Path Analysis

SDS 390 Structural Equation Modeling

Monday Feb 11, 2019

Agenda for today

- Missing data overview
- Path analysis
  - Path analysis piece from lab 1

Missing Data

If you missed Brittney Bailey's excellent talk!

Types of Missing Data Patterns

1. MCAR - Missing completely at random
   - The missingness is totally random and you can treat your data as basically a smaller random sample
2. MAR – Missing at random
   - The missingness on some Y is related to another variable(s) **X** you have in your dataset
3. MNAR
   - The missingness on Y is related to THAT variable, Y
   - Big thumbs down 

Little's MCAR test

```r
library(BaylorEdPsych)
library(mvnmle)

LittleMCAR(mydata)
```

	procrast_1	procrast_2	procrast_3	procrast_4	procrast_5
1	NA	2	1	2	4
2	1	1	4	4	1
3	4	NA	1	5	5
4	NA	1	1	3	2
5	NA	4	3	5	1
6	NA	5	2	2	3
7	3	2	4	2	3
8	3	2	1	2	4
9	3	5	3	3	4
10	4	NA	3	2	3
11	4	3	1	4	4
12	4	2	2	4	3
13	2	NA	2	4	4
14	4	5	4	1	3
15	5	1	3	4	4
16	5	NA	1	1	4

Little's MCAR test

```
> LittleMCAR(mydata)
this could take a while$chi.square
[1] 10.40739

$df
[1] 8

$p.value
[1] 0.2375885

$missing.patterns
[1] 3

$amount.missing
                procrast_1 procrast_2 procrast_3 procrast_4 procrast_5
Number Missing        4.00       4.00          0          0          0
Percent Missing       0.25       0.25          0          0          0
```

Dealing with Missing Data

Value imputation methods - replacing the value

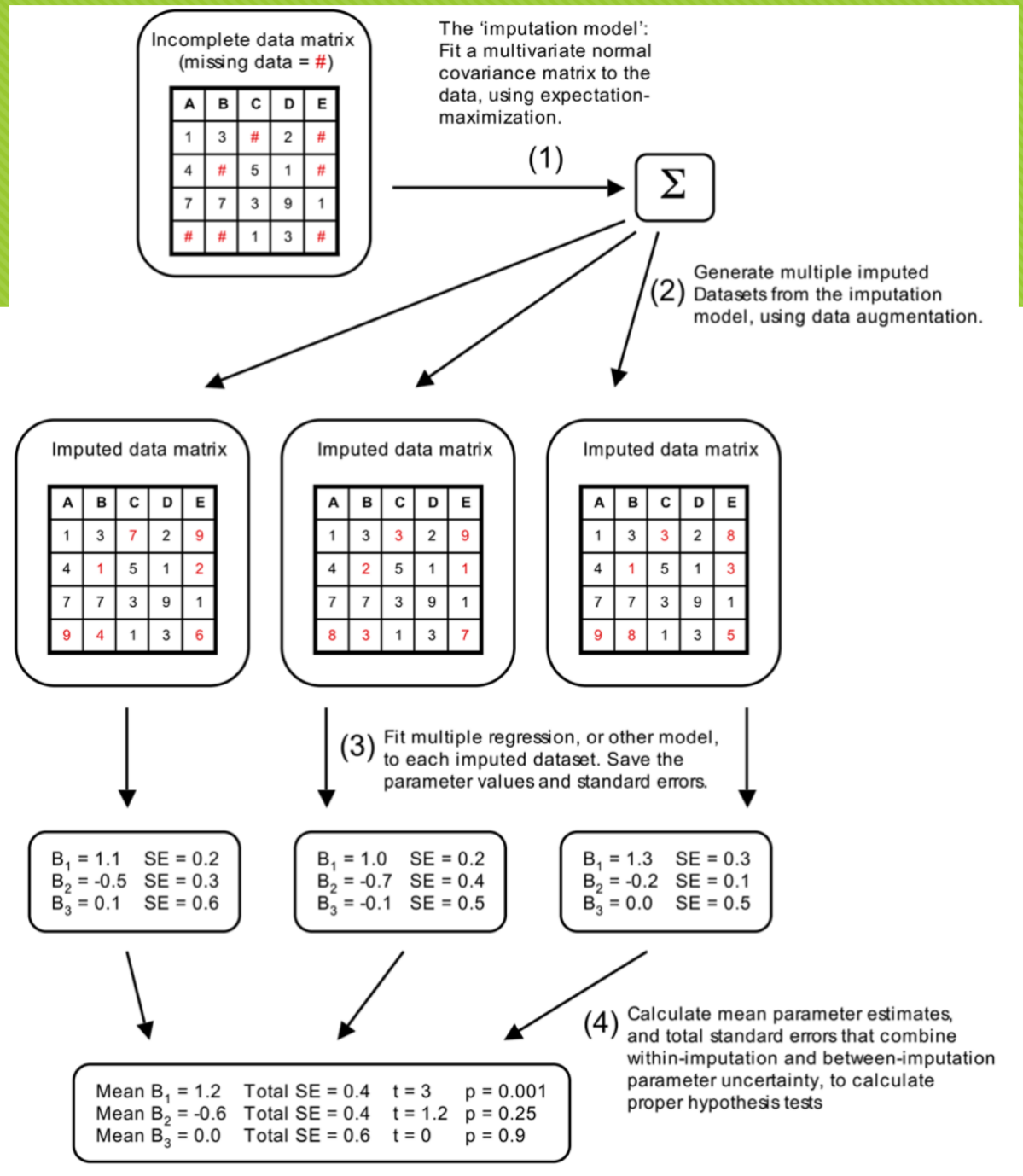
- Mean value imputation
- Last response carried forward
- Regression imputation
- Stochastic regression imputation
- Random hot-deck imputation
- Expectation-maximization (EM)

