Structural Equation Modeling

Justin Wild
Indiana University

March 27, 2018

About the presenter

- Education Policy Studies & Inquiry Methodology doctoral candidate
  - a researcher
  - focus on methods & methodologies
  - not a programming expert
- Contributors
  - From IU's SSRC members
    - Scott Michael
    - Michael Frisby
- Inspirational
  - Yves Rosseel's keynote speech at *useR!* conference, 2017

Talk Overview

- **What is SEM?**
- SEM software packages
- SEM examples: running the software
- Wrapping up
- Slides there will be no time for

What is SEM? (Yves Rosseel, 2017)

- SEM = structural equation modeling
  - Multivariate statistical modeling technique
  - Test a model of the data
    - We propose a data-generating model
    - The model may or may not fit the data
- Assumptions (same as regression and most other assumptions of statistical techniques)
  - No outliers
  - Normality (but can be 'handled' e.g. robust estimation techniques)
  - Linearity
  - Identification (see slide 33)
- What is unique about SEM?
  - May contain latent (unobserved) variables
    - Can be hypothetical 'constructs' (e.g. reading ability, depression, etc.)
    - Can be random effects (e.g. intercepts, slopes)
    - …
  - Allows for indirect effects (mediation), reciprocal effects,…
  - Drawings!!!

SEM: regression example

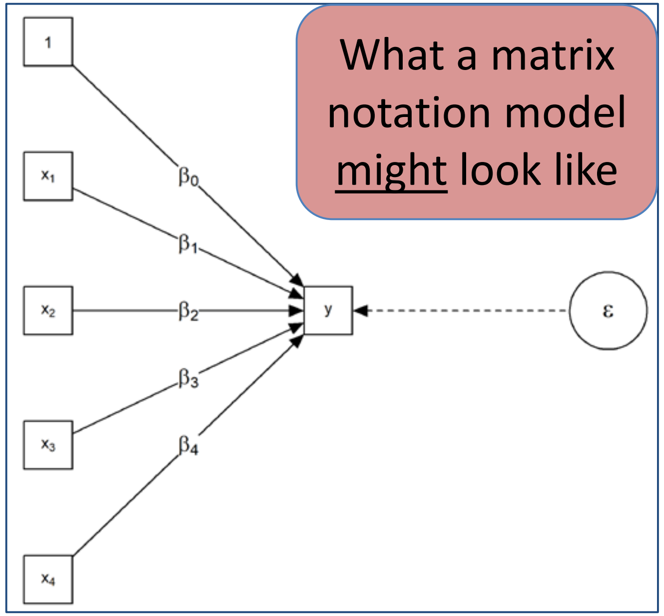
- Univariate linear regression

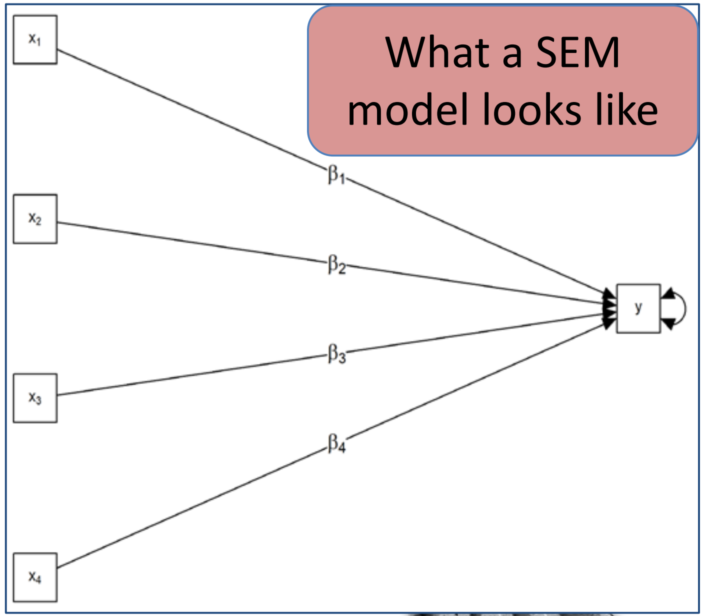
Equation: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$

Matrices: $\underset{n \times 1}{\boldsymbol{Y}} = \underset{n \times p'}{\boldsymbol{X}} \underset{p' \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\boldsymbol{e}}$

Matrix notation: $[y_1] = [1 \quad x_{11} \quad x_{12} \quad x_{13} \quad x_{14}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + [e_1]$

SEM matrix notation: $[y_1] = [\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + [e_1]$



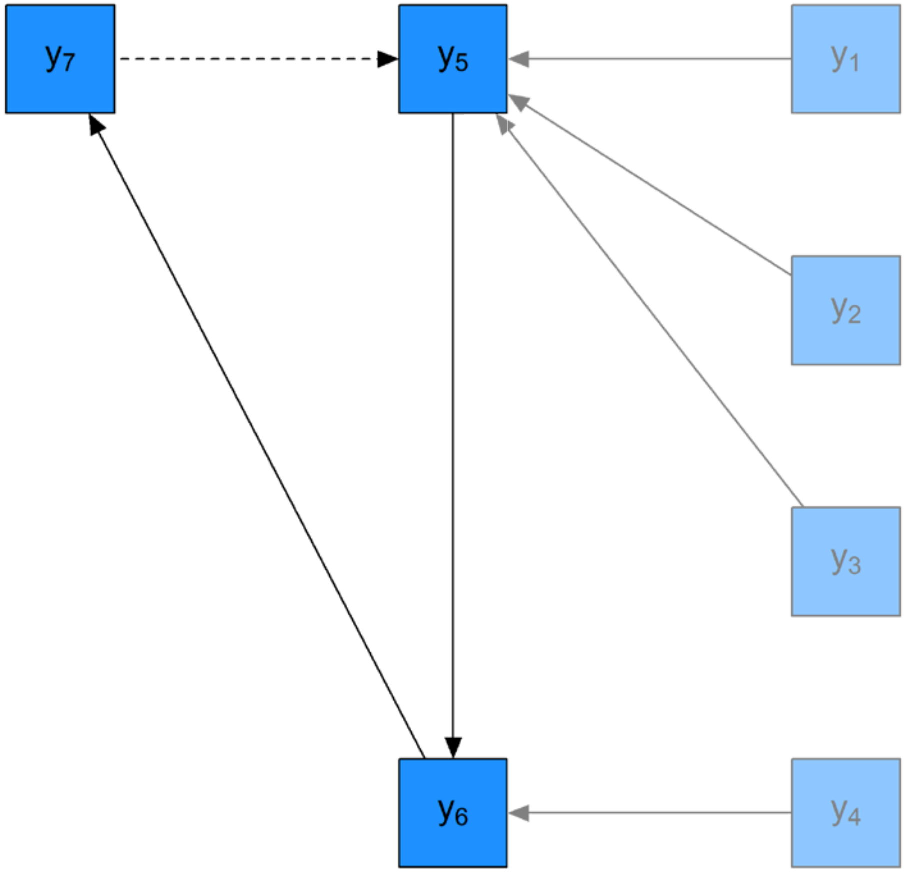


SEM: indirect effects & path analysis

- All variables are observed (manifest)
- Notice indirect effects: y_5 via y_6 on y_7
- Notice cycles: y_7 may influence y_5

