residual correlation between men's and women's distress was $r = .64$, $p < .001$.

Unstandardized estimates

Standardized estimates

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  mendistress ~
    menneuro          0.357    0.035   10.196    0.000    0.357    0.398
    womenneuro        0.451    0.036   12.456    0.000    0.451    0.500
    lengthrela       -1.850    0.224   -8.260    0.000   -1.850   -0.336
  womendistress ~
    menneuro          0.085    0.042    2.014    0.044    0.085    0.099
    womenneuro        0.291    0.043    6.697    0.000    0.291    0.341
    lengthrela       -2.300    0.269   -8.544    0.000   -2.300   -0.440

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
 .mendistress ~~
   .womendistress    23.757    2.791    8.512    0.000   23.757    0.639

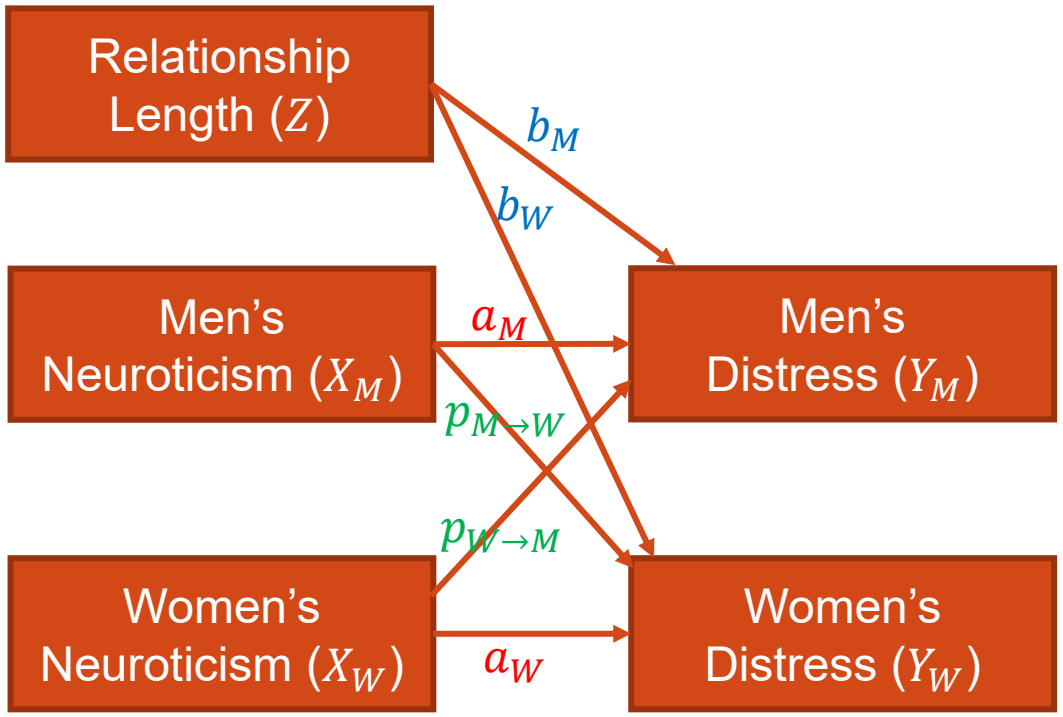
Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .mendistress      30.949    2.768   11.180    0.000   30.949    0.365
   .womendistress    44.694    3.998   11.180    0.000   44.694    0.585

R-Square:
                   Estimate
    mendistress       0.635
    womendistress     0.415
```

- Use unstandardized coefficients to test role differences when corresponding variables have the same measurement scale.
- Standardized coefficients depend on role-specific variances, so differences between them may reflect differences in variability rather than differences in the underlying slopes.

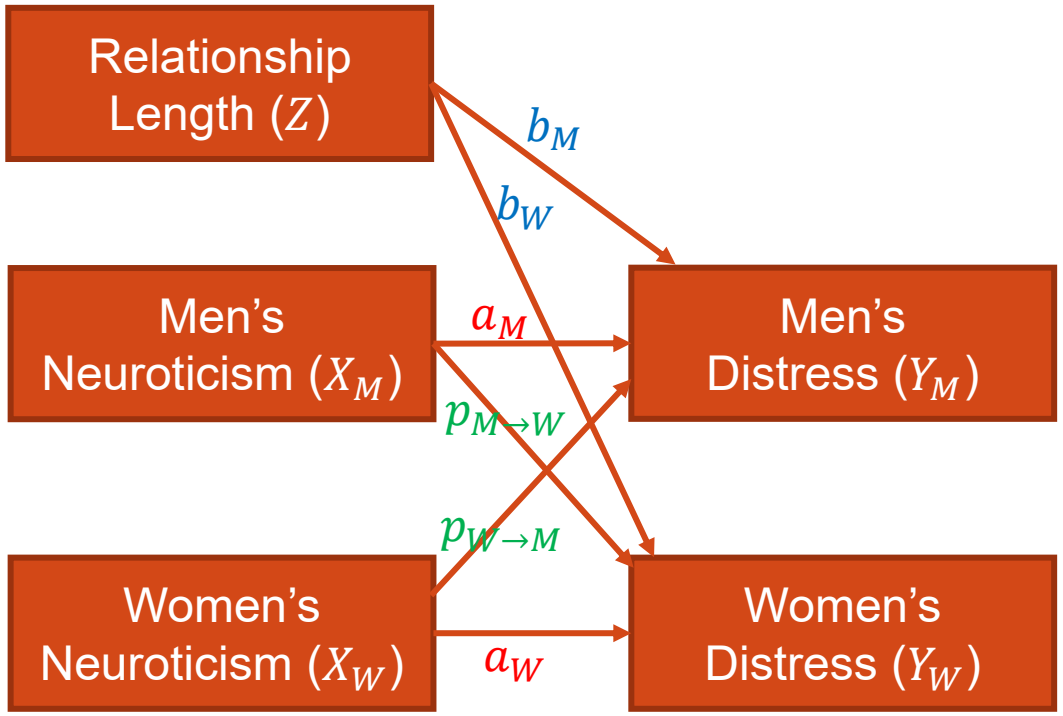
Actor-Partner Interdependence Model

- In SEM, users can define new parameters as functions of estimated parameters and obtain significance tests and confidence intervals for those new parameters.



Potential Defined Parameters:

- For men distress, is the actor effect larger than the partner effect? $a_M - p_{W\to M}$
- For women distress, is the actor effect larger than the partner effect? $a_W - p_{M\to W}$
- Is actor effect larger for men than for women? $a_M - a_W$
- Is partner effect stronger toward men distress than women distress? $p_{W\to M} - p_{M\to W}$



Potential Defined Parameters:

- Does men's neuroticism have a larger association with men's distress than with women's distress? $a_M - p_{M\rightarrow W}$
- Does women's neuroticism have a larger association with men's distress than with women's distress? $a_W - p_{W\rightarrow M}$
- Does relationship length have a larger association with men's distress than with women's distress? $b_M - b_W$
- Is the averaged actor effect significantly different from zero? $(a_M + a_W)/2$
- Is the averaged partner effect significantly different from zero? $(p_{W\rightarrow M} + p_{M\rightarrow W})/2$
- Is the average actor effect larger than the average partner effect? $\frac{a_M+a_W}{2} - \frac{p_{W\rightarrow M}+p_{M\rightarrow W}}{2}$

```
> script3a <- "
+ mendistress ~ am*menneuro + pmw*womenneuro + cm*lengthrela
+ womendistress ~ pwm*menneuro + aw*womenneuro + cw*lengthrela
+ mendistress ~~ womendistress
+
+ # Define New Parameters
+ apmen := am - pmw                    # Compare the actor and partner effects predicting men's distress
+ apwomen := aw - pwm                  # Compare the actor and partner effects predicting women's distress
+ diffa := am - aw                     # Compare actor effects across roles
+ diffp := pmw - pwm                   # Compare partner effects across roles
+ diffmenneuro := am - pwm             # Compare the effects of men's neuroticism on men's and women's distress
+ diffwomenneuro := pmw - aw           # Compare the effects of women's neuroticism on men's and women's distress
+ difflengthrela := cm - cw            # Compare the effects of relationship length on men's and women's distress
+ aveactor := (am + aw) / 2            # Average actor effect
+ avepartner := (pmw + pwm) / 2        # Average partner effect
+ diffactorpartner := aveactor - avepartner # CCompare the average actor and partner effects
+ "
> apim3a <- sem(script3a, data=dat3)
> summary(apim3a)
```

Assign a label to each regression coefficient

Use := to define new parameters.
Use # to add comments within the lavaan model syntax.

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  mendistress ~
    menneuro  (am)    0.357    0.035   10.196    0.000
    womenner (pmw)    0.451    0.036   12.456    0.000
    lengthrl  (cm)   -1.850    0.224   -8.260    0.000
  womendistress ~
    menneuro (pwm)    0.085    0.042    2.014    0.044
    womenner  (aw)    0.291    0.043    6.697    0.000
    lengthrl  (cw)   -2.300    0.269   -8.544    0.000
```

The parameter labels appear beside the corresponding estimates.