Dealing with Missing Data

- Single versus multiple imputation methods
  - See the mice package in R
- Estimation Methods
  - Full Information Maximum Likelihood (FIML)
  - More information on estimation later in the course

MICE package

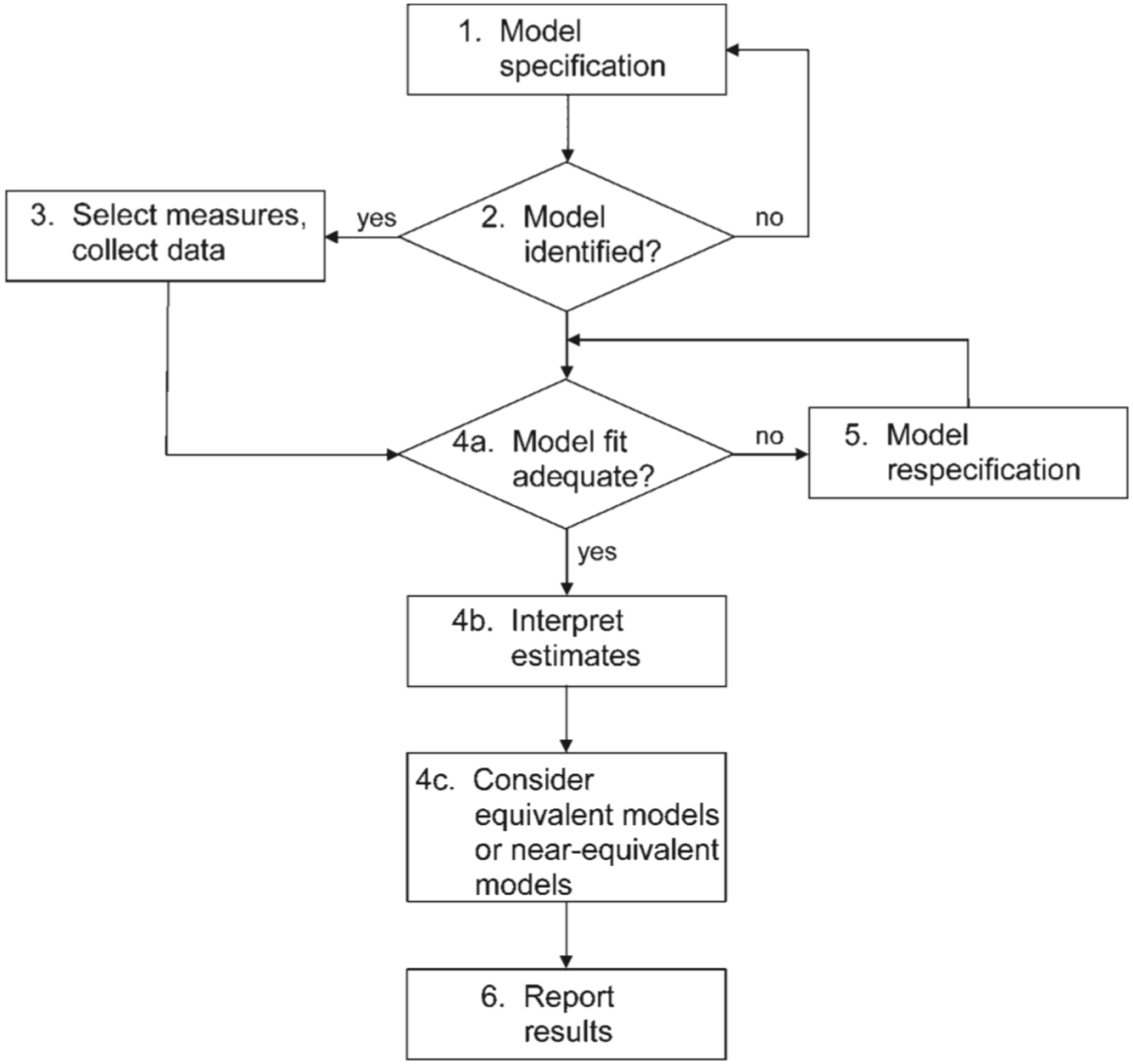
```
tempData <-
mice(data,m=5,maxit=50,meth='pmm',seed=500)
summary(tempData)

Multiply imputed data set
Call:
mice(data = data, m = 5, method = "pmm", maxit = 50,
seed = 500)
Number of multiple imputations:  5
Missing cells per column:
  Ozone Solar.R    Wind    Temp
     37       7       7       5
Imputation methods:
  Ozone Solar.R    Wind    Temp
  "pmm"   "pmm"   "pmm"   "pmm"
VisitSequence:
  Ozone Solar.R    Wind    Temp
      1       2       3       4
PredictorMatrix:
        Ozone Solar.R Wind Temp
Ozone       0       1    1    1
Solar.R     1       0    1    1
Wind        1       1    0    1
Temp        1       1    1    0
Random generator seed value:  500
```



Path Analysis

Six Steps in SEM

1. Specification
2. Model identification
3. Estimation
4. Re-specification



1. Specification

- We use diagrams to visualize the relationships between our variables. The relationships are also represented with equations.

- Pieces of model diagrams:
  - Observed variable in squares of rectangles
  - Latent variables with circles or ovals
  - Hypothesized directional effects with straight, single headed arrows
  - Covariances/correlations with curved, double headed arrows

- Variations on these pieces including triangles for means/intercepts, tiny arrows, etc.

1. Specification

Latent or Measured??