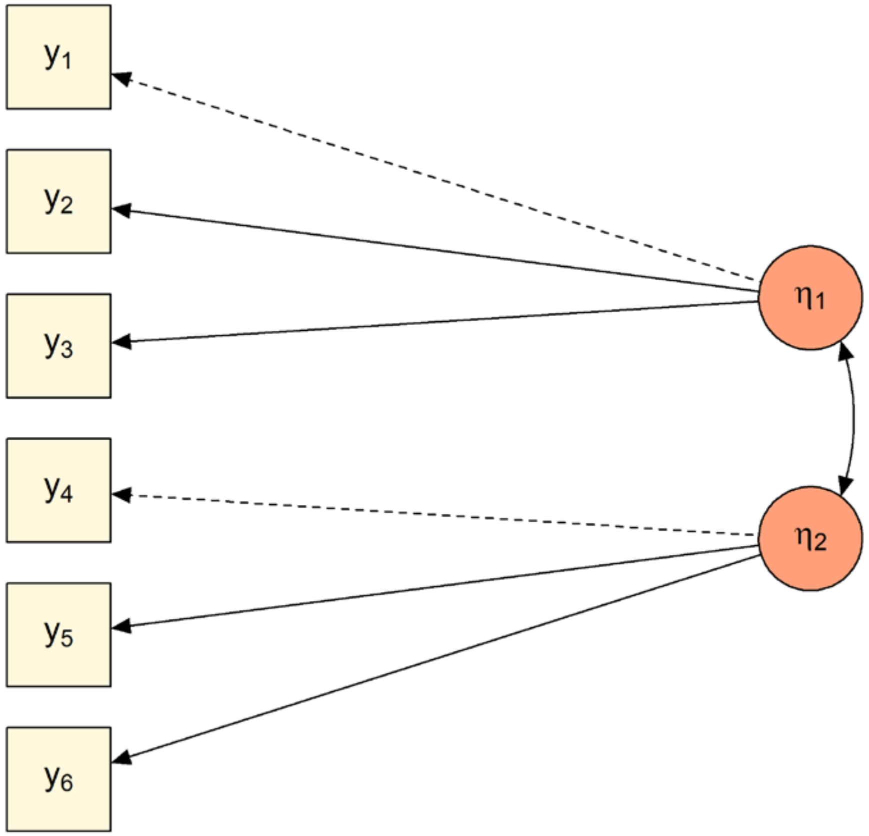
y_1 = read for school
y_2 = read with parent
y_3 = access to other media
y_4 = access to books

y_5 = reading motivation
y_6 = reading frequency
y_7 = reading ability



SEM: latent variables & path analysis

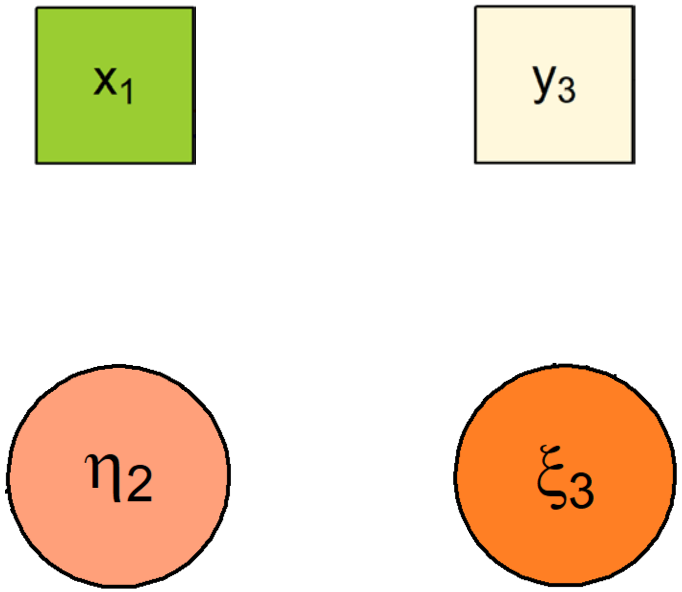
- Specify a measurement model of latent variables (η) of interest and their observed (manifest) indicators (y)
- Confirmatory Factor Analysis
  - Fit indices: statistics that provide information about how well the models fits the data
  - Modification indices: statistics that show relationships that improve the fit of your model
    - If we use this information to change the model there should be some theoretical justification



Some notes about the language of SEM

- Observed (manifest) variables
  - Roman letters: e.g., x_1 or y_3
  - Represented as boxes

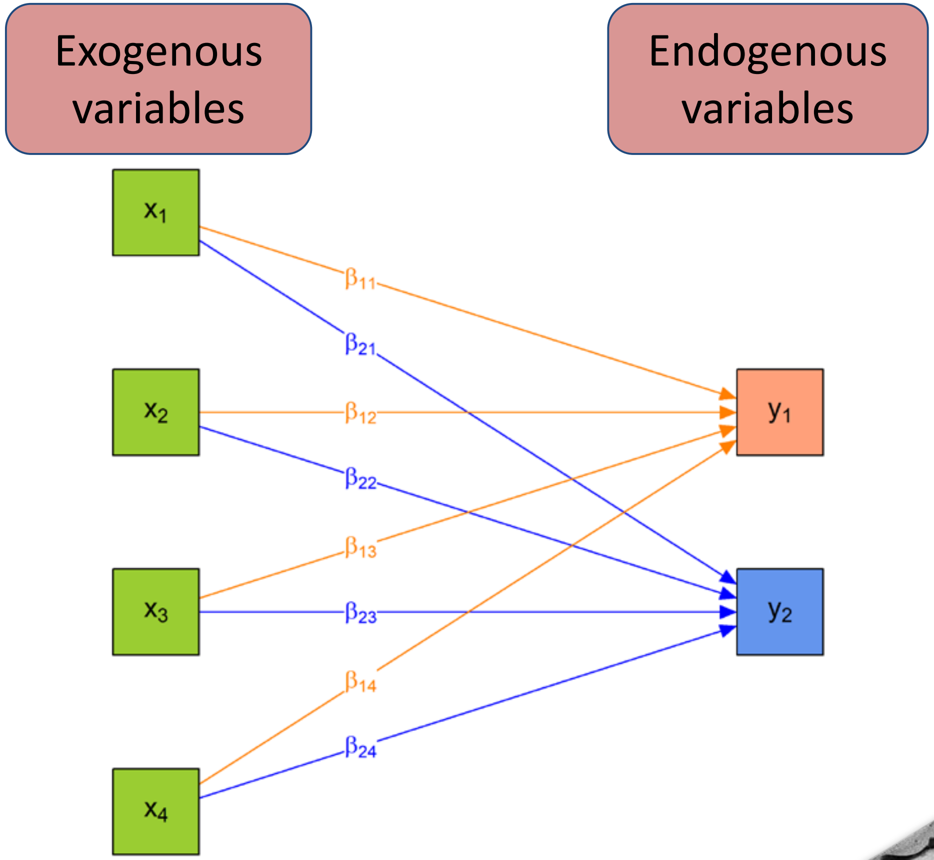
- Latent variables
  - Greek letters: e.g., η_2 or ξ_3
  - Represented as circles



Some notes about the language of SEM

- Exogenous variables
  - "of external origin"
  - Like an independent variable
  - No straight arrows pointing to it

- Endogenous variables
  - "of internal origin"
  - Like a dependent variable
  - Straight arrows point to it (and possibly out of it)