Defined Parameters:

	Estimate	Std.Err	z-value	P(>|z|)
apmen	-0.093	0.047	-1.972	0.049
apwomen	0.206	0.057	3.629	0.000
diffa	0.066	0.054	1.231	0.218
diffp	0.366	0.053	6.848	0.000
diffmenneuro	0.273	0.033	8.156	0.000
diffwomenneuro	0.159	0.034	4.624	0.000
difflengthrela	0.450	0.214	2.106	0.035
aveactor	0.324	0.029	11.208	0.000
avepartner	0.268	0.029	9.304	0.000
diffactorprtnr	0.057	0.023	2.506	0.012

1. For men's distress, the actor effect was significantly smaller than the partner effect, $a_M - p_{W\to M} = -0.093, p = .049$.
2. For women's distress, the actor effect was significantly larger than the partner effect, $a_W - p_{M\to W} = 0.206, p < .001$.
3. The actor effects did not differ significantly between men and women, $a_M - a_W = 0.066, p = .22$.
4. The partner effect predicting men's distress was significantly larger than the partner effect predicting women's distress, $p_{W\to M} - p_{M\to W} = 0.366, p < .001$.
5. Men's neuroticism had a significantly larger association with men's distress than with women's distress, $a_M - p_{M\to W} = 0.273, p < .001$.
6. Women's neuroticism had a significantly larger association with men's distress than with women's distress, $a_W - p_{W\to M} = 0.159, p < .001$.
7. The association of relationship length with men's distress was significantly less negative than its association with women's distress, $b_M - b_W = 0.450, p = .035$.
8. The mean actor effect was 0.324, which was significantly different from zero, $p < .001$.
9. The mean partner effect was 0.268, which was significantly different from zero, $p < .001$.
10. The mean actor effect was significantly larger than the mean partner effect, difference = 0.057, $p = .012$.

All comparisons are based on unstandardized coefficients. A positive difference indicates that the first coefficient is numerically larger than the second.

Actor-Partner Interdependence Model

- To estimate the APIM in indistinguishable dyads using `gls()`, the data must be arranged in pairwise format.

		X_A	Y	X_P	
Twin Pair ID	**Member Within Pair**	**Self: WST**	**Self: Academic Success**	**Partner: WST**	**Partner: WST**
1	1	45	55	42	64
1	2	42	64	45	55
2	1	56	46	58	54
2	2	58	54	56	46
3	1	46	46	40	37
3	2	40	37	46	46

WST = Willingness to Succeed Together

Example: Predicting a person's academic success from

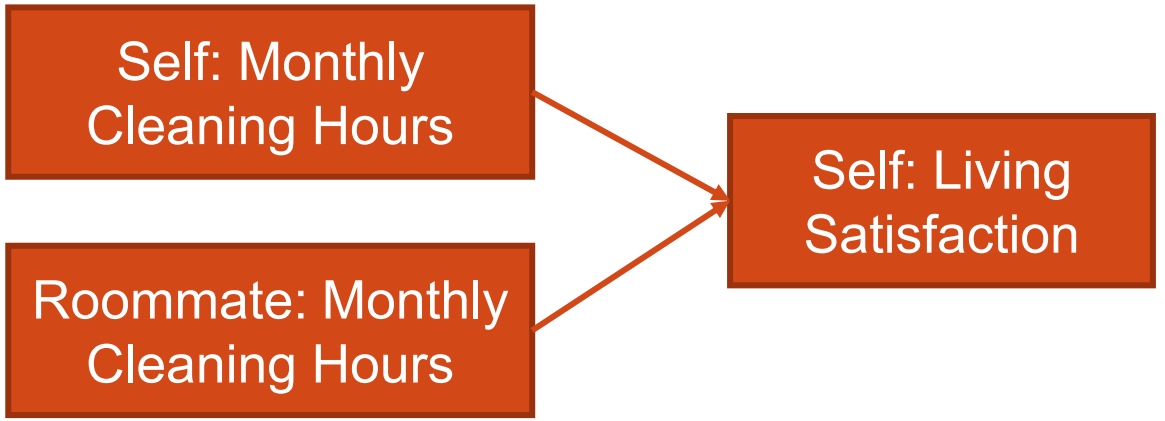
- The person's willingness to succeed together (Actor Effect, X_A)
- The twin's willingness to succeed together (Partner Effect, X_P)

Each twin appears once as the focal person. The self and partner values are reversed in the other twin's rows.

Actor-Partner Interdependence Model

Data from 1,000 Roommates in 500 pairs. Examine whether a person's own monthly cleaning hours and the roommate's monthly cleaning hours predict that person's living satisfaction.





How to make pairwise data

X_{1j} = focal person's score; X_{2j} = partner's score

$\bar{X}_j$ = Average score in dyad j

1. Select the target individual-level variable.
2. Create the dyad mean of that variable.
3. Multiply the dyad mean by 2 and subtract the focal person's score.

$$\bar{X}_j = \frac{X_{1j}+X_{2j}}{2}$$

$$X_{2j} = 2\bar{X}_j - X_{1j}$$

Calculate the dyad mean and attach it to both members.

Target variable

Dyad ID variable

```
> dat4 <- read.table("roommatecleaning.csv", sep=",",
header=TRUE)
> head(dat4)
  coupleid subjectid cleanhrs sat
1        1         1      3.3  50
2        1         2      4.5  56
3        2         3      2.0  59
4        2         4      3.5  36
5        3         5      3.0  61
6        3         6      3.9  59
```

```
> avecleanhrs <- ave(dat4$cleanhrs, dat4$coupleid)
> dat4$cleanhrspartner <- avecleanhrs*2 - dat4$cleanhrs
> head(dat4)
  coupleid subjectid cleanhrs sat cleanhrspartner
1        1         1      3.3  50             4.5
2        1         2      4.5  56             3.3
3        2         3      2.0  59             3.5
4        2         4      3.5  36             2.0
5        3         5      3.0  61             3.9
6        3         6      3.9  59             3.0
```

- This shortcut assumes exactly two members per dyad.
- If either member has a missing value, `ave()` returns `NA` for the dyad mean, so the derived partner value is also `NA`. This is expected. Do not use `na.rm = TRUE`.

```
> library(psych)
> describe(dat4)
               vars    n   mean     sd median trimmed    mad min    max range  skew kurtosis   se
coupleid          1 1000 250.50 144.41  250.5  250.50 185.32   1  500.0 499.0  0.00    -1.20 4.57
subjectid         2 1000 500.50 288.82  500.5  500.50 370.65   1 1000.0 999.0  0.00    -1.20 9.13
cleanhrs          3 1000   3.26   1.00    3.2    3.24   1.04   0    6.3   6.3  0.15    -0.14 0.03
sat               4 1000  50.02  12.70   50.0   50.06  13.34   6   89.0  83.0 -0.05    -0.14 0.40
cleanhrspartner   5 1000   3.26   1.00    3.2    3.24   1.04   0    6.3   6.3  0.15    -0.14 0.03
```

`cleanhrs` and `cleanhrspartner` have identical descriptive statistics because the partner variable contains the same scores rearranged across members. Neither variable has missing values.

Include both the actor and partner predictors.

```
> library(nlme)
> apim4 <- gls(sat ~ 1 + cleanhrs + cleanhrspartner,
+              data = dat4, correlation = corCompSymm(form = ~1|coupleid))
```

`1` specifies the intercept; it is included by default and may be omitted.

Allow the residuals of the two roommates within each dyad to be correlated.

```
> apim4 <- gls(sat ~ 1 + cleanhrs + cleanhrspartner,
+              data = dat4, correlation = corCompSymm(form = ~1|coupleid))
> summary(apim4)
Generalized least squares fit by REML
  Model: sat ~ 1 + cleanhrs + cleanhrspartner
  Data: dat4
       AIC      BIC    logLik
  7339.369 7363.893 -3664.685

Correlation Structure: Compound symmetry
 Formula: ~1 | coupleid
 Parameter estimate(s):
      Rho
0.3421043

Coefficients:
                  Value Std.Error   t-value p-value
(Intercept)    34.12828 2.1089358 16.182701       0
cleanhrs       -2.20285 0.3532515 -6.235925       0
cleanhrspartner  7.08331 0.3532515 20.051761       0

 Correlation:
               (Intr) clenhrs
cleanhrs       -0.890
cleanhrspartner -0.890  0.632

Standardized residuals:
        Min          Q1         Med          Q3         Max
-3.08051766 -0.67341594 -0.02880145  0.68980361  2.66457362

Residual standard error: 9.754758
Degrees of freedom: 1000 total; 997 residual
```

Intercept = predicted living satisfaction when both roommates spend 0 hours per month cleaning.

Actor Effect = -2.20, $p < .001$.
Holding the roommate's cleaning hours constant, each additional hour a person spends cleaning per month is associated with a 2.20-point decrease in that person's living satisfaction.

Partner Effect = 7.08, $p < .001$.
Holding the person's own cleaning hours constant, each additional hour the roommate spends cleaning per month is associated with a 7.08-point increase in the person's living satisfaction.

```
Coefficients:
                  Value Std.Error   t-value p-value
(Intercept)    34.12828 2.1089358 16.182701       0
cleanhrs       -2.20285 0.3532515 -6.235925       0
cleanhrspartner  7.08331 0.3532515 20.051761       0
```

In this simulated example, people dislike doing more cleaning themselves but appreciate having a roommate who cleans more.

In `gls()`, linear combinations of the estimated regression coefficients can be tested using the `multcomp` package. The produce is lest direct than defining parameters in `lavaan`.