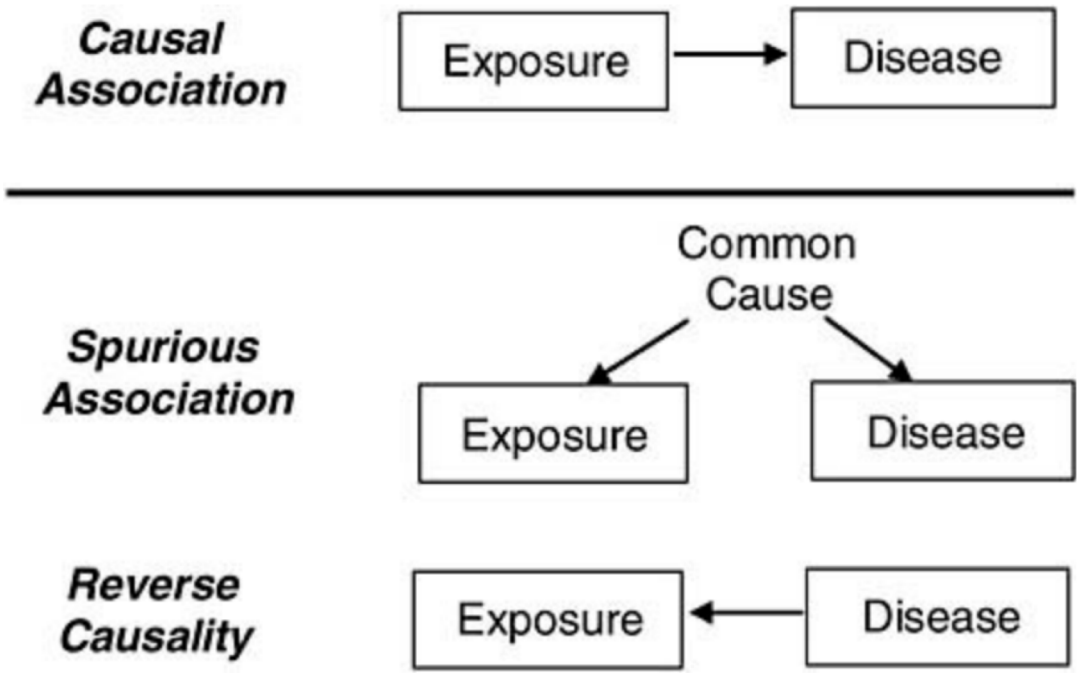
- Imagine you would like to use procrastination as a cause of anxiety.
- We know that procrastination is measured with 20 items.
- We two choices when specifying procrastination
  1. **Measured variable** by using the "scale score" of procrastination in the model
  2. **Latent variable** by representing all of the procrastination items as indicators of procrastination.

- Path Analysis is what we call a Structural Equation Model that only has measured variables and no latent variables.

1. Specification

- Directionality

- 5 general causality conditions
  1. Temporal precedence
  2. Association
  3. Isolation
  4. Correct effect priority
  5. Known distributional form



1. Specification

- Directionality

- 5 general causality conditions
  1. **Temporal precedence** – cause happens, measured before, effect
  2. **Association** – causation implies correlation!
  3. **Isolation** – no confounding variables
  4. **Correct effect priority** – correct causal direction
  5. **Known distributional form** – statistical assumptions hold when estimating the effect

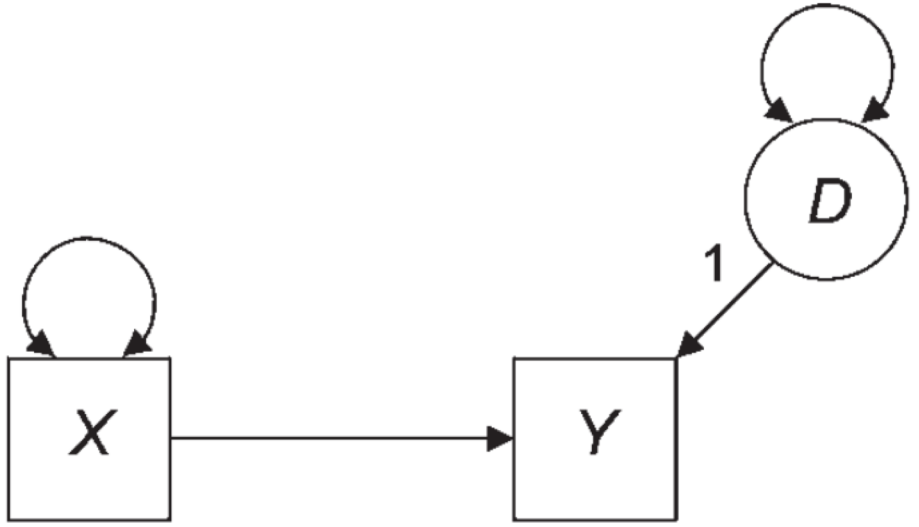
1. Specification

- **Endogenous variables** – other variables in the model are specified to cause them.
- **Exogenous variables** – there are no specified causes of these variables.