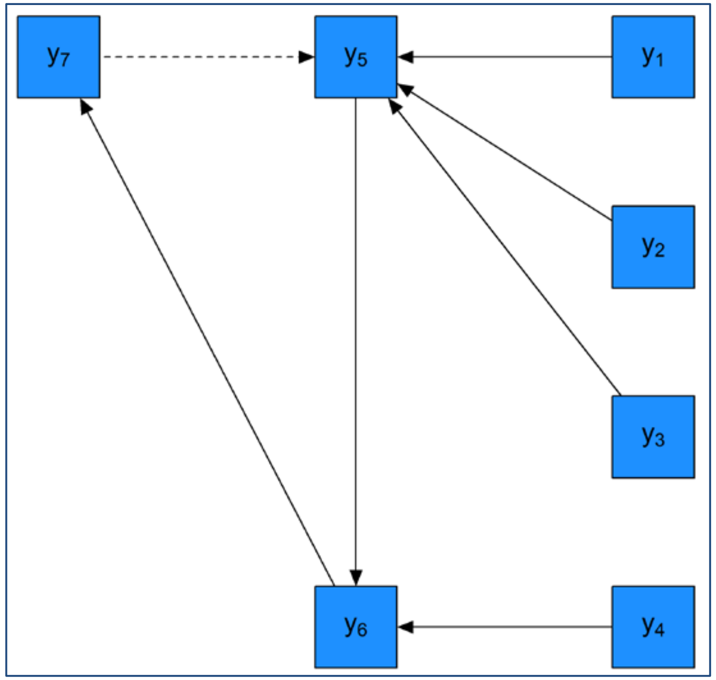


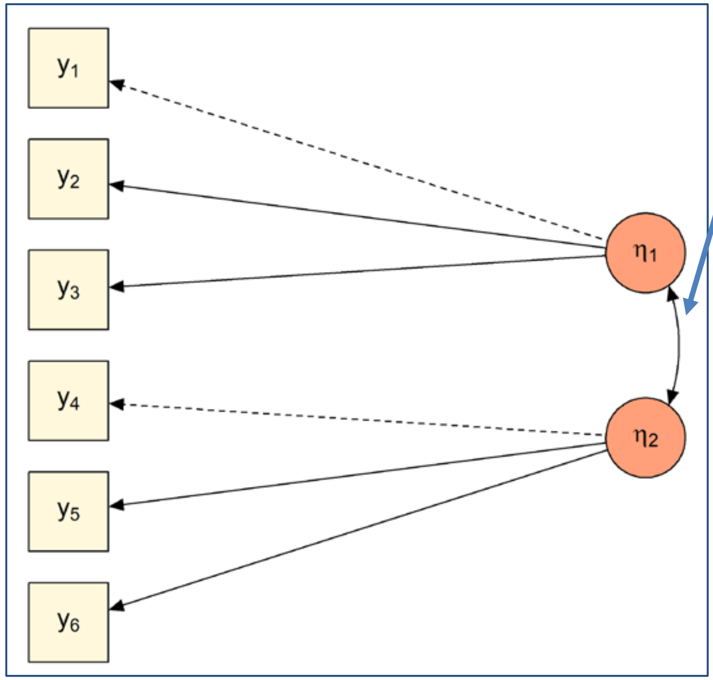
Some notes about the language of SEM

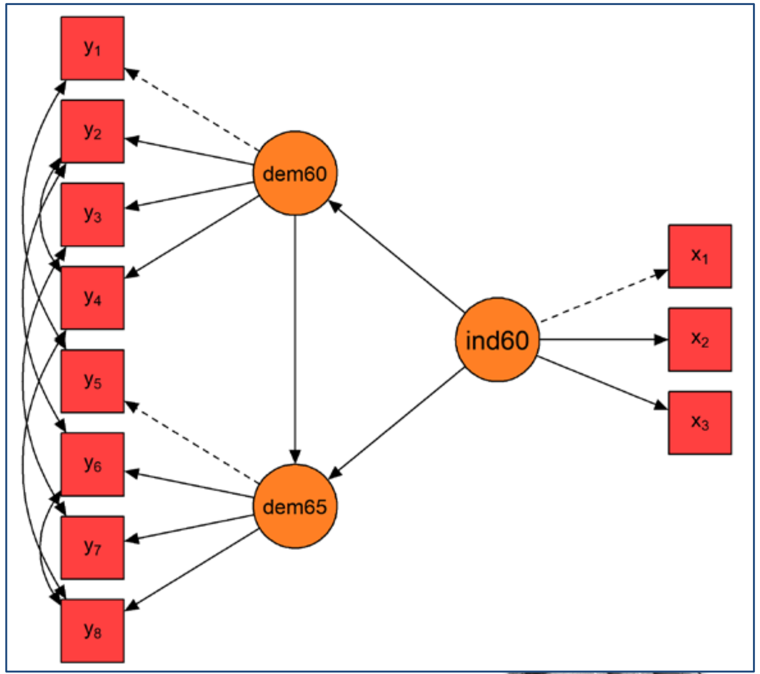
The phrase: “effect of (something) on (something else)”

e.g., “The effect of η_1 on y_1”

Curved arrows between variables indicate covariance







Example: Industrialization and Political Democracy (Bollen, 1989)

Expert ratings of the freedom of the press in 1960

The freedom of political opposition in 1960

The fairness of elections in 1960

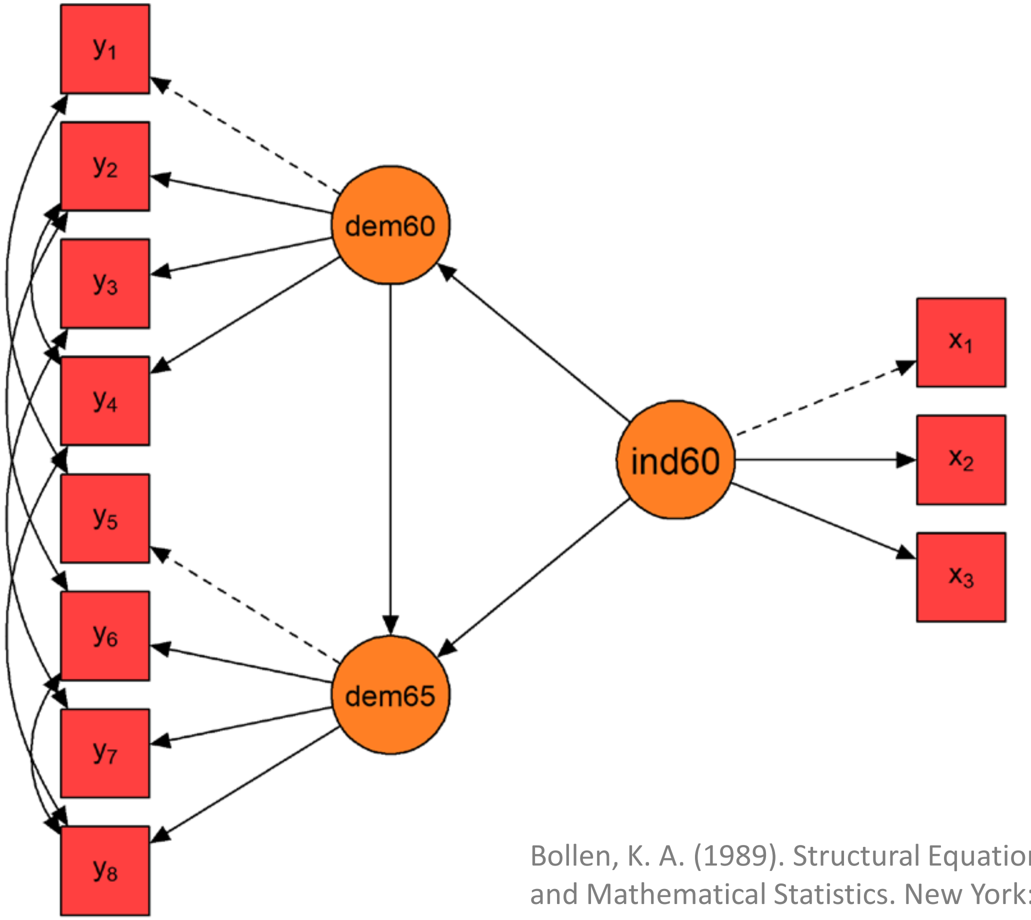
The effectiveness of the elected legislature in 1960

Expert ratings of the freedom of the press in 1965

The freedom of political opposition in 1965

The fairness of elections in 1965

The effectiveness of the elected legislature in 1965



The gross national product (GNP) per capita in 1960

The inanimate energy consumption per capita in 1960

The percentage of the labor force in industry in 1960

Bollen, K. A. (1989). Structural Equations with Latent Variables. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

