Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

Educational Psychology 960
Structural Equation Modeling

David Kaplan

Department of Educational Psychology

THE UNIVERSITY of WISCONSIN MADISON

Spring 2024

Introduction

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- Structural equation models seek to represent hypotheses about raw data or summary statistics derived from empirical measurements in terms of a smaller number of“structural” parameters defined by a hypothesized underlying model.
- Summary statistics include means, variances, and covariances.
- This definition is quite broad and includes many special cases, including factor analysis, path analysis (and regression) as well as general structural models for continuous and categorical observed and latent variables.
- SEM has psychometric, biometric, econometric origins.

History of SEM

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- Psychometric origins in Factor Analysis
  - Spearman (British School) and the underlying structure of mental abilities.
  - Thurstone (American School). Multiple dimensions of intelligence.
- Statistical factor analysis
  - Lawley (Maximum likelihood estimation)
  - Joreskog (and development of LISREL)
  - Anderson and Rubin

History of SEM

- Biometric and econometric origins in path analysis
  - Originated by Sewell Wright
- Simultaneous Equation Modeling
  - Models of the US economy
  - Issues of identification and estimation
  - Dynamic issues
- Sociological and Educational Applications
  - Wisconsin Status Attainment Model
  - Sewell, Hauser, Portes, Duncan etc.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

History of SEM

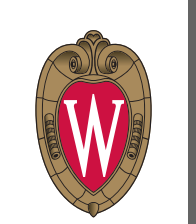
Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- The term “SEM” is usually reserved for the merging of path analysis and factor analysis
- “JKW” model (Joreskog, Keesling, Wiley)
- Addition of latent variables to address problems of measurement error.
- LISREL (Joreskog and Sorbom)

Developments in SEM

- Analysis of non-normal data
- Incorporating missing data
- Incorporating multilevel features and complex sampling designs
- Analysis of growth and change over time.
- Incorporating finite mixtures – merging categorical and continuous latent variables. One size doesn't fit all.
- Bayesian methods

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM



Model Specification

- Science achievement model

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

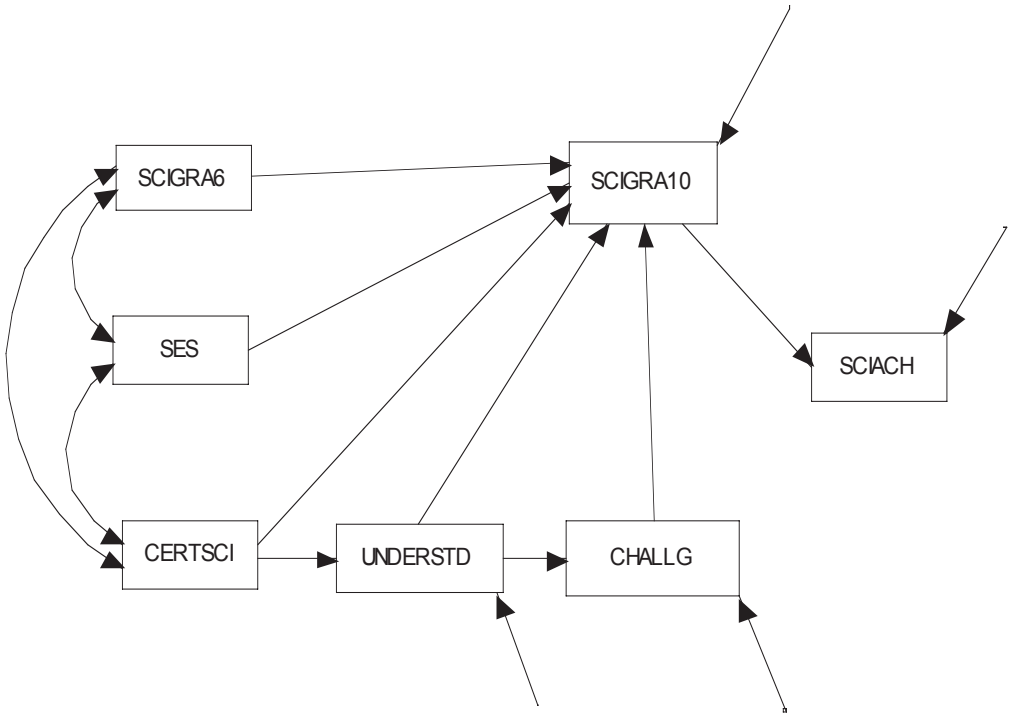


Figure 2.1 Education indicators model: Initial Specification

Model Specification

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The structural form of the model can be written in matrix notation as

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}, \tag{1}$$

- α is a vector of structural intercepts, $\mathbf{B}$ and $\boldsymbol{\Gamma}$ represent the structural relationships among the variables and ζ is a vector of residuals .