For example, we want to test whether the actor and partner effects differ:

$$H_0: B_{\text{actor}} = B_{\text{partner}}$$

$$H_0: B_{\text{actor}} - B_{\text{partner}} = 0$$

$$H_0: (0)B_1 + (1)B_2 + (-1)B_3 = 0$$

Therefore, use the contrast coefficients (0, 1, -1) in `glht()` in the `multcomp` package to test whether the actor and partner effects differ significantly.

Specify the contrast coefficients in a matrix. Each row represents one contrast.

```
> library(multcomp)
> diffhouse <- matrix(c(0, 1, -1), nrow=1, ncol=3)
> ct <- glht(apim4, linfct = diffhouse)
> summary(ct)
```

Test whether the contrast differs significantly from zero.

```
         Simultaneous Tests for General Linear Hypotheses

Fit: gls(model = sat ~ cleanhrs + cleanhrspartner, data = dat4, correla
tion = corCompSymm(form = ~1 |
    coupleid))

Linear Hypotheses:
       Estimate Std. Error z value Pr(>|z|)
1 == 0   -9.286      0.303  -30.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Actor effect – Partner effect = -9.29, $p < .001$.

The actor effect was -2.20 whereas the partner effect was 7.08. The partner effect was significantly larger than the actor effect by 9.29 points, $p < .001$.

Actor-Partner Interdependence Model

- What happens if we collect data from only one roommate in each pair and estimate the actor effect without including the partner predictor?

Select one member from each dyad.

```
> dat4a <- dat4[!duplicated(dat4$coupleid),]
> out4a <- lm(sat ~ cleanhrs, data=dat4a)
> summary(out4a)

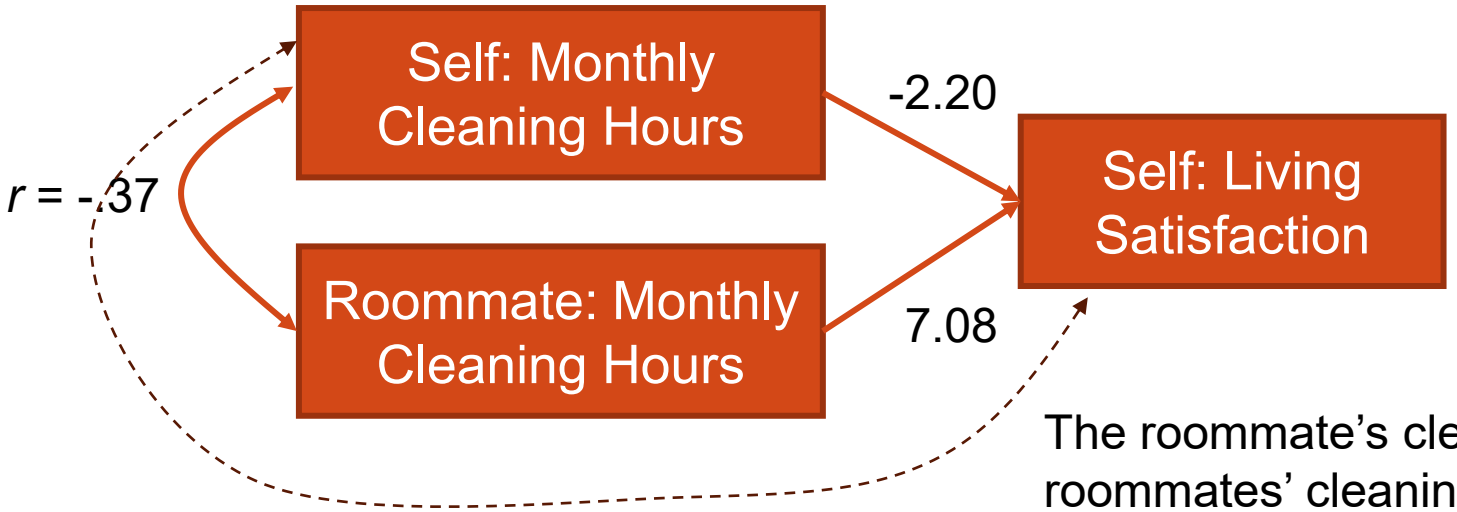
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   64.306      1.768   36.38  < 2e-16 ***
cleanhrs      -4.302      0.524   -8.21 1.91e-15 ***
```



The actor-effect estimate is biased. When the partner predictor is omitted, part of its association with the outcome may be attributed to the actor predictor. In this example, the estimated actor effect becomes more negative, changing from -2.20 to -4.30.

Actor-Partner Interdependence Model

- The direction of omitted-variable bias can be predicted from the partner effect and the correlation between actor and partner predictors.



The roommate's cleaning has a positive effect on satisfaction, but roommates' cleaning hours are negatively related. When the roommate's cleaning is omitted, some of that negative pattern is incorrectly attributed to the person's own cleaning.

Test of Distinguishability

- As discussed earlier, distinguishability can be evaluated by testing whether allowing model parameters to differ across roles improves model fit.

Assume that one roommate is listed first and the other second on the room-registration form. We examine whether this listing order meaningfully distinguishes the two roommates.

```
> dat4$listorder <- rep(c(1,0), 500)
> head(dat4)
  coupleid subjectid cleanhrs sat cleanhrspartner listorder
1        1         1      3.3  50            4.5         1
2        1         2      4.5  56            3.3         0
3        2         3      2.0  59            3.5         1
4        2         4      3.5  36            2.0         0
5        3         5      3.0  61            3.9         1
6        3         6      3.9  59            3.0         0
```

1 = listed first
0 = listed second

Indistinguishable Dyads

```
> apim4 <- gls(sat ~ cleanhrs + cleanhrspartner,
+              data = dat4,
+              correlation = corCompSymm(form = ~1|coupleid),
+              method="ML")
```

Allow the intercept and the actor and partner effects to differ by listing order

Distinguishable Dyads

```
> apim4b <- gls(sat ~ listorder*cleanhrs + listorder*cleanhrspartner,
+              data = dat4,
+              correlation = corCompSymm(form = ~1|coupleid),
+              weights = varIdent(form = ~1|listorder),
+              method="ML")
```

Fit both models using maximum likelihood when comparing models with different predictors.

Allow residual variances to differ by listing order

```
> apim4 <- gls(sat ~ cleanhrs + cleanhrspartner,
+              data = dat4,
+              correlation = corCompSymm(form = ~1|coupleid),
+              method="ML")
> summary(apim4b)
Generalized least squares fit by maximum likelihood
  Model: sat ~ listorder * cleanhrs + listorder * cleanhrspartner
  Data: dat4
       AIC      BIC    logLik
  7343.451 7387.621 -3662.726

Correlation Structure: Compound symmetry
 Formula: ~1 | coupleid
 Parameter estimate(s):
      Rho
0.3397006
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | listorder
 Parameter estimates:
        1         0
1.0000000 0.9936617
```

The estimated residual standard deviations were nearly identical across listing positions.

Listing order:
- main effect → not significant
- interaction effects → not significant

```
Coefficients:
                            Value Std.Error   t-value p-value
(Intercept)              34.82988 2.5655442 13.576022  0.0000
listorder                -1.36655 2.9576876 -0.462033  0.6442
cleanhrs                 -2.64097 0.4689311 -5.631889  0.0000
cleanhrspartner           7.28000 0.4693825 15.509733  0.0000
listorder:cleanhrs        0.88262 0.6224620  1.417942  0.1565
listorder:cleanhrspartner -0.39779 0.6224576 -0.639063  0.5229

 Correlation:
                          (Intr) lstrdr clenhrs clnhrsp listrdr:clnhrs
listorder                 -0.571
cleanhrs                  -0.819  0.468
cleanhrspartner           -0.812  0.463  0.369
listorder:cleanhrs         0.408 -0.709 -0.658  -0.020
listorder:cleanhrspartner  0.401 -0.709 -0.021  -0.659    0.034

Standardized residuals:
       Min         Q1        Med         Q3        Max
-3.1395726 -0.6693044 -0.0370902  0.6791248  2.6433791

Residual standard error: 9.753441
Degrees of freedom: 1000 total; 994 residual
```