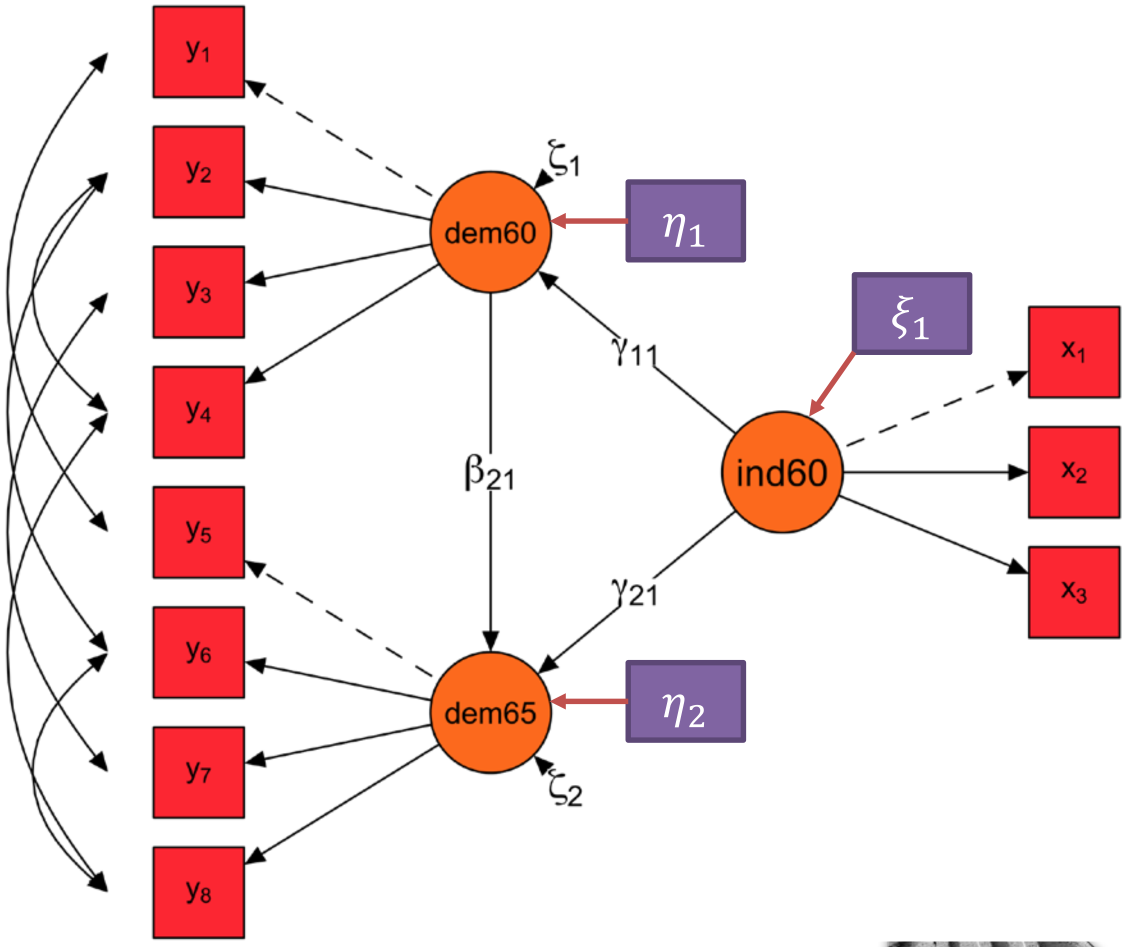
Example: Industrialization And Political Democracy

Latent Variable Model

Vector of endogenous latent variables	eta	$\boldsymbol{\eta}$
Matrix of coefficients	beta	$\boldsymbol{\beta}$
Matrix of coefficients of effect of exogenous latent variables on endogenous latent variables	gamma	$\boldsymbol{\Gamma}$
Vector of exogenous latent variables	xi	$\boldsymbol{\xi}$
Vector of errors	zeta	$\boldsymbol{\zeta}$
Covariance Matrix of latent exogenous variables	phi	$\boldsymbol{\Phi}$
Covariance Matrix of equation errors	psi	$\boldsymbol{\Psi}$

$$\boldsymbol{\eta} = \boldsymbol{\beta}\,\boldsymbol{\eta} + \boldsymbol{\Gamma}\,\boldsymbol{\xi} + \boldsymbol{\zeta}$$

$$\begin{bmatrix} dem60 \\ dem65 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} dem60 \\ dem65 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} [ind60] + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$



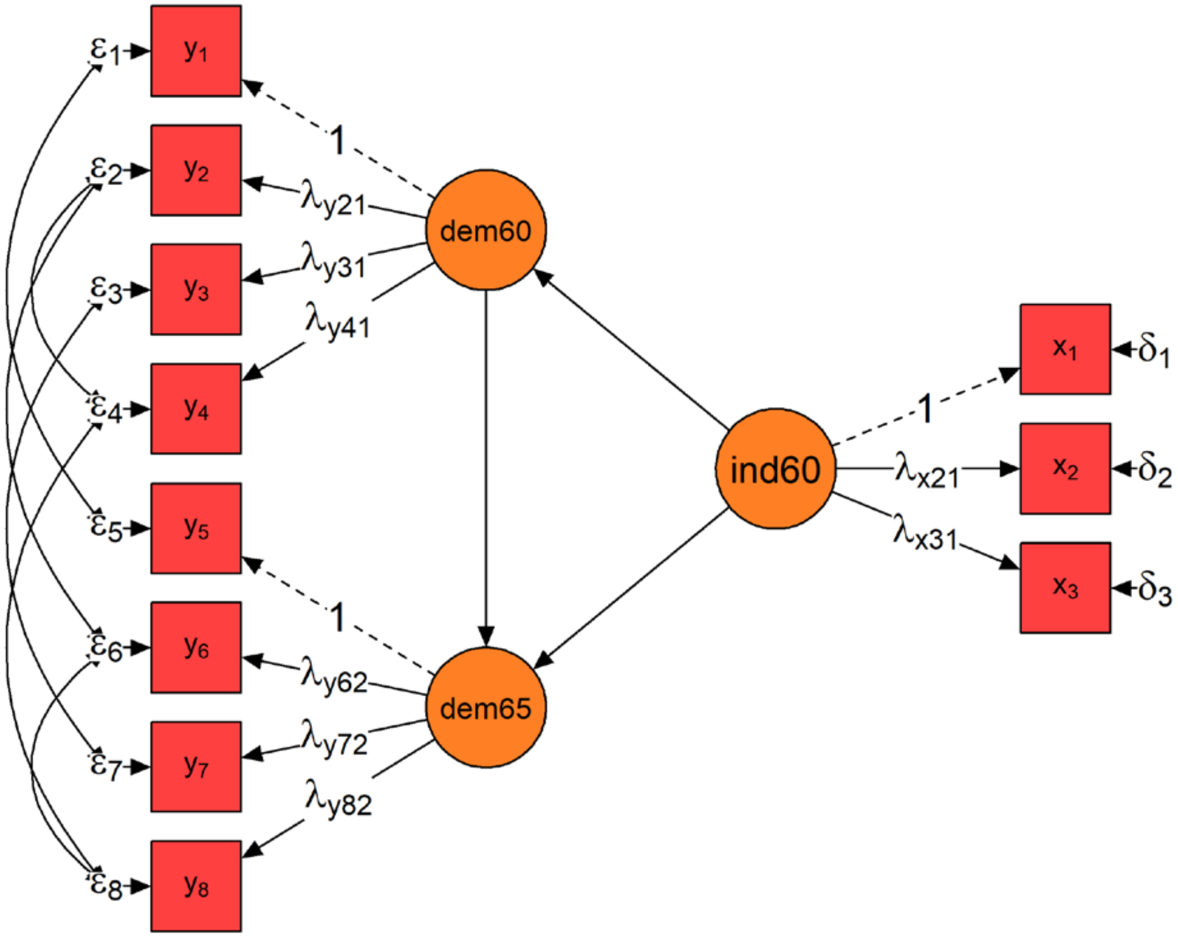
Example: Industrialization And Political Democracy

$\mathbf{x} = \Lambda_x \xi + \delta$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_{x21} \\ \lambda_{x31} \end{bmatrix} [ind60] + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}$$

$\mathbf{y} = \Lambda_y \xi + \epsilon$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{y21} & 0 \\ \lambda_{y31} & 0 \\ \lambda_{y41} & 0 \\ 0 & 1 \\ 0 & \lambda_{y62} \\ 0 & \lambda_{y72} \\ 0 & \lambda_{y82} \end{bmatrix} \begin{bmatrix} dem60 \\ dem65 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \end{bmatrix}$$



Talk Overview

- What is SEM?
- **SEM software packages**
- SEM examples: running the software
- Wrapping up
- Slides there will be no time for

Mplus

- Advantages
  - Single package that can do:
    - SEM
    - Multilevel SEM
    - Bootstrapping
    - Multiple imputations
    - Add weights
    - Latent categorical variables (latent classes)
    - Path diagrams
- Disadvantages
  - Black box: limited to options the developers give you
    - Example: Does not calculate average variance extracted (AVE) and composite reliability (CR), sometimes used as reliability measures
  - Not free
  - No multilevel path diagrams