- The pattern of zero and non-zero elements in these matrices, in turn, are imposed by the underlying, substantive theory.

- For example, in the model shown in Figure 2.1, an element of $\mathbf{B}$ would be the path relating SIGRA10 to CHALLG. An element in $\boldsymbol{\Gamma}$ would be the path relating SIGRA10 to SES.

Model Specification

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Three types of parameters
  - Free (to be estimated)
  - Fixed (usually to zero)
  - Constrained (to be equal to other parameters)

Model Identification

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- Identification refers to whether the parameters of the model can be uniquely determined by the sample data.
- If the parameters of the model are not identified, estimation of the parameters is not possible.
- Example
  - $x + y = 4$
  - There are an infinite number of solutions. Not identified!
- Now let $x = y$
  - $2x = 4$
  - One and only one solution - identified!

Model Identification

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- We say that the elements in Ω (model parameters) are identified if they can be expressed uniquely in terms of the elements of the covariance matrix Σ (the data).
- The covariance matrix is a symmetric matrix with variances of the variables on the diagonal and covariances on the off-diagonal. If the data were standardized, this matrix would be the *correlation matrix* **R**.
- If all elements in Ω are identified we say that the model is identified.
- Note that here identification has nothing to do with data.

Rules of Identification

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- First, the disturbance term is a latent variable.
  - It has no natural metric so we fix the metric to the metric of the observed dependent variable $\mathbf{y}$ hence the multiplicative constant 1.0.
- Second, the matrix $\mathbf{B}$ has zero diagonal elements because an endogenous variable can't cause itself.
  - This is the normalization rule.

Rules of Identification

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The Counting Rule (necessary but not sufficient)
  - Let $s = p + q$, be the total number of endogenous and exogenous variables, respectively. Then the number of non-redundant elements in $\mathbf{\Sigma}$ is equal to $1/2s(s+1)$.
  - Let t be the total number of parameters in the model that are to be estimated (i.e. the free parameters). The counting rule states that a necessary condition for identification is that $t \leq 1/2s(s+1)$.
  - If the equality holds, then the model may be just *identified*. If t is strictly less than $1/2s(s+1)$, then the model may be *overidentified.* If t is greater than $1/2s(s+1)$, then the model may be *not identified.*

Rules of Identification

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- Example of Counting Rule
  - The total number of variables, s, in this model are seven. Thus, we obtain 28 elements in $\boldsymbol{\Sigma}$.
  - There are 10 variances and covariances of exogenous variables (including disturbances), and 8 path coefficients. Using the counting rule, we obtain $s - t$ = 28 - 18 = 10.
  - Because t is strictly less than the number of elements in $\boldsymbol{\Sigma}$, We say that the model is over-identified. The 10 overidentifying elements come from the 10 restrictions placed on the model.

Rules of Identification

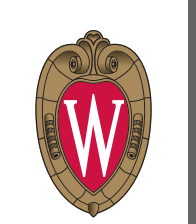
Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Aside: Counting rule and degrees-of-freedom
  - Overidentifying restrictions are the same as df.
  - Represents the number of *degrees of disconfirmation*.
  - A model with many dfs that fits the data is to be preferred over a model with fewer df.

Rules of Identification

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The Recursive Rule (sufficient)
  - A sufficient condition for identification is that $\mathbf{B}$ is triangular and that $\mathbf{\Psi}$ is a diagonal matrix, i.e. recursive models are identified.
  - Note that when $\mathbf{B} = \mathbf{0}$, we have a regression model, and so regression models are identified (in fact, just identified).
  - In other words, all recursive models are identified.