```
> anova(apim4, apim4b)
       Model df      AIC      BIC    logLik   Test  L.Ratio p-value
apim4      1  5 7338.147 7362.686 -3664.074
apim4b     2  9 7343.451 7387.621 -3662.726 1 vs 2 2.696192  0.6099
```

The distinguishable-dyad model did not fit significantly better than the indistinguishable-dyad model, $\Delta\chi^2(4) = 2.70, p = .61$. The simpler model that treats the roommates as indistinguishable by listing order is therefore preferred.

Centering Parameterization

- The APIM models can be expressed as follows:
  - Distinguishable dyads

$$Y_{1j} = i_1 + a_1 X_{1j} + p_{2\to1} X_{2j} + e_{1j}$$
$$Y_{2j} = i_2 + a_2 X_{2j} + p_{1\to2} X_{1j} + e_{2j}$$

$$\begin{bmatrix} e_{1j} \\ e_{2j} \end{bmatrix} \overset{\text{i.i.d.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$$

  - Indistinguishable dyads

$$Y_{Aj} = i + aX_{Aj} + pX_{Pj} + e_{Aj}$$
$$Y_{Pj} = i + aX_{Pj} + pX_{Aj} + e_{Pj}$$

$$\begin{bmatrix} e_{Aj} \\ e_{Pj} \end{bmatrix} \overset{\text{i.i.d.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \sigma_{AP} \\ \sigma_{AP} & \sigma^2 \end{bmatrix} \right)$$

A and *P* arbitrarily identify the focal person and partner within dyad *j*. Each member serves once as the focal person.

Centering Parameterization

- An alternative parameterization transforms the two members' predictor scores into the dyad mean and the half-difference score:
  - Distinguishable dyads

$$\bar{X}_j = (X_{1j} + X_{2j})/2 \qquad Y_{1j} = i_1 + g_1\bar{X}_j + f_1 d_j + e_{1j}$$
$$d_j = (X_{1j} - X_{2j})/2 \qquad Y_{2j} = i_2 + g_2\bar{X}_j + f_2 d_j + e_{2j}$$

$$\begin{bmatrix} e_{1j} \\ e_{2j} \end{bmatrix} \overset{\text{i.i.d.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}\right)$$

It can be shown that:

$$a_1 = \frac{g_1 + f_1}{2} \qquad p_{2\to1} = \frac{g_1 - f_1}{2}$$
$$a_2 = \frac{g_2 - f_2}{2} \qquad p_{1\to2} = \frac{g_2 + f_2}{2}$$

The centering and APIM parameterizations are mathematically equivalent; they express the same model using different coefficients.

Centering Parameterization

$$Y_{1j} = i_1 + g_1\bar{X}_j + f_1 d_j + e_{1j}$$

$$Y_{2j} = i_2 + g_2\bar{X}_j + f_2 d_j + e_{2j}$$

- This parameterization separates the predictor into two components:
  - Dyad-mean effect. g_1 and g_2 indicate how the average predictor score within the dyad is associated with the outcomes of Role 1 and Role 2, respectively.
    - Example: How is the couple's average income associated with the marital satisfaction of the husband and the wife?
  - Difference effect. f_1 and f_2 indicate how the difference between the two members' predictor scores is associated with the outcomes of Role 1 and Role 2, respectively.
    - Example: How is the extent to which the husband's income exceeds the wife's income associated with the marital satisfaction of the husband and the wife?

```
> dat3 <- read.table("spousedistress.csv", sep=",", header=TRUE)
> dat3$aveneuro <- with(dat3, (menneuro + womenneuro)/2)
> dat3$diffneuro <- with(dat3, (menneuro - womenneuro)/2)
> library(lavaan)
> script3c <- '
+ mendistress ~ aveneuro + diffneuro + lengthrela
+ womendistress ~ aveneuro + diffneuro + lengthrela
+ mendistress ~~ womendistress
+ '
> apim3c <- sem(script3c, data=dat3)
> summary(apim3c, standardize=TRUE, rsquare=TRUE)
```

Calculate the average and difference in neuroticism between roles

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  mendistress ~
    aveneuro          0.808    0.053   15.172    0.000    0.808    0.614
    diffneuro        -0.093    0.047   -1.972    0.049   -0.093   -0.076
    lengthrela       -1.850    0.224   -8.260    0.000   -1.850   -0.336
  womendistress ~
    aveneuro          0.376    0.064    5.875    0.000    0.376    0.301
    diffneuro        -0.206    0.057   -3.629    0.000   -0.206   -0.176
    lengthrela       -2.300    0.269   -8.544    0.000   -2.300   -0.440
```

Couple's average neuroticism

- Significantly positively associated with men's distress, $b = 0.81, p < .001$.
- Significantly positively associated with women's distress, $b = 0.38, p < .001$.

Men-women neuroticism half-difference

- Significantly negatively associated with men's distress, $b = -0.09, p = .049$.
- Significantly negatively associated with women's distress, $b = -0.21, p < .001$.

Potentially defined parameters:

- Role difference in the effect of couple's average neuroticism
- Role difference in the effect of men-women neuroticism half-difference

Centering Parameterization

- Indistinguishable dyads
  - APIM parameterization

$$Y_{Aj} = i + aX_{Aj} + pX_{Pj} + e_{Aj}$$
$$Y_{Pj} = i + aX_{Pj} + pX_{Aj} + e_{Pj}$$

$$\begin{bmatrix} e_{Aj} \\ e_{Pj} \end{bmatrix} \overset{\text{i.i.d.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \sigma_{AP} \\ \sigma_{AP} & \sigma^2 \end{bmatrix}\right)$$

  - Centering parameterization

$$\bar{X}_j = (X_{Aj} + X_{Pj})/2$$
$$d_{Aj} = X_{Aj} - \bar{X}_j$$
$$d_{Pj} = X_{Pj} - \bar{X}_j$$
$$d_{Pj} = -d_{Aj}$$

$$Y_{Aj} = i + g\bar{X}_j + f\,d_{Aj} + e_{Aj}$$
$$Y_{Pj} = i + g\bar{X}_j + f\,d_{Pj} + e_{Pj}$$

$$\begin{bmatrix} e_{Aj} \\ e_{Pj} \end{bmatrix} \overset{\text{i.i.d.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \sigma_{AP} \\ \sigma_{AP} & \sigma^2 \end{bmatrix}\right)$$

It can be shown that: $a = \frac{g + f}{2}$ $p = \frac{g - f}{2}$

The two parameterizations are mathematical equivalent.

Centering Parameterization

$$Y_{Aj} = i + g\bar{X}_j + f\,d_{Aj} + e_{Aj}$$
$$Y_{Pj} = i + g\bar{X}_j + f\,d_{Pj} + e_{Pj}$$

- The parameterization separates predictor into two components:
  - Dyad-mean effect. g indicates how the average predictor score within the dyad is associated with each member's outcome.
    - Example: How is the average SES of the two siblings associated with each sibling's jealousy toward the other?
  - Within-dyad deviation effect. f indicates how being higher or lower than the dyad average is associated with the outcome.
    - Example: How is a sibling's SES relative to the average SES of the sibling pair associated with that sibling's jealousy toward the other?

```
> dat4 <- read.table("roommatecleaning.csv", sep=",", header=TRUE)
> dat4$avecleanhrs <- ave(dat4$cleanhrs, dat4$coupleid)
> dat4$diffcleanhrs <- dat4$cleanhrs - dat4$avecleanhrs
```

Calculate the dyad mean and each roommate's deviation from the dyad mean in monthly cleaning hours.

```
> apim4c <- gls(sat ~ avecleanhrs + diffcleanhrs,
+               data = dat4,
+               correlation = corCompSymm(form = ~1|coupleid),
+               method="ML")
> summary(apim4c)
Generalized least squares fit by maximum likelihood
  Model: sat ~ avecleanhrs + diffcleanhrs
  Data: dat4
       AIC      BIC    logLik
  7338.147 7362.686 -3664.074

Correlation Structure: Compound symmetry
 Formula: ~1 | coupleid
 Parameter estimate(s):
      Rho
0.3412184
Coefficients:
                Value Std.Error    t-value p-value
(Intercept)  34.12828 2.1078779  16.190823       0
avecleanhrs   4.88046 0.6379030   7.650795       0
diffcleanhrs -9.28616 0.3031664 -30.630591       0

 Correlation:
             (Intr) avclnh
avecleanhrs  -0.986
diffcleanhrs  0.000  0.000

Standardized residuals:
       Min         Q1        Med         Q3        Max
-3.0856783 -0.6745441 -0.0288497  0.6909592  2.6690375

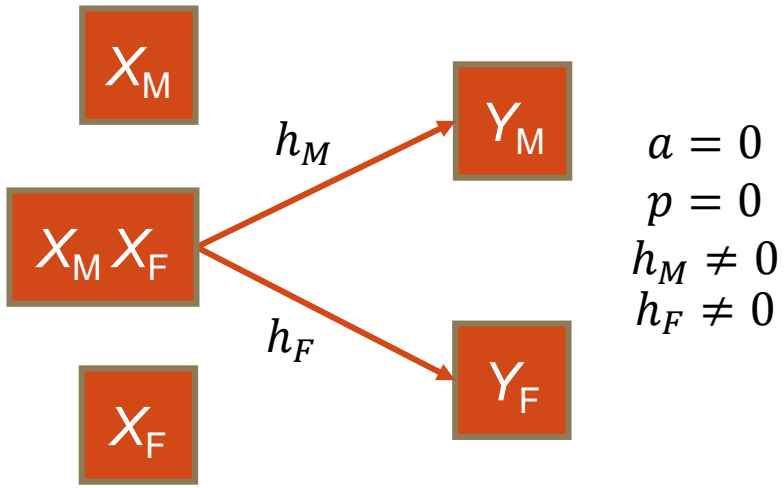
Residual standard error: 9.738443
Degrees of freedom: 1000 total; 997 residual
```

The roommate pair's average monthly cleaning hours were significantly positively associated with living satisfaction, $b = 4.88, p < .001$.
For each one-hour increase in the roommates' average monthly cleaning time, predicted living satisfaction increased by 4.88 points.