- Advantages
  - Open-source; many options for the user
  - Users/coders can (re)write software for perceived needs and omissions in current packages offered
  - Not just statistics
    - Survey api to point directly to surveys
    - Security: use in conjunction with stats
    - Display results in html (e.g. Shiny, Markdown)
    - …
- Disadvantages (pertaining to SEM)
  - While R can do SEM, multilevel SEM, bootstrapping, multiple imputations, weights, path diagrams, etc. no single PACKAGE can do all (yet)
  - Packages:
    - R is a base code for which coders write packages that are then uploaded by users who need to use specific features; therefore some comfort with basic coding is advised

lavaan package in R

- Can do:
  - SEM
  - Bootstrapping
- Cannot (yet) do:
  - hierarchical/multilevel datasets (multilevel cfa, multilevel sem)
  - Latent categorical variables (latent classes)
  - Path diagrams
- More on website: http://lavaan.ugent.be/

lavaan ecosystem (non-random sample)

Path diagram plotting packages

- semPlot
- Onyx
- RAMpath

Extensions

- lavaan.survey: survey ***weights***, clustering, strata, and finite sampling corrections in SEM
- blavaan: Bayesian SEM with a lavaan interface
- semTools: collection of useful functions for SEM including ***multiple imputations***
- simsem: simulation of SEM models

Talk Overview

- What is SEM?
- SEM software packages
- **SEM examples: running the software**
- Wrapping up
- Slides there will be no time for

Recall our example: Industrialization and Political Democracy (Bollen, 1989)

Expert ratings of the freedom of the press in 1960

The freedom of political opposition in 1960

The fairness of elections in 1960

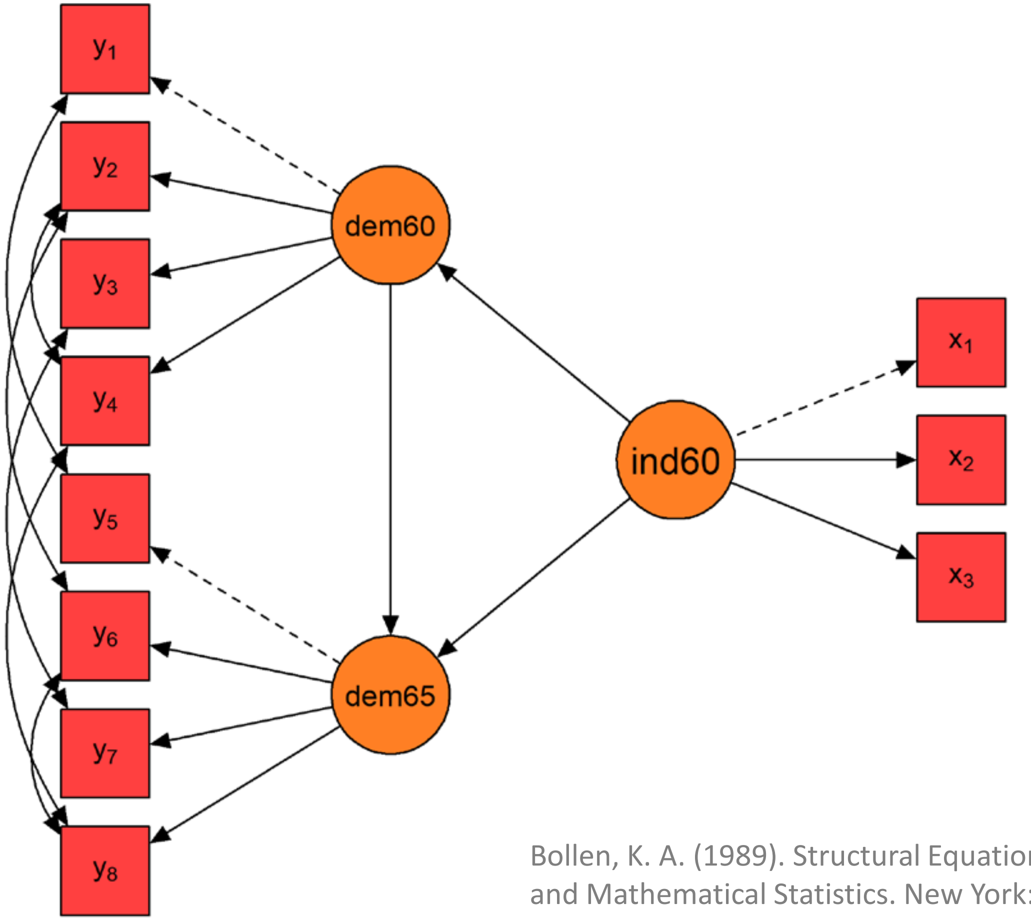
The effectiveness of the elected legislature in 1960

Expert ratings of the freedom of the press in 1965

The freedom of political opposition in 1965

The fairness of elections in 1965

The effectiveness of the elected legislature in 1965



The gross national product (GNP) per capita in 1960

The inanimate energy consumption per capita in 1960

The percentage of the labor force in industry in 1960

Bollen, K. A. (1989). Structural Equations with Latent Variables. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

R and Mplus: code side by side

```
# Political Democracy example
PolDem<-'# measurement model
          ind60 =~ x1 + x2 + x3
          dem60 =~ y1 + y2 + y3 + y4
          dem65 =~ y5 + y6 + y7 + y8

        # regressions / latent variable model
          dem60 ~ g11*ind60
          dem65 ~ g21*ind60 + b21*dem60

        # residual covariances
          y1 ~~ y5
          y2 ~~ y4 + y6
          y3 ~~ y7
          y4 ~~ y8
          y6 ~~ y8

        # Indirect
          indirect := g11*b21

        # Contrasts
          contrast := g21 - g11*b21

        # Total effect
          total    := g21 + g11*b21
        '
fit_PD<-sem(PolDem, data=PoliticalDemocracy, mimic="Mplus", estimator="MLR")

# Summary of results
summary(fit_PD, fit.measures=TRUE, rsquare=TRUE, ci=TRUE)
```

Due to concerns about *lavaan*'s capabilities, Yves has created 'mimic' arguments in the code that will calculate the answers to mirror Mplus, LISREL, and EQS

```
TITLE:     Political Democracy Example;

DATA:      FILE IS "C:/Users/statmath/Downloads/PD.dat";

VARIABLE:  NAMES ARE y1 y2 y3 y4 y5 y6 y7 y8 x1 x2 x3;
           USEVARIABLES ARE y1 y2 y3 y4 y5 y6 y7 y8 x1 x2 x3;

ANALYSIS:  ESTIMATOR IS MLR;

MODEL:     ind60 BY x1 x2 x3;
           dem60 BY y1-y4;
           dem65 BY y5 y6 y7 y8;

           dem60 ON ind60;
           dem65 ON ind60 dem60;

           y1 WITH y5;
           y2 WITH y4 y6;
           y3 WITH y7;
           y4 WITH y8;
           y6 WITH y8;

MODEL INDIRECT:
           dem65 IND dem60 ind60;


OUTPUT:    STDYX TECH1 TECH4 MODINDICES CINTERVAL;
```

R and Mplus

Let's see this example in R

Another example: Maslach Burnout Inventory (MBI)

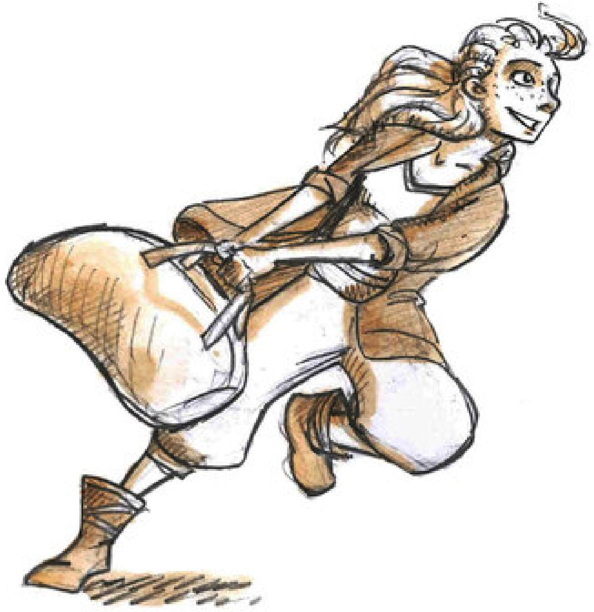
- This study uses a 22-item instrument with uniform Likert responses 0 (*feeling has never been experienced*) to 6 (*feeling experienced daily*) with three latent subscales:
  - Emotional Exhaustion
    - Example item: I feel emotionally drained from my work
  - Depersonalization
    - Example item: I feel I treat some recipients as if they were impersonal 'objects'
  - Personal Accomplishment
    - Example item: I can easily understand how my recipients feel about things

NOTE: 'Recipient' is someone who receives one's care or service related to their job

Byrne, B. M. (2012). *Structural Equation Modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge. (Chapter 4 example)

Maslach, C., & Jackson, S. E. (1981). *Maslach Burnout Inventory manual*. Palo Alto, CA: Consulting Psychologists Press.