Rules of Identification for Non-Recursive Models

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

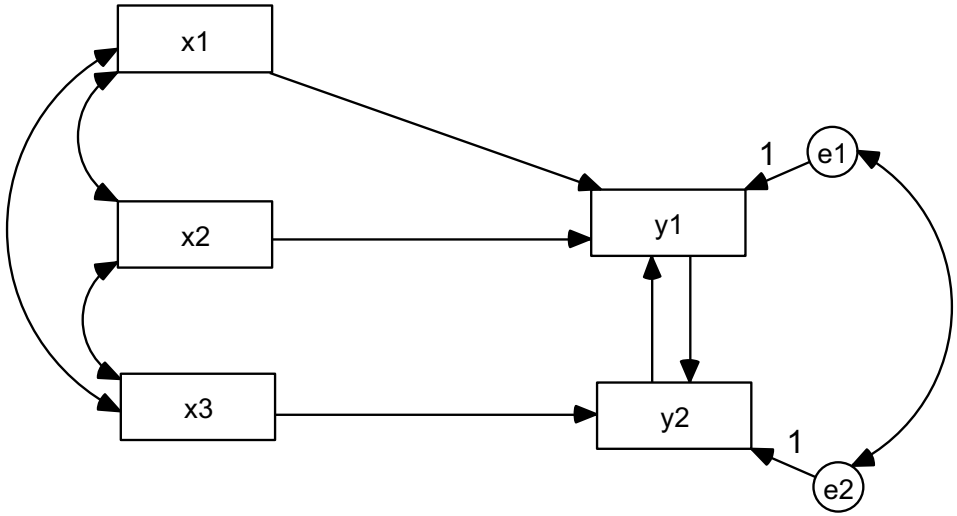


Figure 2.3. Prototype non-recursive path model

Rules of Identification for Non-Recursive Models

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Identification of non-recursive models
  - The number of variables (exogenous and endogenous) excluded (restricted) from any of the equations in the model must be at least $p - 1$.
  - Each equation must have at least one dedicated exogenous variable.
  - This is referred to as the "order condition" and is only a necessary condition for the identification of an equation of the model.

Rules of Identification for Non-Recursive Models

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- Practical relevance of identification
  - Not much
  - Useful for theory and to understand degrees-of-freedom
  - Not necessary to place in a substantive paper. Possibly relevant for methodological papers.

Some Useful Rules of Matrix Algebra

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

1. We can add and subtract matrices if they are the same dimension. If a matrix is $p \times p$ it can only be added or subtracted from a $p \times p$ matrix.

2. Matrices can be multiplied only if their inner orders are the same. Ex: A $p \times q$ matrix can be multiplied by any $q \times k$ matrix to yield a $p \times k$ matrix.

3. Matrix division (inversion) is simply matrix multiplication by the inverse of another matrix. Ex: To divide a matrix $\mathbf{A}$ by a matrix $\mathbf{B}$, we obtain the inverse of the matrix $\mathbf{B}$ and then multiple, $\mathbf{C} = \mathbf{A}\mathbf{B}^{-1}$

Some useful rules of matrix algebra (cont'd)

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

4. The matrix $\mathbf{B}$ must be invertible, and so it can't be singluar. That would be like dividing by zero.

5. The *determinant* of a matrix is a numerical assessment of the invertibility of a matrix. The determinant $|\mathbf{B}|$ can't be zero otherwise it is singluar.

6. The trace of a matrix $tr(\mathbf{A})$ is the sum of the diagonal elements of a matrix. The matrix must be square.

7. An identity matrix $\mathbf{I}$ is a square matrix with 1's in the diagonal and 0's in the off diagonal. If $\mathbf{I}$ is $p \times p$, then $tr(\mathbf{I}) = p$.

Estimation of Model Parameters

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Notation
  - $\boldsymbol{\Sigma}$ is the matrix of population variances and covariances of the data.
  - $\mathbf{S}$ is the matrix of sample variances and covariances.
  - $\boldsymbol{\Sigma}(\boldsymbol{\Omega})$ is the **model-based** matrix of population variances and covariances.
- The parameters of the model are
  - The variances and covariances of exogenous variables contained in $\boldsymbol{\Phi}$,
  - The variances and covariances of disturbance terms contained in $\boldsymbol{\Psi}$, and
  - The regression coefficients contained in $\mathbf{B}$ and $\boldsymbol{\Gamma}$.