- Elemental Models
  1. Single cause
  2. Correlated causes
  3. Indirect effects (mediation model)

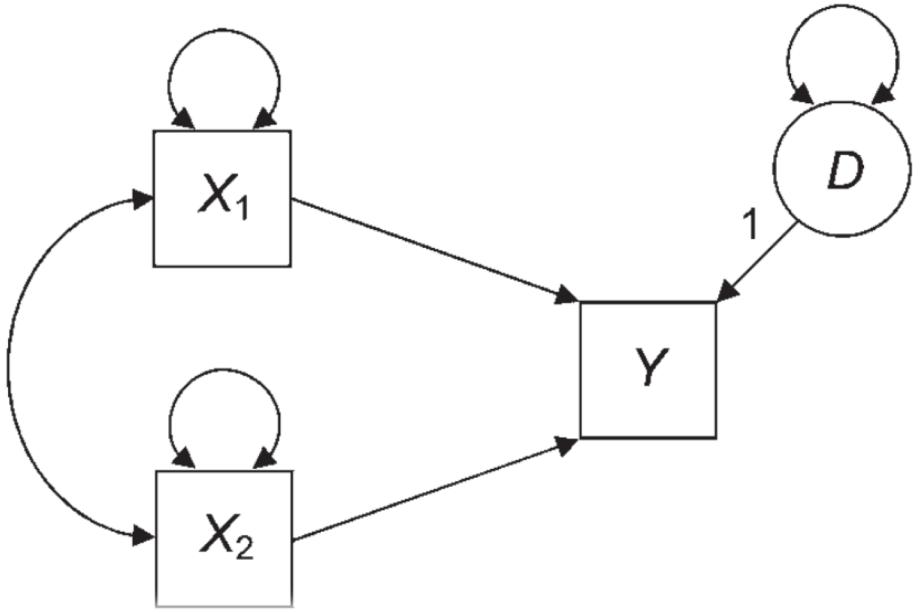
Single Cause

- Simple linear regression
  - Making causal assumptions



Correlated Causes

- Multiple regression
  - Making causal assumptions



Practice

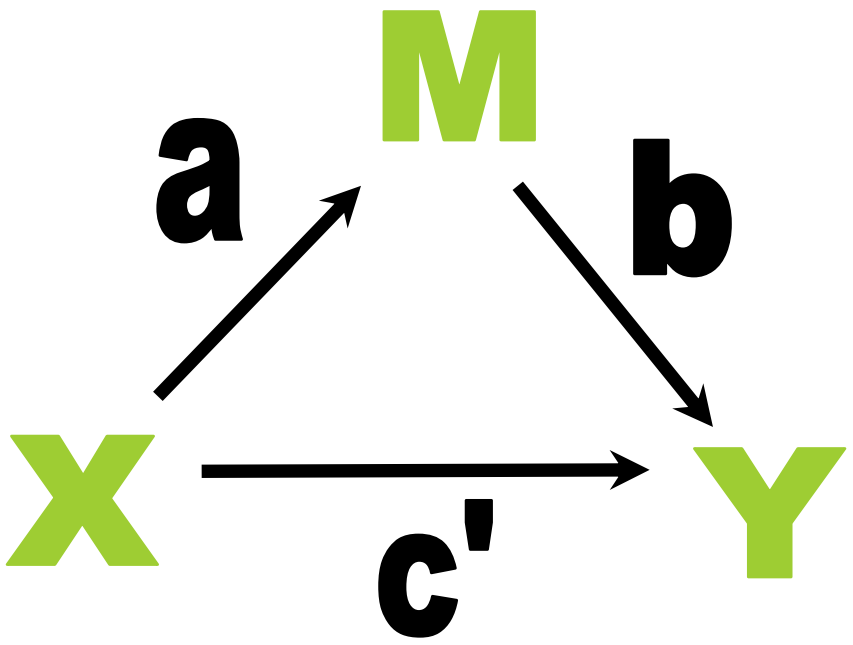
- With a partner, draw a SEM diagram for each scenario below, specifying the causal relationships