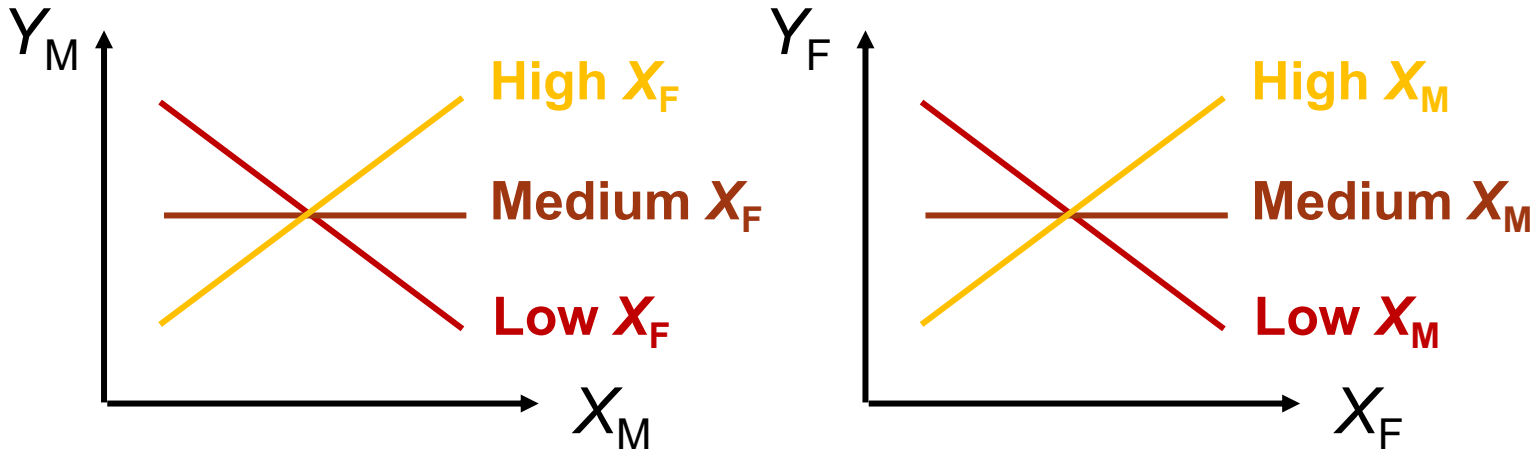
A roommate's deviation from the dyad's average monthly cleaning hours was significantly negatively associated with living satisfaction, $b = -9.29, p < .001$.
Holding the roommates' average cleaning time constant, a person who cleaned more than their roommate reported lower living satisfaction.

Interaction Effect

- The predictors from the two dyad members may interact in predicting each member's outcome.
  - Example: The association between a person's political ideology and marital satisfaction may depend on the partner's political ideology.



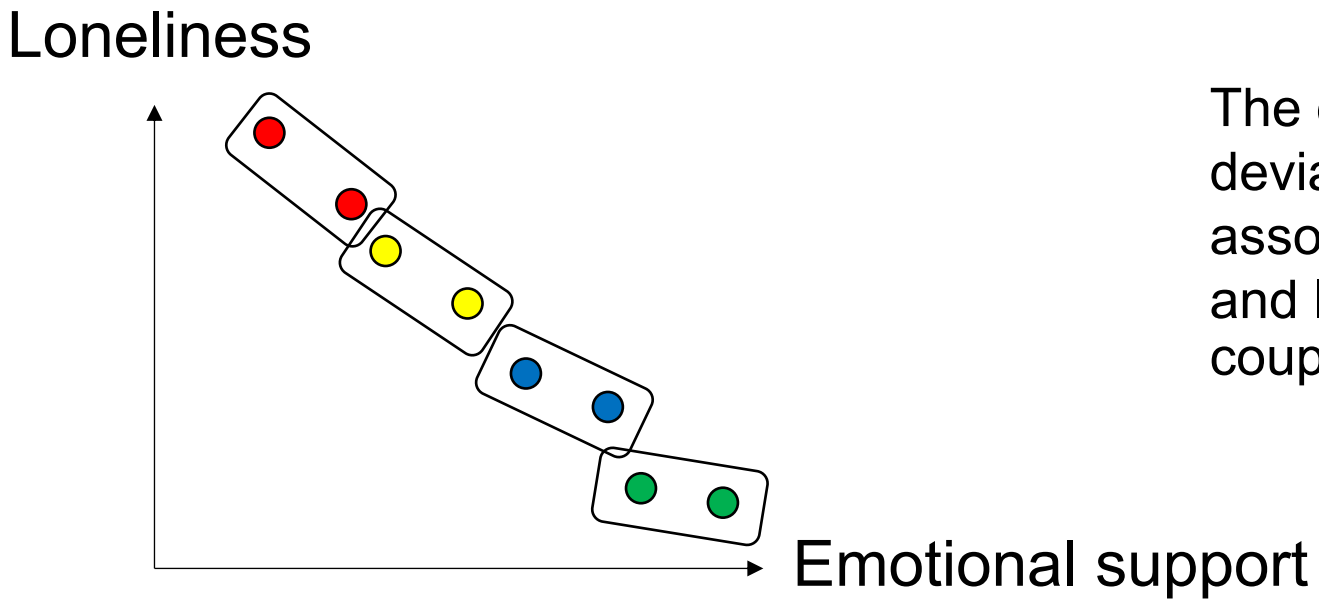
$a = 0$
$p = 0$
$h_M \neq 0$
$h_F \neq 0$



The actor effect varies across levels of the partner's predictor.

Interaction Effect

- In the centering parameterization, the dyad mean may interact with a member's deviation from the dyad mean in predicting that member's outcome.



The dyad mean moderates the within-dyad deviation effect: the negative within-couple association between emotional support and loneliness is stronger when the couple's average emotional support is low.

Interaction Effect

- The basic logic of the interaction effect is the same as in ordinary regression: an interaction is represented by the product of two predictors.
- I will focus on the interaction between predictors from different roles. Interactions between role and other predictors, including dyad-level predictors, are also possible.
- Distinguishable dyads:

$$Y_{1j} = i_1 + a_1 X_{1j} + p_{2\to1} X_{2j} + h_1 X_{1j} X_{2j} + e_{1j}$$

$$Y_{2j} = i_2 + p_{1\to2} X_{1j} + a_2 X_{2j} + h_2 X_{1j} X_{2j} + e_{2j}$$

$$Y_{1j} = i_1 + a_1 X_{1j} + p_{2\to1} X_{2j} + h_1 X_{1j} X_{2j} + e_{1j} \qquad Y_{2j} = i_2 + p_{1\to2} X_{1j} + a_2 X_{2j} + h_2 X_{1j} X_{2j} + e_{2j}$$

Find the effects of X_{1j} on Y_{1j} and Y_{2j} when X_{2j} is held constant at s:

$$Y_{1j} = (i_1 + p_{2\to1} s) + (a_1 + h_1 s) X_{1j} + e_{1j}$$

Conditional actor effect on Role 1

$$Y_{2j} = (i_2 + a_2 s) + (p_{1\to2} + h_2 s) X_{1j} + e_{2j}$$

Conditional partner effect on Role 2

Find the effects of X_{2j} on Y_{1j} and Y_{2j} when X_{1j} is held constant at m:

$$Y_{1j} = (i_1 + a_1 m) + (p_{2\to1} + h_1 m) X_{2j} + e_{1j}$$

Conditional partner effect on Role 1

$$Y_{2j} = (i_2 + p_{1\to2} m) + (a_2 + h_2 m) X_{2j} + e_{2j}$$

Conditional actor effect on Role 2

The interaction effects on the outcomes of Roles 1 and 2 are represented by h_1 and h_2, respectively.

Simple intercept

Simple slope

Interaction Effect

A blind-dating company measures each customer's extraversion, randomly matches men and women for blind dates, and asks each person how much they liked their date. The dataset contains 300 blind-date pairs.



Hypotheses

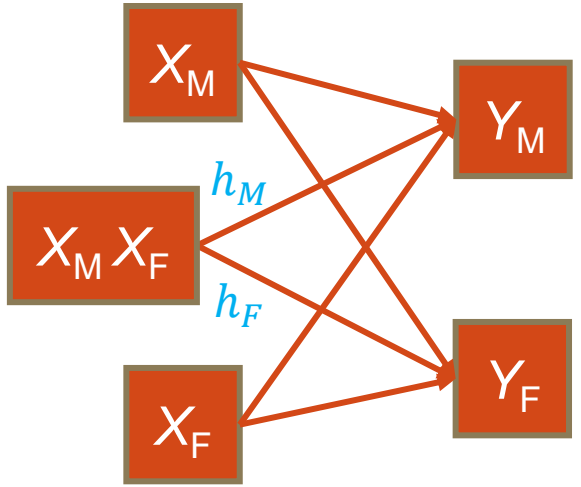
Actor Effect

- High partner extraversion → positive actor effect
- Low partner extraversion → negative actor effect

Partner Effect

- High actor extraversion → positive partner effect
- Low actor extraversion → negative partner effect

We expect the actor–partner interaction to be statistically significant for both men's and women's outcomes.



```
> dat5 <- read.csv("blinddate.csv", header=TRUE)
> library(psych)
> describe(dat5)
          vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
menext       1 300 50.47 10.23     50   50.56 11.86  22  79    57 -0.01    -0.34 0.59
womenext     2 300 51.09 10.07     52   51.37  8.90  17  76    59 -0.32     0.09 0.58
menlike      3 300 51.02 10.11     51   50.92 10.38  19  83    64  0.09     0.08 0.58
womenlike    4 300 50.91  8.88     51   50.88  8.90  23  75    52 -0.03    -0.12 0.51
```