Another example: Maslach Burnout Inventory (MBI)



Back to the software !!!

Talk Overview

- What is SEM?
- SEM software packages
- SEM examples: running the software
- **Wrapping up**
- Slides there will be no time for

Advantages of SEM

- Confirmatory approach: test your model/theory
- Provides goodness-of-fit measures
- Flexible statistical modeling approach
- SEM can handle:
  - Missing data (fiml, multiple imputations)
  - (in)equality constraints
  - Mixed categorical data (binary, ordinal, count,...)
  - Mixed discrete and continuous latent variables
  - Clustered (multilevel) data
- Many other approaches are special cases (e.g. generalized linear mixed models)

Disadvantages of SEM

- Modeling flexibility can be overwhelming (e.g. there are connections everywhere!!!)
- Dedicated software needed (not available in SPSS)
- ‘Specifying’ model in software can be challenging
- As a statistical field
  - Better (?) inference for small samples
  - Understanding output can be a challenge without ‘complete’ knowledge of the model and techniques
- Not well connected to other branches of statistics (mixed methods,…)

Uses of SEM

- Widely used in social sciences: e.g. education, psychology
- Now being used in:
  - Medical sciences
  - Nueroimaging
  - Biology, ecology (e.g. climate change)
  - Operation research
  - …and growing
- SEM software is used for ‘standard’ analyses (e.g. regression) when there is need for:
  - Robust standard errors
  - Diagnostics
  - (in)equality restraints: needing a parameter to be larger
  - …

Thank you

Questions?

There are additional resources in the slides that follow

email: jcwild@Indiana.edu

Talk Overview

- What is SEM?
- SEM software packages
- SEM examples: running the software
- Wrapping up
- **Slides there will be no time for**

SEM is like a car…

- Remember, in part 1 of the slides one of the disadvantages of SEM was noted as “not well connected to other branches of statistics.” Specific to this, Yves Rosseel mentioned in his Keynote that the uniqueness of latent variables in the model separate SEM from other techniques like regression, ANOVA, etc., and that has led to a specific SEM journal:

Structural Equation Modeling: A Multidisciplinary Journal (since 1994)

- You can imagine that a branch of statistics that has had a specific journal solely dedicated to it is quite large. If SEM is like a car, some are drivers who can do some maintenance themselves, such as rotating tires, changing the oil and brake fluid, but may need to take their car to a mechanic for more complicated repairs. In other words, some researchers specialize in estimation, others in interpretation, others in fit indices, etc. Have fun exploring your interest.

Other interesting things that we can't get into

- SEM and regression used to analyze the same data:
  - https://pdfs.semanticscholar.org/5280/080a464f14ffe4fdf7ef501a6ff581cb34e8.pdf
  - https://ac.els-cdn.com/S0261517709000521/1-s2.0-S0261517709000521-main.pdf?_tid=8c07ce88-d5d8-11e7-b8dd-00000aab0f27&acdnat=1512051358_f1bf721894435f4c1ff7e404b914085b
  - https://repositories.lib.utexas.edu/bitstream/handle/2152/22516/XIAO-MASTERSREPORT-2013.pdf?sequence=1&isAllowed=y

- SEM and ordinal data:
  - https://www.academia.edu/7605083/Evaluating_estimation_methods_for_ordinal_data_in_structural_equation_modeling?auto=download
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3103762/ (estimating covariances with ordinal and mixed data)
  - http://cmss.univnt.ro/wp-content/uploads/vol/split/vol_IV_issue_1/CMSS_vol_IV_issue_1_art.002.pdf

- Latent Categorical Variables discussion & examples:
  - https://stats.idre.ucla.edu/mplus/seminars/intromplus-part2/mplus-class-notesanalyzing-data-latent-class-and-other-mixture-models/
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3823694/

- Yves Rosseel keynote:
  - https://channel9.msdn.com/Events/useR-international-R-User-conferences/useR-International-R-User-2017-Conference/KEYNOTE-Structural-Equation-Modeling-models-software-and-stories

Identification

- Identification: the model must *receive as many or more* **inputs** than *requested* **parameters**
  - Given the sample covariance matrix and the implied model, can a unique set of parameters be found?
    - Underidentified: df < 0 (not all requested parameters can be estimated)
    - Just identified: df = 0 (parameters estimations are unique)
    - Over identified: df > 0 (more than one parameter estimation)

  - Resource: https://stat.utexas.edu/images/SSC/Site/AMOS_Tutorial.pdf (pp. 11-12)

  - Tests for identification
    - *t*-rule (necessary, but not sufficient)
    - Null B rule (sufficient, but not necessary)
    - Recursive rule (sufficient, but not necessary)
    - Rank and Order (for non-recursive models)
    - Rank condition (necessary and sufficient)

  - Resource: https://stat.utexas.edu/software-faqs/lisrel/146-training/software/655-lisrel-assessing-model-identification

How does SEM estimate the parameters?

- OLS: minimize the residual sum of squares
- SEM: minimize the difference between sample covariance and the model predicted covariance

- H_0: Observed covariance is a function of a set of parameters *if* model is correct and parameters are known, then residuals = **0**

- $\boldsymbol{\Sigma = \Sigma(\Theta)}$
  - $\boldsymbol{\Sigma}$ **:** population covariance matrix of observed variables
  - $\boldsymbol{\Theta}$ **:** vector of model parameters
  - $\boldsymbol{\Sigma(\Theta)}$ **:** covariance as a function of model parameters

Estimation (continued)

- Begin with a **simple** model: $y_1 = x_1 + \zeta_1$
- If the model is correct and the population parameters are known, then:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta) \ or \begin{bmatrix} var(y_1) & cov(y_1, x_1) \\ cov(y_1, x_1) & var(x_1) \end{bmatrix} = \begin{bmatrix} \phi_{11} + \psi_{11} & \phi_{11} \\ \phi_{11} & \phi_{11} \end{bmatrix}$$

- But we never know population parameters; the observed sample covariance matrix:

$$\boldsymbol{S} = \begin{bmatrix} var(y_1) & cov(y_1, x_1) \\ cov(y_1, x_1) & var(x_1) \end{bmatrix}$$

- The fitted/model implied covariance matrix is given by:

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \hat{\phi}_{11} + \hat{\psi}_{11} & \hat{\phi}_{11} \\ \hat{\phi}_{11} & \hat{\phi}_{11} \end{bmatrix}$$

Estimation (continued)

- SEM chooses values of $\hat{\boldsymbol{\Sigma}}$ to minimize:

$$\boldsymbol{S} - \hat{\boldsymbol{\Sigma}} = \begin{bmatrix} var(y_1) - (\hat{\phi}_{11} + \hat{\psi}_{11}) & cov(y_1, x_1) - \hat{\phi}_{11} \\ cov(y_1, x_1) - \hat{\phi}_{11} & var(x_1) - \hat{\phi}_{11} \end{bmatrix}$$

- Estimate the $\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$ so that $\hat{\boldsymbol{\Sigma}}$ is as "close" to $\boldsymbol{S}$ as possible

- Three common types of fitting functions:
  - Maximum likelihood
  - Unweighted least squares
  - Generalized least squares

Fit indices

- Chi-square
  - Only statistical test of model fit; H_0: $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{0}$
  - However,
    - Depends heavily on sample size leading to small misspecifications leading to rejection
      - Simulation studies can help determine the sample size needed for a model
    - Also, depends on parameter values, number of indicators per factor…
    - In other words, it has been known to allow serious misspecification while rejecting irrelevant misspecification

Fit indices

- Comparing worst fit (baseline) to “best” fit or model fit
  - TLI (Tucker-Lewis index):

$$p_2 = \frac{\chi_b^2/df_b - \chi_m^2/df_m}{\chi_b^2/df_b - 1}$$

  - CFI (Comparative fit index):

$$CFI = \frac{(\chi_b^2 - df_b) - (\chi_m^2 - df_m)}{(\chi_b^2 - df_b)}$$

- Must consider:
  - Choice of baseline
  - Estimation method
- Index interpretation (remember: E(χ^2)=df)
  - < 1 : misspecification
  - =1 : ideal fit
  - > 1 : overfitting
  - **> 0.9 : good fit**