Estimation of Model Parameters

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The goal is to obtain estimates $\hat{\mathbf{\Omega}}$ of the parameter vector $\mathbf{\Omega} = (\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Psi}, \mathbf{\Phi})$ that minimize a discrepancy function $F(\mathbf{S}, \hat{\mathbf{\Sigma}})$ where $\hat{\mathbf{\Sigma}}$ is the covariance matrix based on the estimates of the model – the so-called *fitted covariance matrix*.
- This is the goal of all statistics.
- A discrepancy function has three properties
  - $F(\mathbf{S}, \mathbf{\Sigma}(\hat{\mathbf{\Omega}})) \geq \mathbf{0}$
  - $F(\mathbf{S}, , \mathbf{\Sigma}(\hat{\mathbf{\Omega}})) = \mathbf{0}$ iff $\hat{\mathbf{\Sigma}}$ = $\mathbf{S}$
  - $F(\mathbf{S}, \mathbf{\Sigma}(\hat{\mathbf{\Omega}}))$ is a continuous function of $\hat{\mathbf{\Sigma}}$ and $\mathbf{S}$

Estimation of Model Parameters

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- *Maximum likelihood estimation* is a classic approach to estimating the parameters of the path model.
- ML is based on the multivariate normal distribution, and thus rests on multivariate normality as an assumption.
- The multivariate normal distribution can be written as follows:

$$f(\mathbf{z}) = (2\pi)^{-(p+q)/2}|\mathbf{\Sigma}(\mathbf{\Omega})|^{1/2}\exp\left[\left(-\frac{1}{2}\mathbf{z}'\mathbf{\Sigma}(\mathbf{\Omega})^{-1}\mathbf{z}\right)\right] \quad (2)$$

- Under the assumption of independent observations

$$f(\mathbf{z}_1, \mathbf{z}_2, \ldots \mathbf{z}_N) = f(\mathbf{z}_1)f(\mathbf{z}_2)\ldots f(\mathbf{z}_N) \quad (3)$$

Estimation of Model Parameters

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The product, referred to as the likelihood can be written as

$$L(\mathbf{\Omega}) = (2\pi)^{-N(p+q)/2}|\mathbf{\Sigma}(\mathbf{\Omega})|^{-N/2}\exp\left[\left(-\frac{1}{2}\sum_{i=1}^{N}\mathbf{z_i'}\mathbf{\Sigma}(\mathbf{\Omega})^{-1}\mathbf{z_i}\right)\right] \quad (4)$$

- Working with logs is easier

$$logL(\Omega) = \frac{-N(p+q)}{2}\log(2\pi) - \left(\frac{N}{2}\right)log|\mathbf{\Sigma}(\mathbf{\Omega})| \quad (5)$$

$$-\left(\frac{N}{2}\right)tr\left[\mathbf{S}\mathbf{\Sigma}(\mathbf{\Omega})^{-1}\right), \quad (6)$$

where *tr* is the so-called *trace* of the matrix which is the sum of the diagonal elements.

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- After some algebra, we then maximize the likelihood with respect to the model parameters. Note that model parameters are not in the first term, so we maximize

$$\log L(\boldsymbol{\Omega}) = -\frac{N}{2}\{\log|\boldsymbol{\Sigma}(\boldsymbol{\Omega})| + tr[\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1}]\} \tag{7}$$

- But this isn't a true discrepancy function! So, we subtract constants $\log|\mathbf{S}|$ and p that do not enter into the maximization.

- This then leads to the ML fitting function

ML Fitting Function

$$F_{ML} = \log|\boldsymbol{\Sigma}(\boldsymbol{\Omega})| + tr[\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1}] - \log|\mathbf{S}| - p. \tag{8}$$

- Maximization of this function via matrix calculus leads to the maximum likelihood estimates $\hat{\boldsymbol{\Omega}}$.

Estimation of Model Parameters

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- Differentiation of the F_{ML} function can take place twice.
- The solution to the second order derivatives of F_{ML} yield the covariance matrix of the estimates
- The standard errors are obtained from the square roots of the diagonal elements of the inverse of the information matrix.

Estimation of Model Parameters

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Properties of ML
  - If the assumptions of ML hold, then ML estimates
    - yield unbiased estimates
    - are asymptotically normal
    - have minimum sampling variability
  - When these assumptions do not hold, we need new estimation methods.
  - Major developments in SEM were focused on the fact that these assumptions almost never hold.