- Because the raw extraversion scale has no naturally meaningful zero point or measurement unit, standardizing the scores makes the coefficients easier to interpret.
- The same mean and standard deviation should be used to standardize extraversion for both roles. Otherwise, a one-unit increase would represent different raw-score changes for men and women, making role comparisons difficult.
- Because the sample sizes are equal across roles, the pooled mean and pooled within-role standard deviation can be calculated as follows.

$$\bar{X}_{\text{Pooled}} = \frac{\bar{X}_M + \bar{X}_F}{2} = \frac{50.47 + 51.09}{2} = 50.78$$

$$SD_{\text{Pooled}} = \sqrt{\frac{SD_M^2 + SD_F^2}{2}} = \sqrt{\frac{(10.23)^2 + (10.07)^2}{2}} = 10.15$$

$$z_{X_M} = \frac{X_M - \bar{X}_{\text{Pooled}}}{SD_{\text{Pooled}}}$$

$$z_{X_F} = \frac{X_F - \bar{X}_{\text{Pooled}}}{SD_{\text{Pooled}}}$$

```
> meanmenext <- mean(dat5$menext)
> meanwomenext <- mean(dat5$womenext)
> pooledmean <- (meanmenext + meanwomenext)/2
>
> varmenext <- var(dat5$menext)
> varwomenext <- var(dat5$womenext)
> pooledsd <- sqrt((varmenext + varwomenext)/2)
>
> dat5$zmenext <- (dat5$menext - pooledmean)/pooledsd
> dat5$zwomenext <- (dat5$womenext - pooledmean)/pooledsd
>
> dat5$intext <- dat5$zmenext*dat5$zwomenext
```

Calculate the pooled mean.

Calculate the pooled within-role standard deviation.

Standardize extraversion scores for both roles using the same pooled mean and standard deviation.

Create the interaction term by multiplying the two standardized extraversion scores.

```
> library(lavaan)
> script5 <- '
+ menlike ~ zmenext + zwomenext + intext
+ womenlike ~ zmenext + zwomenext + intext
+ menlike ~~ womenlike
+ '
> int5 <- sem(script5, data=dat5)
```

Regress each person's liking on men's extraversion, women's extraversion, and their interaction.

```
> summary(int5, rsquare=TRUE)

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  menlike ~
    zmenext           2.991    0.504    5.938    0.000
    zwomenext         0.404    0.514    0.787    0.431
    intext            4.000    0.498    8.035    0.000
  womenlike ~
    zmenext          -0.061    0.466   -0.132    0.895
    zwomenext         0.935    0.475    1.968    0.049
    intext            3.468    0.460    7.535    0.000
```

- The interaction effect between men's and women's extraversion significantly predicted men's liking, $b = 4.00, p < .001$.
- The interaction effect between men's and women's extraversion significantly predicted women's liking, $b = 3.47, p < .001$.

Women's Extraversion	Men Actor Effect on Men's Liking	Men Partner Effect on Women's Liking
-1	$2.99 - 4.00 = -1.01$	$-0.06 - 3.47 = -3.53$
0	2.99	-0.06
1	$2.99 + 4.00 = 6.99$	$-0.06 + 3.47 = 3.41$

Men's Extraversion	Women Actor Effect on Women's Liking	Women Partner Effect on Men's Liking
-1	$0.94 - 3.47 = -2.53$	$0.40 - 4.00 = -3.60$
0	0.94	0.40
1	$0.94 + 3.47 = 4.41$	$0.40 + 4.00 = 4.40$

- The results generally support the hypothesized crossover pattern. When the other person's extraversion is low, the conditional actor and partner effects are negative. When the other person's extraversion is high, the conditional actor and partner effects are positive.
- At the other person's average extraversion, the actor effects are positive whereas the partner effects are close to zero.

Defining Conditional Actor and Partner Effects

```
> script5p <- '
+ menlike ~ am*zmenext + pm*zwomenext + intm*intext
+ womenlike ~ pw*zmenext + aw*zwomenext + intw*intext
+ menlike ~~ womenlike
+
+ # Men's actor effect, conditional on women's extraversion
+ am_womenhigh := am + intm
+ am_womenmid  := am
+ am_womenlow  := am - intm
+
+ # Women's actor effect, conditional on men's extraversion
+ aw_menhigh := aw + intw
+ aw_menmid  := aw
+ aw_menlow  := aw - intw
+
+ # Women-to-men partner effect, conditional on men's extraversion
+ pm_menhigh := pm + intm
+ pm_menmid  := pm
+ pm_menlow  := pm - intm
+
+ # Men-to-women partner effect, conditional on women's extraversion
+ pw_womenhigh := pw + intw
+ pw_womenmid  := pw
+ pw_womenlow  := pw - intw
+ '
```

Define conditional actor and partner effects at low, average, and high extraversion.

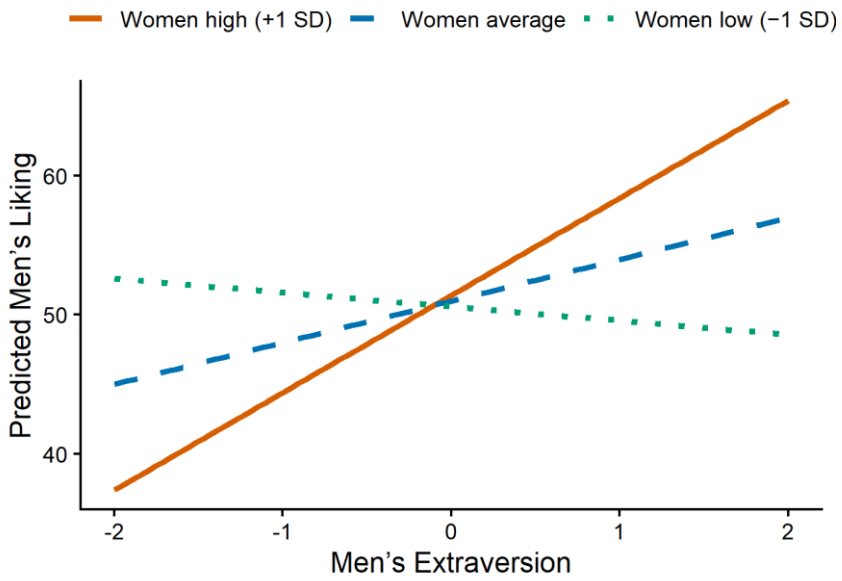
Defined Parameters:

	Estimate	Std.Err	z-value	P(>|z|)
am_womenhigh	6.990	0.718	9.729	0.000
am_womenmid	2.991	0.504	5.938	0.000
am_womenlow	-1.009	0.698	-1.447	0.148
aw_menhigh	4.402	0.691	6.367	0.000
aw_menmid	0.935	0.475	1.968	0.049
aw_menlow	-2.533	0.630	-4.023	0.000
pm_menhigh	4.404	0.748	5.888	0.000
pm_menmid	0.404	0.514	0.787	0.431
pm_menlow	-3.595	0.681	-5.278	0.000
pw_womenhigh	3.406	0.664	5.128	0.000
pw_womenmid	-0.061	0.466	-0.132	0.895
pw_womenlow	-3.529	0.645	-5.472	0.000

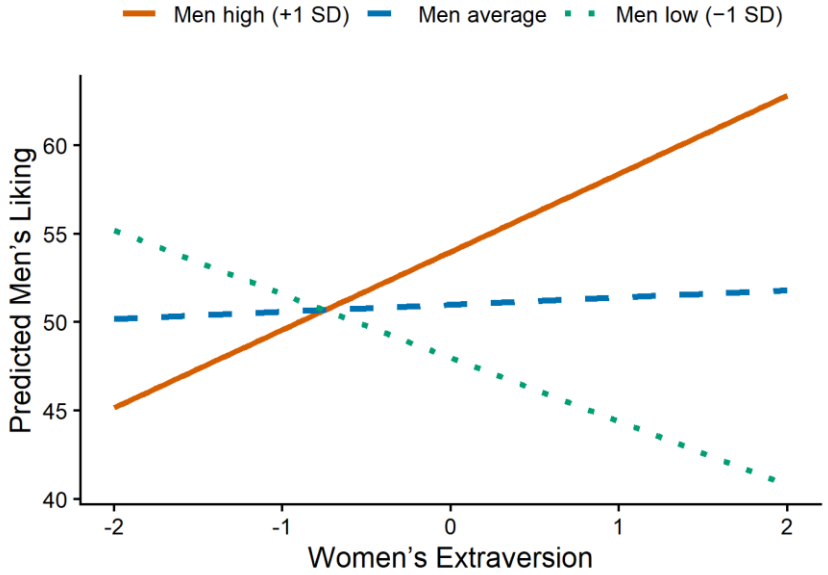
Women's Extraversion	Men Actor Effect on Men's Liking	Men Partner Effect on Women's Liking
-1	$-1.01, p = .15$	$-3.53, p < .001$
0	$2.99, p < .001$	$-0.06, p = .90$
1	$6.99, p < .001$	$3.41, p < .001$

Men's Extraversion	Women Actor Effect on Women's Liking	Women Partner Effect on Men's Liking
-1	$-2.53, p < .001$	$-3.60, p < .001$
0	$0.94, p = .049$	$0.40, p = .43$
1	$4.41, p < .001$	$4.40, p < .001$