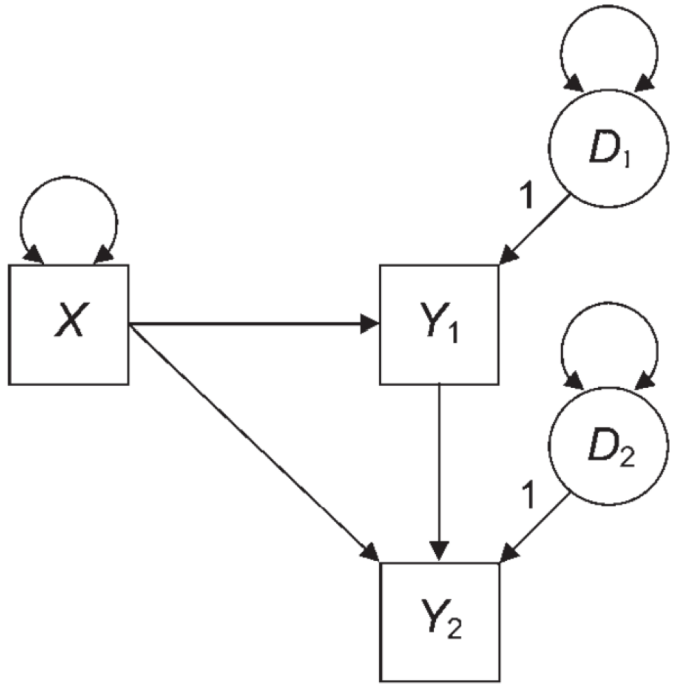
1. A study showed that women who work in the production of computer chips have abnormally high numbers of miscarriages. The union claimed that exposure to chemicals used in production caused the miscarriages. Another possible explanation is that these workers spend most of their work time standing up.

2. A study shows that there is a positive correlation between the size of a hospital (measured by its number of beds X) and the median number of days Y that patients remain in the hospital. Does this mean that you can shorten a hospital stay by choosing a small hospital? If not, think of an omitted variable to add to the model.