GLS and ULS

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The idea behind the generalized least squares (GLS) estimator is to correct for heteroskedastic disturbances.
- Recall from regression that this means that the variance of the residuals varies with values of the predictor.
- This is a violation of ML assumptions.
- The GLS estimator is a member of the family of weighted least squares (WLS) estimators that can be written generally as

$$F_{WLS} = [\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\Omega})]' \mathbf{W}^{-1} [\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\Omega})] \tag{9}$$

  where $\mathbf{W}^{-1}$ is a weight matrix that weights the deviations $\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\Omega})$ in terms of their variances and covariances with other elements.
- Notice that F_{WLS} is a proper fitting function.

Estimation of Model Parameters

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- A critical consideration of WLS estimators is the choice of a weight matrix $\mathbf{W}^{-1}$.
- One choice could be $\mathbf{W}^{-1} = \mathbf{I}$, the identity matrix. With the identity matrix as the choice for the weight matrix, WLS reduces to unweighted least squares (ULS). Similar to OLS.
- Although ULS is known to yield unbiased estimates of model parameters, it is not the most efficient choice of estimators with respect to yielding estimates with minimum sampling variability.

Estimation of Model Parameters

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- One can choose $\mathbf{W}^{-1} = \mathbf{S}^{-1}$. This is the most common choice for $\mathbf{W}^{-1}$.
- This choice leads to the GLS estimator

$$F_{GLS} = \frac{1}{2} tr[\mathbf{S}^{-1}(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\Omega}))]^2 \tag{10}$$

$$= \frac{1}{2} tr[\mathbf{I} - \mathbf{S}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Omega})]^2 \tag{11}$$

- We will consider other WLS estimators later.

Model and Parameter Testing

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- ML and GLS estimation of the path model allows us to explicitly test the hypothesis that the model fits the data.
- Consider the log-likelihood under the null hypothesis

$$logL(\mathbf{\Omega}) = -\frac{N}{2} log|\mathbf{\Sigma}(\mathbf{\Omega})| + tr[\mathbf{S}\mathbf{\Sigma}(\mathbf{\Omega})^{-1}] \tag{12}$$

- The log likelihood under the alternative hypothesis is

$$logL_{Ha} = -\frac{N}{2} \log|\mathbf{S}| + tr(\mathbf{S}\mathbf{S}^{-1}) \tag{13}$$

$$= -\frac{N}{2} \log|\mathbf{S}| + tr(\mathbf{I}) \tag{14}$$

$$= -\frac{N}{2} \log|\mathbf{S}| + p. \tag{15}$$

Model and Parameter Testing

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The likelihood ratio (*LR) test and is expressed as*

Likelihood Ratio Test

$$-2\log\frac{L_0}{L_1} = -2\log L_0 + 2\log L_1 \tag{16}$$

$$= N[\log|\boldsymbol{\Sigma}(\boldsymbol{\Omega})| + tr(\boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1}\mathbf{S}) - log|\mathbf{S}| + p] \tag{17}$$

$$= N \times F_{ML}. \tag{18}$$

Model and Parameter Testing

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Notice from the last equality in Equation (19) is $N \times F_{ML}$.
- The large sample distribution of the likelihood ratio test is chi-square with degrees-of-freedom given by the difference in the number non-redundant elements in $\boldsymbol{\Sigma}$ and the number of free parameters in the model.
- The likelihood ratio chi-square test is used to test the null hypothesis that the population covariance matrix possesses the structure implied by the model against the alternative hypothesis that $\boldsymbol{\Sigma}$ is an arbitrary symmetric positive definite matrix.

Model and Parameter Testing

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Consider a second hypothesis, say H_{02} that differs from H_{01} in that a single restriction $\theta_j = 0$ is relaxed.
- An example would be to simple free a path that is presently equal to zero.
- This would improve the fit, and so F_{ML} would be smaller.
- The change in the chi-square value can be written as

$$\Delta\chi^2 = N(F_{ML_1} - F_{ML_2}) \tag{19}$$

Model and Parameter Testing

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The second method of evaluating the components of the model concerns whether the restrictions placed on the model hold in the population.
- Recall that the restrictions are typically in the form of paths fixed to zero, meaning zero in the population.
- A test can be formed, referred to as the Lagrange Multiplier (LM) test, that assesses the validity of the restrictions in the model.
- The LM test is asymptotically distributed as chi-square with degrees-of-freedom equaling the difference between the degrees-of-freedom of the more restrictive model and the less restrictive model.
- The LM test is also referred to as the modification index.