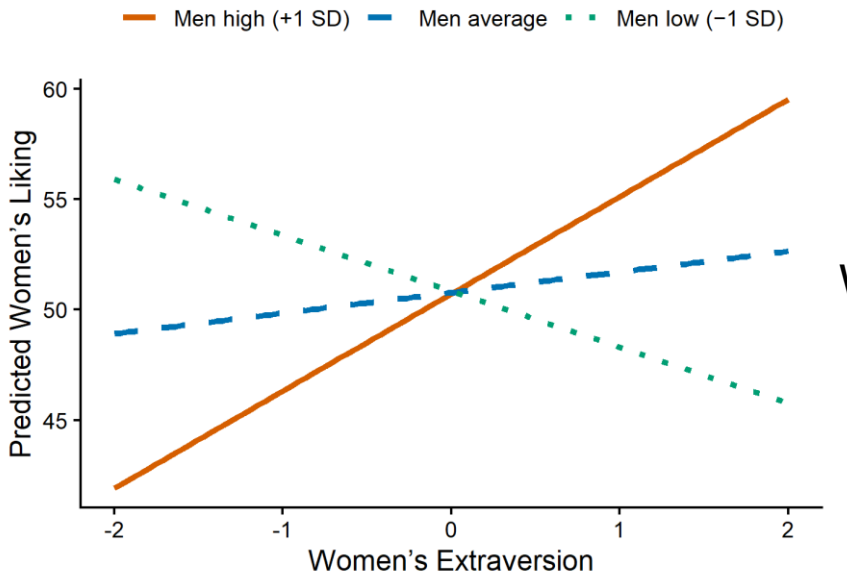
Please refer to the illustrative R code for creating interaction plots.



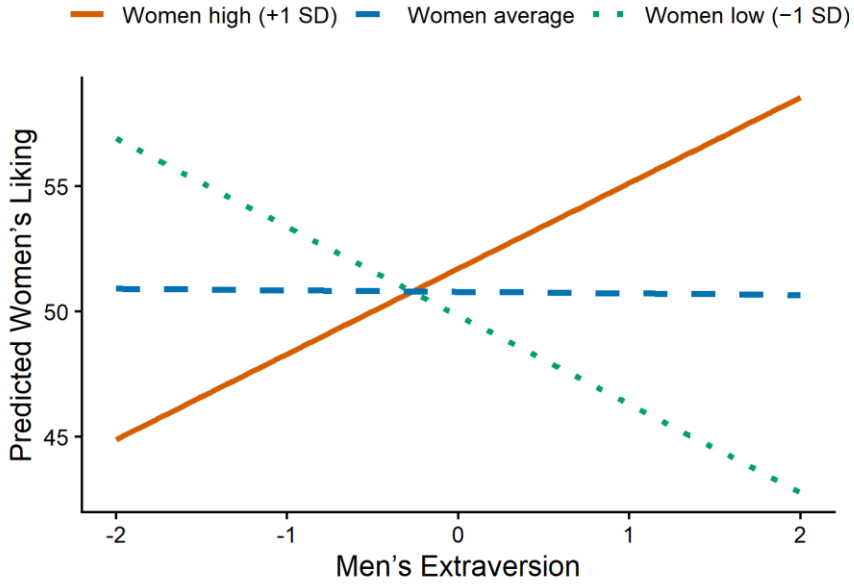
Men Actor Effect on Men's Liking



Women Partner Effect on Men's Liking



Women Actor Effect on Women's Liking



Men Partner Effect on Women's Liking

Interaction Effect

- Indistinguishable dyads:

$$Y_{Aj} = i + aX_{Aj} + pX_{Pj} + hX_{Aj}X_{Pj} + e_{Aj}$$

$$Y_{Pj} = i + aX_{Pj} + pX_{Aj} + hX_{Aj}X_{Pj} + e_{Pj}$$

Find the effects of X_{Aj} on Y_{Aj} when X_{Pj} is held constant at s:

$$Y_{Aj} = (i + ps) + (a + hs)X_{Aj} + e_{1j}$$ Conditional actor effect

The interaction effect on the outcomes is represented by h.

Find the effects of X_{Pj} on Y_{Aj} when X_{Aj} is held constant at m:

$$Y_{Aj} = (i + am) + (p + hm)X_{Pj} + e_{1j}$$ Conditional partner effect

Simple intercept

Simple slope

Interaction Effect

Data from 1,000 Roommates in 500 pairs. Examine whether a person's own monthly cleaning hours and the roommate's monthly cleaning hours predict that person's living satisfaction.



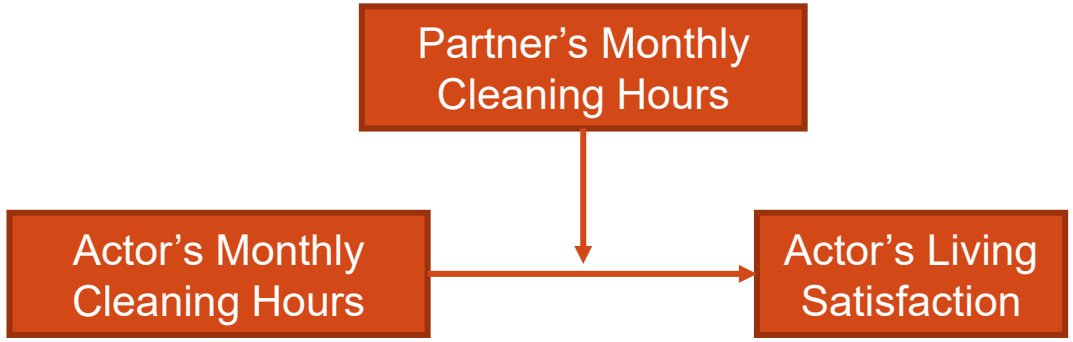
Hypotheses

Actor Effect

- When the partner cleans more, the negative actor effect will be weaker.
- When the partner cleans less, the negative actor effect will be stronger.

Partner Effect

- When the actor cleans more, the positive partner effect will be stronger.
- When the actor cleans less, the positive partner effect will be weaker.



```
> dat4 <- read.table("roommatecleaning.csv", sep=",", header=TRUE)
> dat4$avecleanhrs <- ave(dat4$cleanhrs, dat4$coupleid)
> dat4$diffcleanhrs <- dat4$cleanhrs - dat4$avecleanhrs
> dat4$cleanhrspartner <- dat4$avecleanhrs*2 - dat4$cleanhrs
```

Compute the average cleaning hours within each roommate pair, each person's deviation from the pair mean, and the partner's cleaning hours.

```
> quantile(dat4$cleanhrs)
  0%  25%  50%  75% 100%
 0.0  2.6  3.2  3.9  6.3
```

Because monthly cleaning hours are measured in meaningful units, we use the 25th, 50th, and 75th percentiles to represent low, medium, and high levels of cleaning, respectively.

As in `lm()`, the `*` operator in `gls()` specifies both the main effects and their interaction.

```
> library(nlme)
> out4i <- gls(sat ~ cleanhrs*cleanhrspartner,
+              data = dat4, correlation = corCompSymm(form = ~1|coupleid))
```

```
Coefficients:
                              Value Std.Error   t-value p-value
(Intercept)                45.53177  4.091386 11.128691  0.0000
cleanhrs                   -5.71393  1.138436 -5.019105  0.0000
cleanhrspartner             3.57223  1.138436  3.137843  0.0018
cleanhrs:cleanhrspartner    1.11993  0.345484  3.241623  0.0012
```

The interaction between the actor's and partner's cleaning hours was significant, $b = 1.12$, $p = .001$.

Partner's Cleaning Hours	Actor Effect on Actor's Living Satisfaction
2.6	$-5.71 + (2.6 \times 1.12) = -2.80$
3.2	$-5.71 + (3.2 \times 1.12) = -2.13$
3.9	$-5.71 + (3.9 \times 1.12) = -1.34$

Actor's Cleaning Hours	Partner Effect on Actor's Living Satisfaction
2.6	$3.57 + (2.6 \times 1.12) = 6.48$
3.2	$3.57 + (3.2 \times 1.12) = 7.15$
3.9	$3.57 + (3.9 \times 1.12) = 7.94$

As the partner's cleaning hours increased, the negative actor effect became weaker.
As the actor's cleaning hours increased, the positive partner effect became stronger.

```
Coefficients:
                            Value
(Intercept)              45.53177
cleanhrs                 -5.71393
cleanhrspartner           3.57223
cleanhrs:cleanhrspartner  1.11993
```

Use `glht()` from the `multcomp` package to test the simple slopes.

Each vector assigns weights to the intercept, actor effect, partner effect, and interaction effect, respectively.