Indirect Effects and Basic Mediation

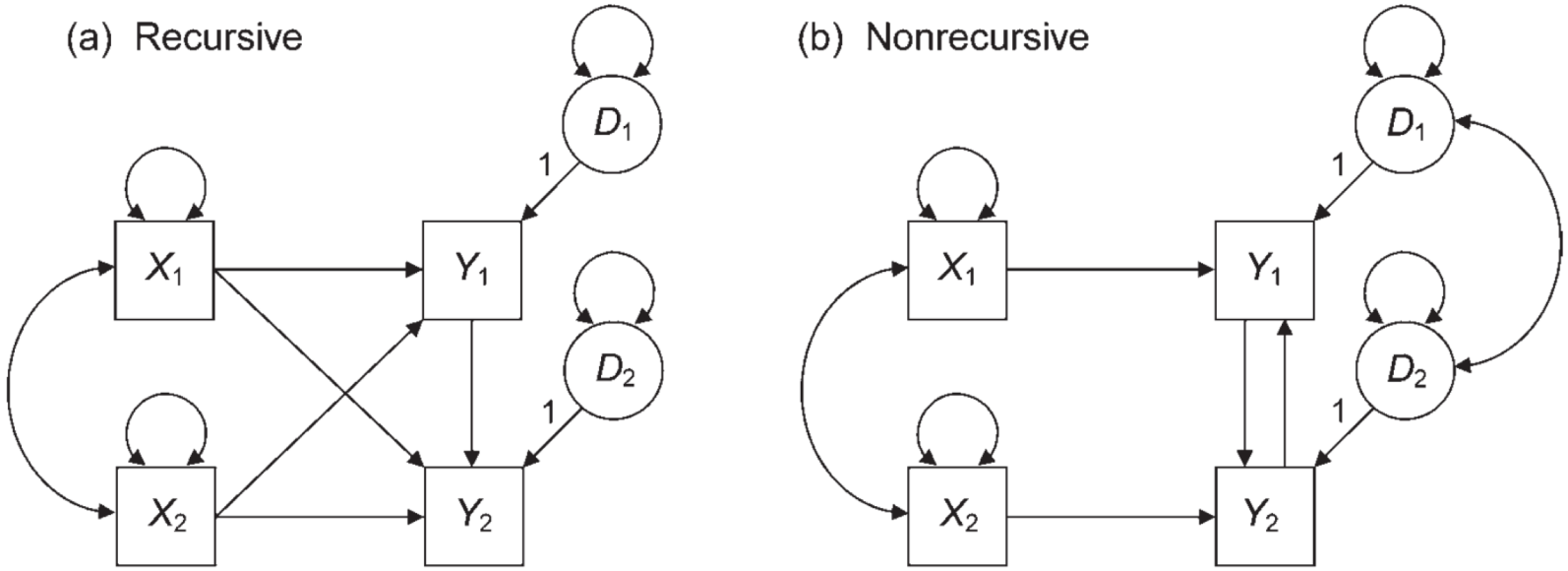
- Also called a mediation model





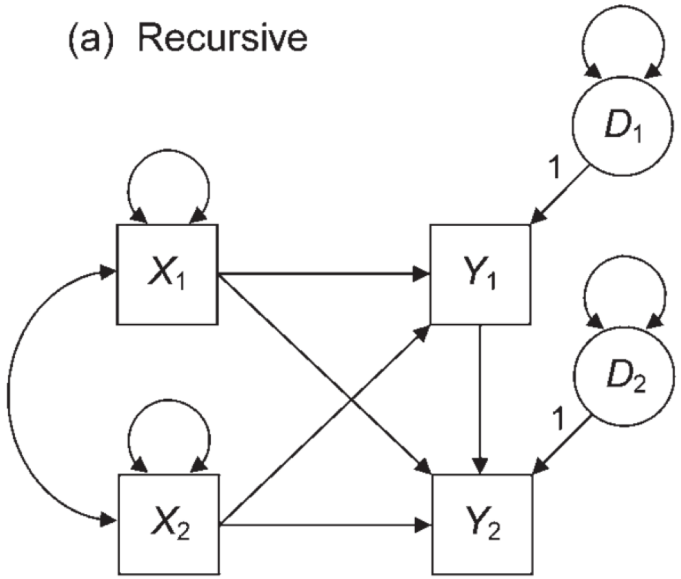