Model and Parameter Testing

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Finally, we can consider evaluating the impact of placing restrictions on the unrestricted model.
- If a small (and perhaps non-significant) path coefficient was restricted to be zero (removed from the model) would that restriction hold in the population?
- The test for the validity of restricting parameters is given by the Wald test (W)
- The Wald test is asymptotically distributed as chi-square with degrees-of-freedom equaling the number imposed restrictions.
- Each of these three approaches are asymptotically related to each other. They give exactly the same value for the change in chi-square in infinitely large samples.

Decomposition of Effects

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- **Direct effect** = path coefficient (regression coefficient of one variable on another.
- **Indirect effect** (mediating effect) = product of coefficients along a path.
- **Total indirect** = sum of all indirect effects.
- **Total effect** = sum of direct + all indirect effects.
- Each of these effects are associated with standard errors that allow significance testing.

Mediation and Moderation

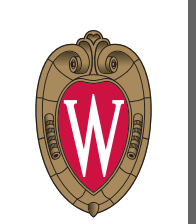
Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- It is important to distinguish between mediation and moderation.
- **Mediation** refers (simply) to the relationship between an exogenous variable X on an outcome Y through another variable, say, M.
- The mediator M is often a *process* variable in an *input-process-output* model.
    - For example, teacher qualifications impact student outcomes through the manner in which those qualifications translate into high quality teaching practices.
- An exogenous variable can have both a significant direct and indirect effect - *partial mediation*.

Mediation and Moderation

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- A **moderator** is basically another exogenous variable that **interacts** in an important way to change the slope of the relationship between the exogenous variable of interest and the outcome.
    - For example, student self-concept in reading might be moderated by gender, such that the relationship between self-concept in reading and reading scores might be flatter for boys than girls.
- Models of **moderated mediation** are also possible, and can be handled in multiple group models simply.



Moderated Mediation

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

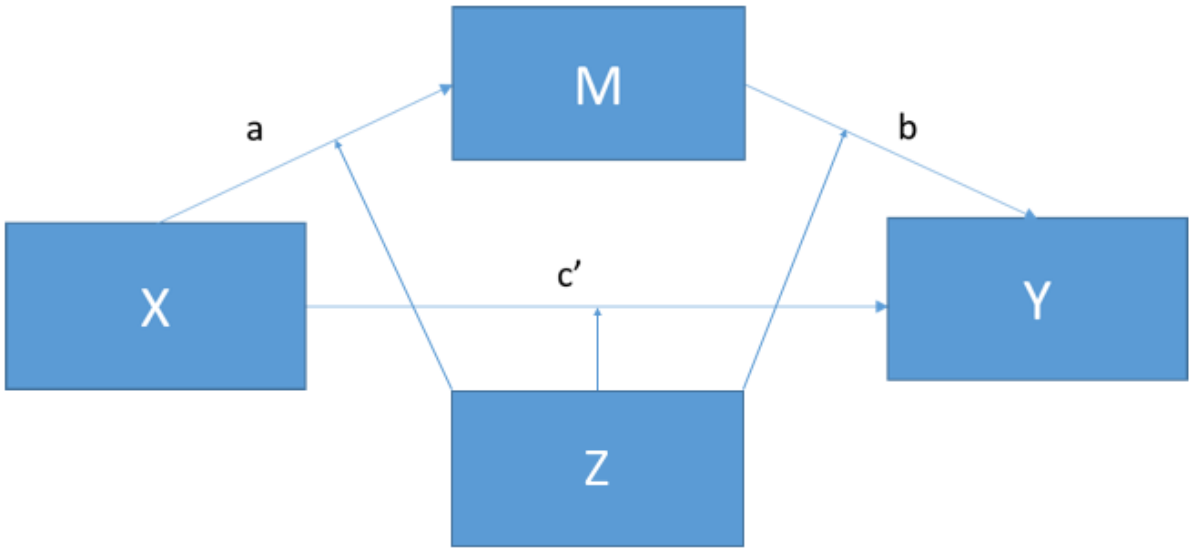
Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM



Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- A conundrum
  - For the properties of ML and GLS to hold, it is desirable to have 'large" sample sizes.
  - Large sample sizes ensure that "asymptotic" properties hold, and so we can trust the estimates, standard errors, and test statistics.
  - However, large sample sizes can also lead to the rejection of the overall fit of the model on the basis of the LR chi-square.
  - We would end up rejecting a model even when the specification errors in the model might be quite small.
  - This is the paradox of NHST in highly parameterized models.
  - What are we to do?