Partner's Cleaning Hours	Actor Effect on Actor's Living Satisfaction
2.6	$-5.71 + (2.6 \times 1.12) = -2.80$
3.2	$-5.71 + (3.2 \times 1.12) = -2.13$
3.9	$-5.71 + (3.9 \times 1.12) = -1.34$

Actor's Cleaning Hours	Partner Effect on Actor's Living Satisfaction
2.6	$3.57 + (2.6 \times 1.12) = 6.48$
3.2	$3.57 + (3.2 \times 1.12) = 7.15$
3.9	$3.57 + (3.9 \times 1.12) = 7.94$

(0, 1, 0, 2.6)

(0, 1, 0, 3.2)

(0, 1, 0, 3.9)

(0, 0, 1, 2.6)

(0, 0, 1, 3.2)

(0, 0, 1, 3.9)

```
> library(multcomp)
> ct1 <- c(0, 1, 0, 2.6)
> ct2 <- c(0, 1, 0, 3.2)
> ct3 <- c(0, 1, 0, 3.9)
> ct4 <- c(0, 0, 1, 2.6)
> ct5 <- c(0, 0, 1, 3.2)
> ct6 <- c(0, 0, 1, 3.9)
> ctall <- rbind(ct1, ct2, ct3, ct4, ct5, ct6)
```

Use unadjusted tests when each simple slope is evaluated separately, as is common in simple-slopes analysis.

```
> ctsimslopes <- glht(out4i, linfct = ctall)
> summary(ctsimslopes, test=adjusted("none"))

         Simultaneous Tests for General Linear Hypotheses

Fit: gls(model = sat ~ cleanhrs * cleanhrspartner, data = dat4,
correlation = corCompSymm(form = ~1 |
    coupleid))

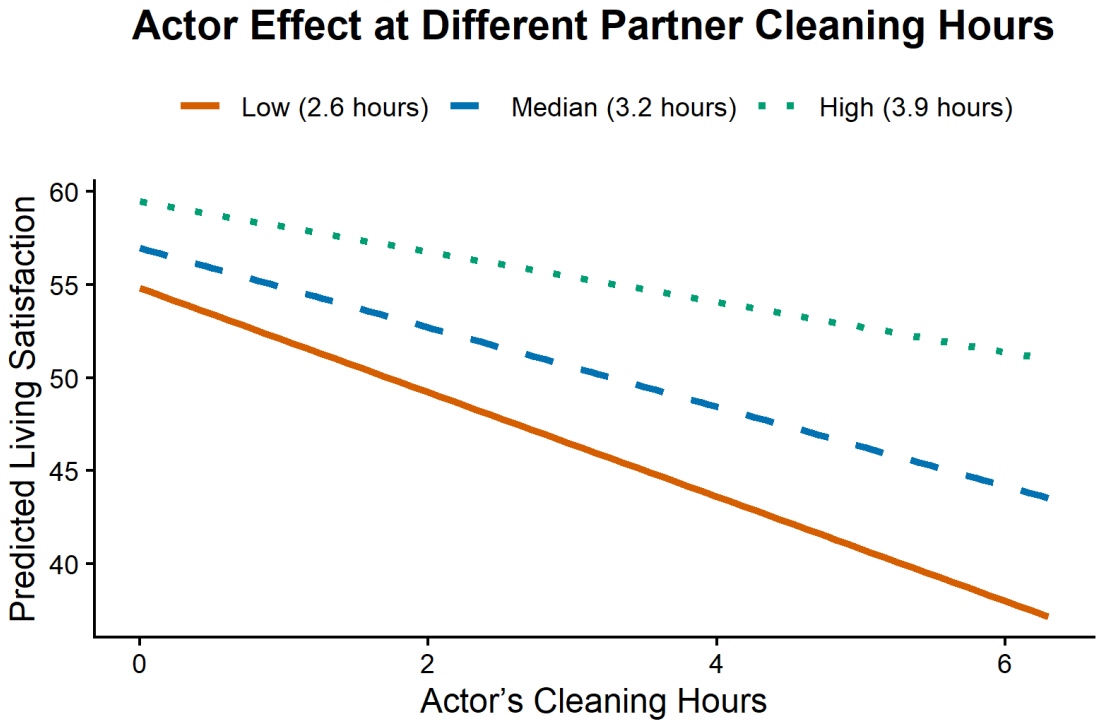
Linear Hypotheses:
         Estimate Std. Error z value Pr(>|z|)
ct1 == 0  -2.8021     0.3963  -7.071 1.54e-12 ***
ct2 == 0  -2.1302     0.3513  -6.064 1.32e-09 ***
ct3 == 0  -1.3462     0.4390  -3.067  0.00217 **
ct4 == 0   6.4840     0.3963  16.361  < 2e-16 ***
ct5 == 0   7.1560     0.3513  20.373  < 2e-16 ***
ct6 == 0   7.9400     0.4390  18.087  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- none method)
```

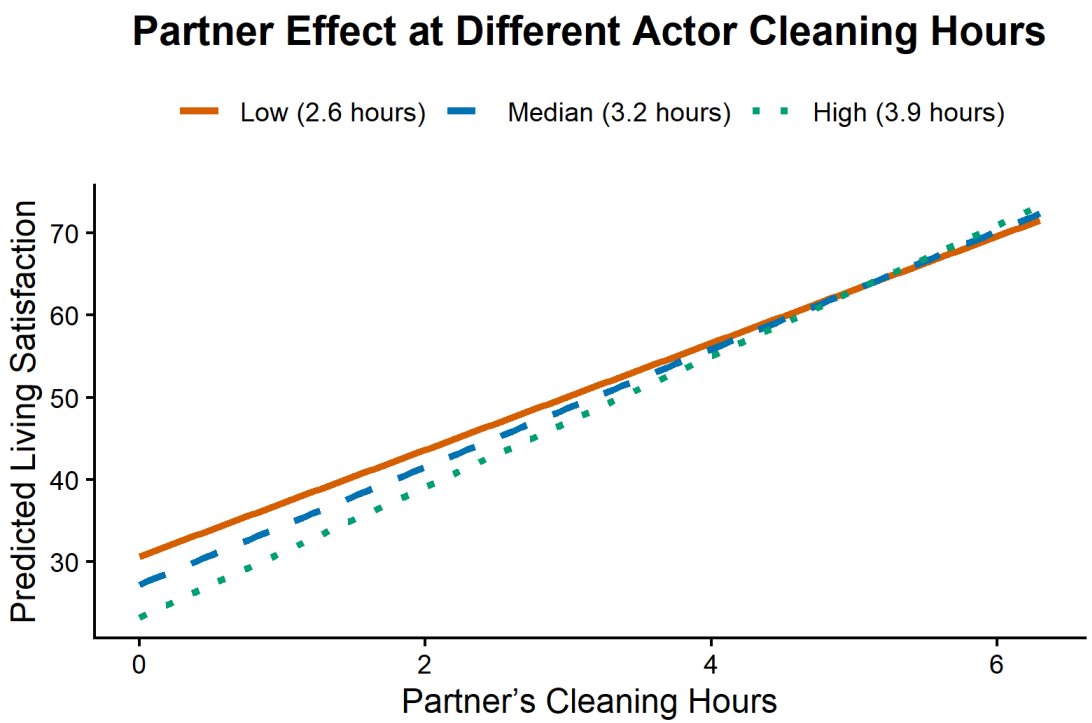
Partner's Cleaning Hours	Actor Effect on Actor's Living Satisfaction
2.6	$-2.80, p < .001$
3.2	$-2.13, p < .001$
3.9	$-1.34, p = .002$

Actor's Cleaning Hours	Partner Effect on Actor's Living Satisfaction
2.6	$6.48, p < .001$
3.2	$7.16, p < .001$
3.9	$7.94, p < .001$

- The actor effect remained significantly negative at all three levels of partner cleaning, but became weaker as partner cleaning increased.
- The partner effect remained significantly positive and became stronger as actor cleaning increased.

Please refer to the illustrative R code for creating interaction plots.





Interaction Effect

- Let us reparameterize the cleaning example using the centering approach and examine the interaction between dyad mean and the deviation score.

$$Y_{Aj} = i + g\bar{X}_j + f\,d_{Aj} + k\bar{X}_j d_{Aj} + e_{Aj}$$

$$Y_{Pj} = i + g\bar{X}_j + f\,d_{Pj} + k\bar{X}_j d_{Pj} + e_{Pj}$$

k represents how the deviation effect changes as the predictor dyad mean increases.

- Although the standard APIM and centering parameterizations are mathematically equivalent for models containing only main effects, an actor-by-partner interaction is not equivalent to a dyad-mean-by-deviation interaction.

```
> out4ic <- gls(sat ~ avecleanhrs*diffcleanhrs,
+               data = dat4, correlation = corCompSymm(form = ~1|coupleid))

Coefficients:
                             Value Std.Error   t-value p-value
(Intercept)              34.12828 2.1089358 16.182701  0.0000
avecleanhrs               4.88046 0.6382231  7.646957  0.0000
diffcleanhrs             -9.82165 1.8965177 -5.178782  0.0000
avecleanhrs:diffcleanhrs  0.16030 0.5604199  0.286034  0.7749
```

- The association between a person's deviation in cleaning hours from the dyad mean and living satisfaction was not significantly moderated by the dyad mean of cleaning hours, $b = 0.16, p = .77$.
- Thus, there was no evidence that the deviation effect varied across levels of the dyad mean. The deviation effect can therefore be summarized by a single coefficient using a model without the interaction term.

Proportion of Variance Explained

- In `lavaan`, specify `rsquare=TRUE` in `summary()` to obtain the proportions of variance explained for in each role's outcome.
- For `gls()`, one simple descriptive pseudo-R^2 is the squared correlation between the observed and predicted outcome scores.
  - For a random-intercept model, I compared this value with the corresponding measure from the `r2mlm` package and obtained similar results.

```
> out4i <- gls(sat ~ cleanhrs*cleanhrspartner,
+              data = dat4, correlation = corCompSymm(form = ~1|coupleid))
> cor(predict(out4i), dat4$sat)^2
[1] 0.4194868
```

Together, the predictors accounted for approximately 42% of the observed variance in living satisfaction.