1. Specification

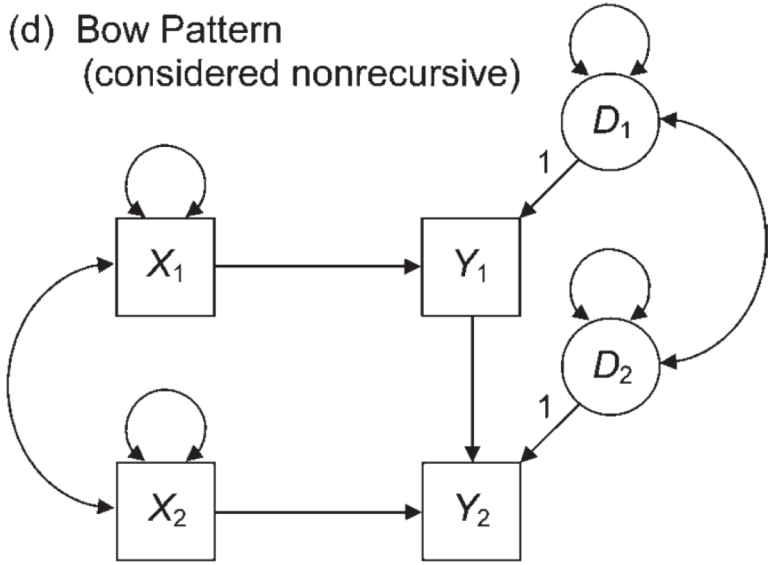
- Recursive versus non-recursive models
  - Recursive models flow neatly from left to right