Fit indices

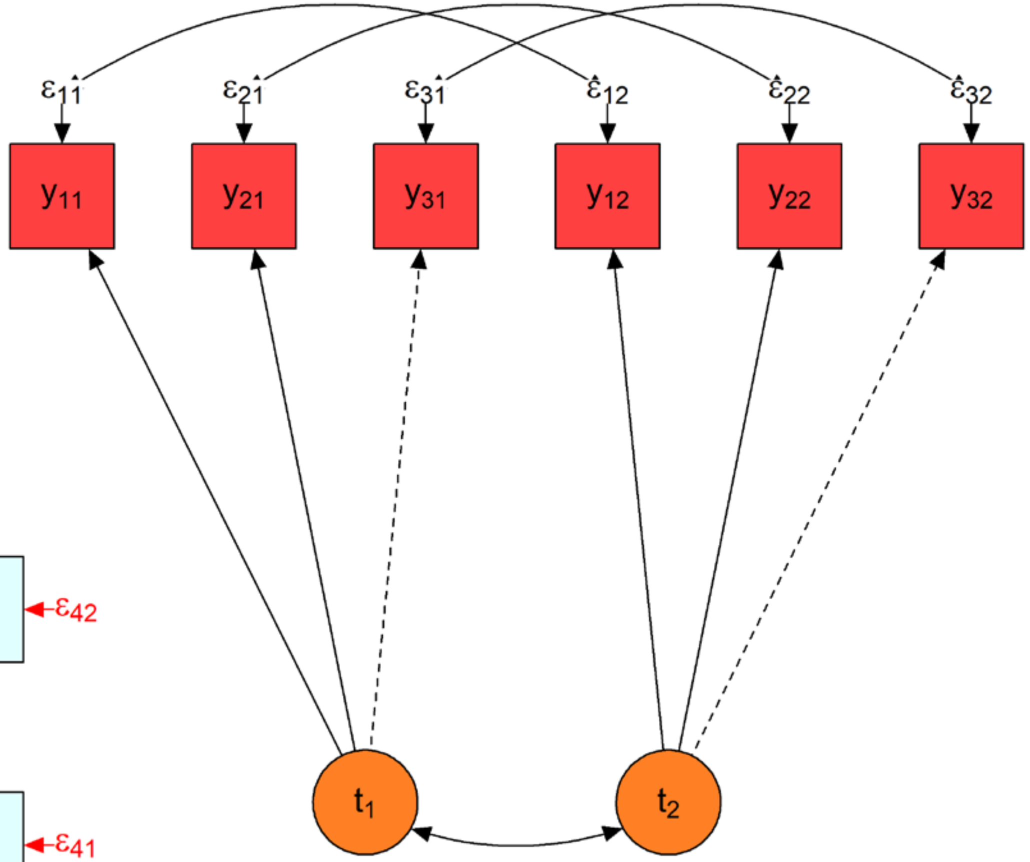
- Root Mean Square Error of Approximation in Population
  - RMSEA:

$$RMSEA = \sqrt{(\frac{\chi^2}{df} - 1)/(N - 1)}$$

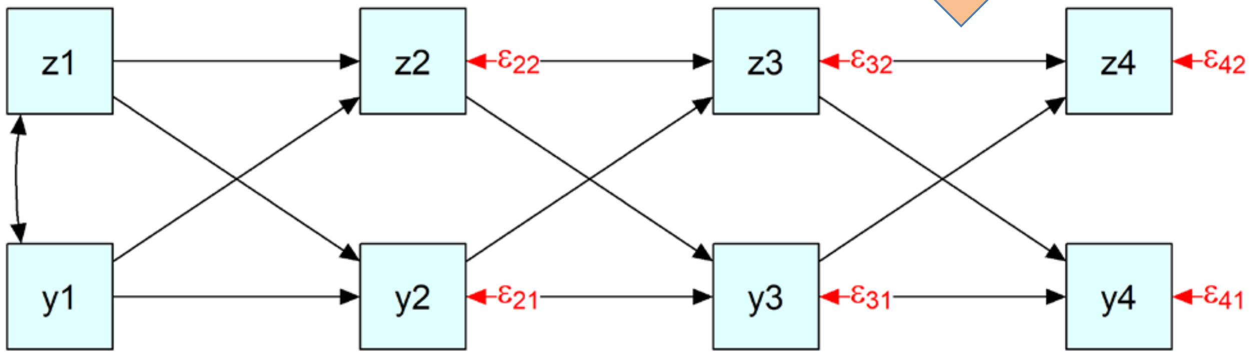
  - 90% confidence intervals available
- Index interpretation (remember: E(χ^2)=df)
  - > .10 : misspecification
  - < .08 : reasonable fit
  - **< .05 : good fit**