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Substantive researchers will attempt to evaluate the fit of the model and in some cases interpret parameter estimates regardless of whether the assumptions have been assessed and/or controlled.
- The problem of evaluating and interpreting structural equation models has dominated the methodological literature for many years.
- We divide alternative fit indices into three general categories:
  - Measures based on comparative fit to a baseline model, including those that add a penalty function for model complexity,
  - Measures based on population errors of approximation, and
  - Cross-validation/predictive adequacy measures.

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Measures Based on Comparative Fit to a Baseline Model
  - The fit of the model is compared to the fit of some baseline model that usually specifies complete independence among the observed variables.
  - The baseline model of complete independence is the most restrictive model possible and hence the fit of the baseline model will usually be quite large.
  - The issue is whether oneÕs model of interest is an improvement relative to the baseline model.
  - These indices are typically scaled to lie between zero and one, with one representing perfect fit relative to this baseline model.
  - The usual rule of thumb for these indices is that 0.95 is indicative of good fit relative to the baseline model.

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The Normed Fit Index

$$NFI = \frac{\chi_b^2 - \chi_t^2}{\chi_b^2} \tag{20}$$

- Values close to zero suggest that the target model is not much better than a model of complete independence among the variables.
- Values close to one suggest that the target model is an improvement over the baseline model.

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The Tucker-Lewis Index (NNFI)

$$TLI = \frac{(\chi_b^2/df_b - \chi_t^2/df_t)}{(\chi_b^2/df_b - 1)} \tag{21}$$

- Takes into account the expected value of the chi-square statistic
- The NFI and TLI assume the null hypothesis is true.
- Because the null is never true the distribution of the test statistic is better approximated by the non-central chi-square distribution.

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The Relative Non-Centrality Index

$$TLI = \frac{(\chi_b^2 - df_b) - (\chi_t^2 - df_t)}{(\chi_b^2 - df_b)} \tag{22}$$

- The RNI can lie outside the 0-1 range.
- Bentler (1990) adjusted the RNI so that it would like in the range of 0-1. This adjusted version is referred to as the *Comparative Fit Index* (CFI).

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Parsimony-Based Comparative Fit Indices
  - There are classes of comparative fit indices that adjust existing fit indexes for the number of parameters that are estimated.
  - The rationale behind these indices is that a model can be made to fit the data by simply estimating more and more parameters.
  - A model that is just-identified fits the data perfectly.
  - Therefore, it makes sense to require that these indices be adjusted for the number of parameters that are estimated.

Evaluating and Modifying SEMs

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- An Example of a Parsimony-Based Index

$$PNFI = \left(\frac{df_t}{df_b}\right) \times NFI \tag{23}$$

- The more parameters estimated in the target model, the less restricted the model becomes relative to the baseline model and the greater the penalty attached to the NFI.

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Measures Based on Errors of Approximation
  - The likelihood ratio chi-square test assesses an exact null hypothesis that the model fits perfectly in the population.
  - If the model is over-identified, then it is quite unlikely that the model will fit the data perfectly even if the entire population were measured.
  - It may be more sensible to assess whether the model fits approximately well in the population.

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Recall that assuming all assumptions hold and that the model is correctly specified, $N * F_{ML}$ has a χ^2 distribution.
- If the model is not correctly specified, then the test statistic follows a "non-central" χ^2 distribution with ν df and non-centrality parameter

$$F_0 * (N - 1) \tag{24}$$

- The population RMSEA is defined as

$$\epsilon = \sqrt{\frac{F_0}{\nu}} \tag{25}$$

- We could substitute F_{ML} for F_0 but F_{ML} can be quite biased if the model is not correctly specified.