1. Specification

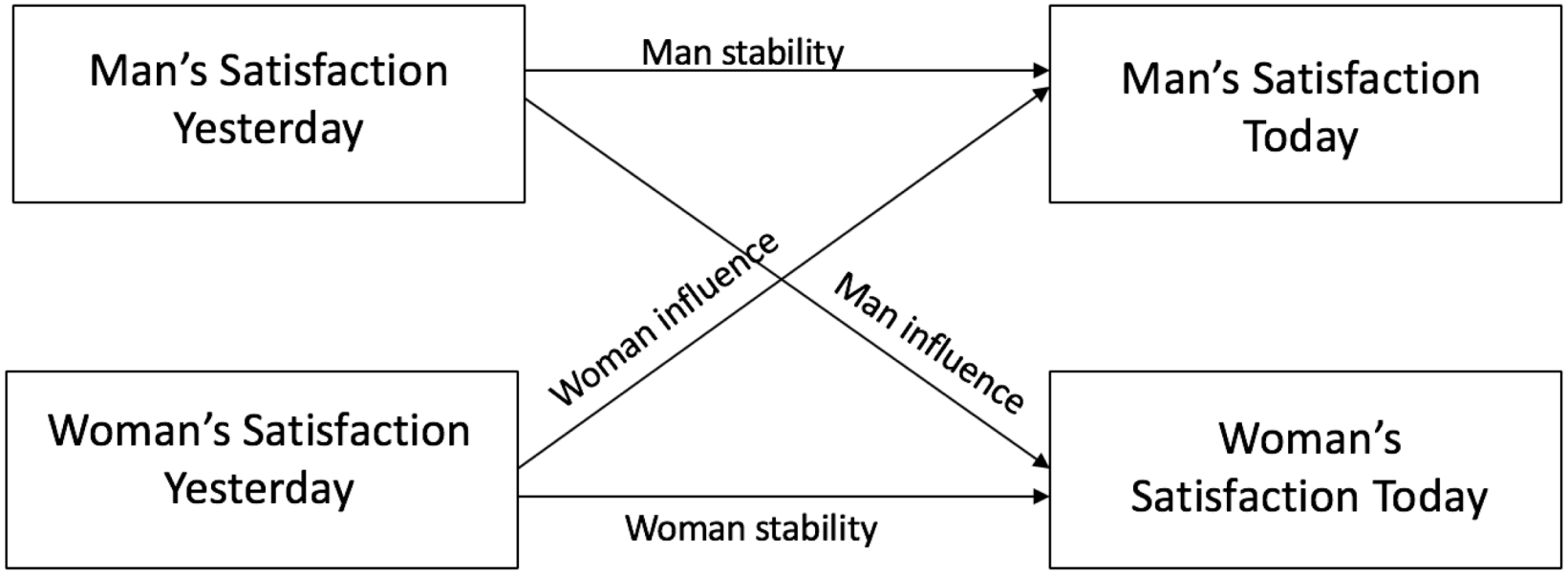
- Recursive versus non-recursive models
  - Recursive models flow neatly from left to right





1. Specification

- Panel Models
  - Stability influence model



2. Model Identification

- Number of observations for SEM is the lower diagonal elements in the covariance matrix plus the variances
- Model degrees of freedom = knowns - unknowns
  - Knowns = $\frac{v(v+1)}{2}$
  - Unknowns =
    - **Paths**
    - **Covariances between the exogenous variables, between the disturbances, and between exogenous variables and disturbances, and**
    - **Variances of the exogenous variables and disturbances of endogenous variables**
    - **Minus the number of linear constraints**

The lavaan package

Away to R! Please pull up Lab 1

Check out the lavaan tutorials