More SEM examples

- Longitudinal: paired t-test using latent variables

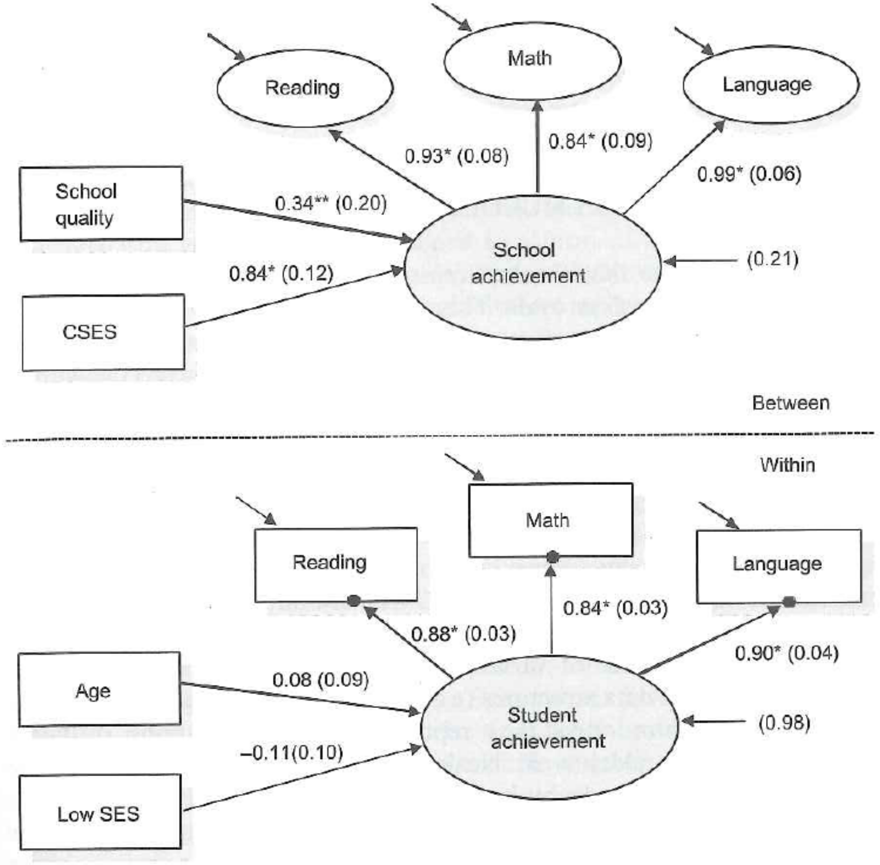


- Panel models with crosslagged effects



Multilevel SEM

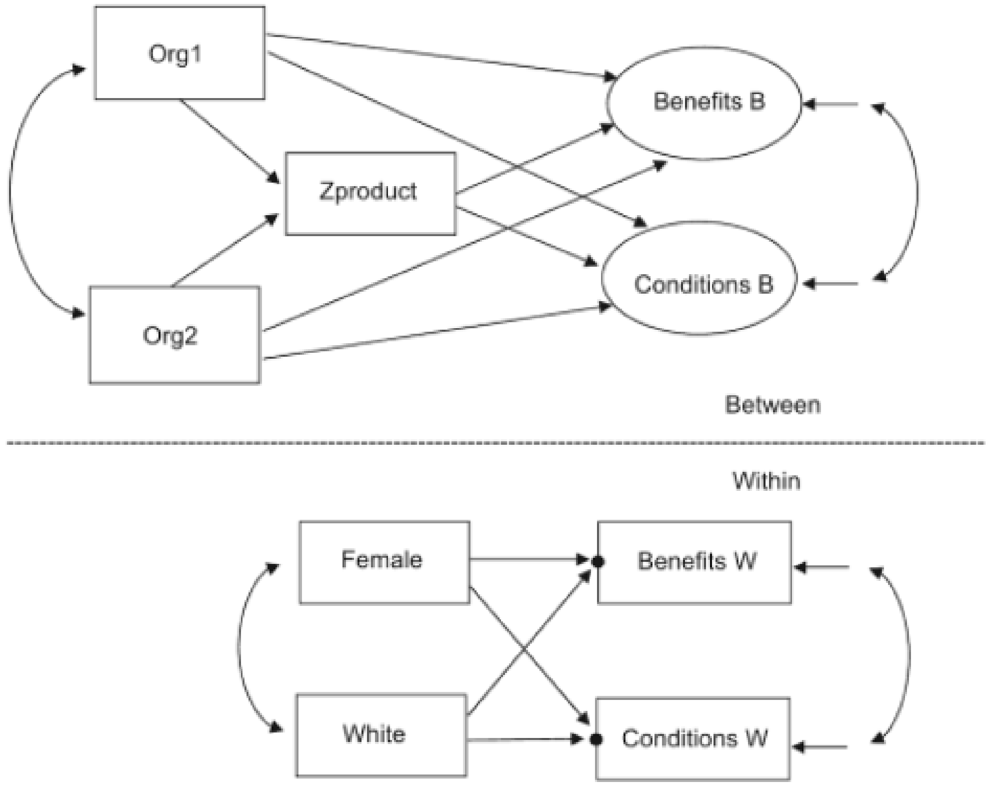
- In education and other research, individuals are (naturally) clustered into groups, justifying multilevel SEM analysis
  - Additionally, you may have variables measured at the group level
- Group level is often called the "between level" or "between model"
  - The model looks at differences between groups
- Individual level is often called the "within level" or "within model"
  - The model looks at differences within each group



Multilevel SEM: another example

A large random sample of employees nested with 160 product and service organizations

- Individual level
  - Satisfaction with **benefits** (10 point scale)
  - Satisfaction with workplace **conditions** (10 point scale)
  - Demographics
    - Sex (**female**: 1=female, 0=male)
    - Ethnicity (**white**: 1=white, 0=not-white)
- Organizational level
  - **Org1** – organizational context (e.g. type)
  - **Org2** – organizational context (e.g. management type)
  - **Zproduct** – standardized measure of productivity



Mplus: comparing single level and multilevel code

Single level

```
TITLE:    Political Democracy Example;

DATA:     FILE IS "C:/Users/statmath/Downloads/PD.dat";

VARIABLE: NAMES ARE y1 y2 y3 y4 y5 y6 y7 y8 x1 x2 x3;
          USEVARIABLES ARE y1 y2 y3 y4 y5 y6 y7 y8 x1 x2 x3;

ANALYSIS: ESTIMATOR IS MLR;

MODEL:    ind60 BY x1 x2 x3;
          dem60 BY y1-y4;
          dem65 BY y5 y6 y7 y8;

          dem60 ON ind60;
          dem65 ON ind60 dem60;

          y1 WITH y5;
          y2 WITH y4 y6;
          y3 WITH y7;
          y4 WITH y8;
          y6 WITH y8;

MODEL INDIRECT:
          dem65 IND dem60 ind60;


OUTPUT:   STDYX TECH1 TECH4 MODINDICES CINTERVAL;
```

Multilevel

```
TITLE:    Multilevel example;

DATA:     FILE IS "C:/Users/statmath/Downloads/ch4mv.dat";
          FORMAT IS 7f8.0,6f8.2;

VARIABLE: NAMES ARE orgid female white satpay morale org1
                    org2 benefit cond resour zproduct lev1wt lev2wt;
          USEVARIABLES ARE benefit cond female white org1 org2 zproduct;
          within  = female white;
          between = org1 org2 zproduct;
          CLUSTER IS orgid;

ANALYSIS: TYPE = twolevel; ESTIMATOR IS MLR;

MODEL:    %Between%
          benefit cond ON   org1;
          zproduct     ON   org2(a);
          benefit cond ON   zproduct(b1 b2);
          benefit      WITH cond;
          %Within%
          benefit cond ON   female white;
          benefit      WITH cond;
          Model constraint: New(ind1 ind2);
          ind1=a*b1; ind2=a*b2;

OUTPUT:   STDYX TECH1 TECH4 TECH8 MODINDICES CINTERVAL;
```

A Brief History of SEM software packages

- Closed-source: main four
  - LISREL ('70s, Karl Jöreskog)
  - EQS ('80s, Peter Bentler)
  - AMOS ('90s, James Arbuckle)
  - Mplus (Bengt Muthén, 1998-now)
- SAS through proc CALIS, proc TCALIS
- SPSS: Bought AMOS and sells as a separate product (have not changed much since)
- Stata module (free): 'gllamm' (Sophia Rabe-Hesketh, Anders Skrondal, Andrew Pickles, since 2002)
- Statistica (SEPATH), Systat (RAMONA), Mx (Michael Neale, free, closed-source, '90s)
- Open-source
  - Interfaces between R and commercial packages:
    - REQS (Patrick Mair, Eric Wu, since 2008)
    - MplusAutomation (Michael Hallquist, since 2010)
- R packages
  - sem (John Fox, since 2001)
  - OpenMX (Steven Boker, Michael Neale, . . . since 2009)
  - lavaan (Yves Rosseel, since 2010)
  - lava (Klaus Holst, since 2012)
  - strum (Nathan Morris, since 2015; for latent variables of genetic analysis)

R: Average Variance Extracted (AVE) & Composite Reliability (CR)

- Since the base R code includes complex mathematical calculations, AVE and CR (often seen reported in journals) can be easily computed (code is available at the bottom of the Democracy example)
  - Where *i* is the number of indicators (observed variables)
  - lambda is the coefficient of effect for the indicator

$$AVE = \frac{\sum_{i=1}^{i} \lambda_i^2}{i}$$

$$CR = \frac{\left(\sum_{i=1}^{i} \lambda_i\right)^2}{\left(\sum_{i=1}^{i} \lambda_i\right)^2 + \left(\sum_{i=1}^{i} 1 - \lambda_i^2\right)}$$

More (interesting) detail about lavaan ecosystem packages

- Onyx (Timo von Oertzen, Andreas M. Brandmaier, Siny Tsang, since 2009): interactive graphical interface for SEM (written in Java)
- lavaan.survey (Daniel Oberski, since 2012): survey weights, clustering, strata, and finite sampling corrections in SEM
- semTools (Sunthud Pornprasertmanit and many others, since 2012): collection of useful functions for SEM including multiple imputations
- simsem (Sunthud Pornprasertmanit and many others, since 2012): simulation of SEM models
- RAMpath (Zhiyong Zhang, Jack McArdle, Aki Hamagami, Kevin Grimm, since 2012) : visualizations of SEM models
- semPlot (Sacha Epskamp, since 2013): visualizations of SEM models
- blavaan (Ed Merkle, Yves Rosseel, since 2015): Bayesian SEM (currently using jags) with a lavaan interface

A Multilevel SEM package in R?

- OpenMX
  - web page is hosted and maintained by the Social Science Research Institute and Quantitative Developmental Systems group at Penn State
  - Pre-built package binaries are hosted by the Human Dynamics Lab in the Department of Psychology at the University of Virginia
  - Specification using syntax of paths **or** matrices
  - Although some forums claim multilevel SEM is possible using this package, I have not had success duplicating their examples
    - https://openmx.ssri.psu.edu/node/4258

Software websites of interest

- Stata's gllamm:
  - http://www.gllamm.org/
- REQS:
  - https://cran.r-project.org/web/packages/REQS/REQS.pdf
  - https://escholarship.org/uc/item/2pk051gv
- Mplus Automation:
  - https://cran.r-project.org/web/packages/MplusAutomation/MplusAutomation.pdf
- SEM Stat Wiki using AMOS:
  - http://statwiki.kolobkreations.com/index.php?title=Main_Page