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- As we are concerned with the model being misspecified, we can obtained the expected value of F_{ML} as

$$E(F_{ML}) = F_0 + \frac{\nu}{N} \tag{26}$$

- Thus a better estimate of F_0 would be

$$\hat{F}_0 = F_{ML} - \frac{\nu}{N} \tag{27}$$

- However, because ν/n could be larger than F_{ML}, this would mean that the fit would be negative and that is nonsensical. Thus, we constrain the lower bound to zero, and so the RMSEA becomes

$$\hat{\epsilon} = \sqrt{\max\left\{0, \frac{\hat{F}_0}{\nu}\right\}} \tag{28}$$

Evaluating and Modifying SEMs

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The Root Mean Square Error of Approximation
  - In utilizing the RMSEA for assessing approximate fit, a formal hypothesis testing framework is employed.

$$H_0 : \epsilon \leq 0.05 \tag{29}$$

  - Practical guidelines recommended by Browne & Cudeck (1993)
    - 0.05 and 0.08 are indicative of fair fit
    - 0.08 and 0.10 are indicative of mediocre fit.
  - Confidence intervals for the RMSEA are available in lavaan.

Model Selection

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Model selection measures
  - Not really model fit in the sense we've been talking about.
  - If we have two competing models for the same outcome, how might we choose among them?
  - We look at two similar looking, but conceptually different measures based on penalization.
    - Akaike Information Criteria (AIC)
    - Bayesian Information Criterion (BIC)

Model Selection

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The AIC

$$AIC = (-2)\textit{log-likelihood} + 2(\textit{number of parameters}) \quad (30)$$
$$= \chi^2 - 2(df) \quad (31)$$

- The model with the lowest AIC can be considered the best model from an out-of-sample predictive point of view.
- It would be the model that would do best in predicting the observations.

Model Selection

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The BIC
  - This statistic rests on the notion of Bayesian model selection
  - Consider two models M_1 and M_2, not necessarily nested.
  - M_1 could be an initially specified path model and M_2 could be the same model with one path deleted.
  - The posterior odds of M_1 relative to M_2 is

$$\frac{p(M_1|\mathbf{Y})}{p(M_2|\mathbf{Y})} = \left[\frac{p(\mathbf{Y}|M_1)}{p(\mathbf{Y}|M_2)}\right]\left[\frac{p(M_1)}{p(M_2)}\right] \tag{32}$$

$$= \mathbf{B}_{12}\pi_{12} \tag{33}$$

    where $\mathbf{B}_{12}$ is called the Bayes factor, and π_{12} is the prior odds of M_1 relative to M_2.

  - In words

$$\textit{posterior odds} = \textit{Bayes factor} \times \textit{prior odds} \tag{34}$$

Model Selection

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- The Bayes factor is simply the ratio of the posterior odds of M_1 to its prior odds.
- When the two models are equally probable, then the Bayes factor is equal to the posterior odds in favor of M_1.
- We typically do not see the Bayes factor used for model comparison when models are not considered equally probable.

Model Selection

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Under conditions where there is little prior information an approximation of the Bayes factor can be written as

$$2 \log B_{12} = \chi^2_{12} - df_{12} \log n, \tag{35}$$

where χ^2_{12} is the conventional likelihood ratio chi-square obtained from testing M_1 against M_2 and df_{12} is the difference in the degrees of freedom associated with each test.

$$BIC = -2 \log(\hat{\theta}|y) + p \log(n), \tag{36}$$

where $-2 \log \hat{\theta}|y$ describes model fit while $p \log(n)$ is a penalty for model complexity, where *p* represents the number of variables in the model and n is the sample size.

- We discuss Bayesian SEM at the end of the course.

Bayesian Information Criterion

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Rules of thumb have been developed to assess the quality of the evidence favoring one hypothesis over another using Bayes factors and the comparison of BIC values from two competing models.

Table: Grades of Evidence Corresponding to Values of the Bayes Factor for M_1 Against M_2, and the BIC Difference (Jeffreys, 1961; modified by Raftery, 1995)

BIC Difference	Bayes Factor	$p(M_1\|y)$	Evidence for M_1
0 to 2	1 to 3	0.50 – 0.75	Weak
2 to 6	3 to 20	0.75 – 0.95	Positive
6 to 10	20 to 150	0.95 – 0.99	Strong
> 10	> 150	$>$ .99	Decisive

Criticism of the BIC

1. The BIC is neither Bayesian nor an information criterion. The AIC is a true information criterion.
2. BIC is better in locating the true model if it is in the set of models being examined, and will find the true model as $n \to \infty$.
3. The BIC and AIC will usually track each other in large samples.
4. **BUT** these are not "goodness-of-fit" measures.
5. Not everyone agrees that the BIC is useful.
6. The general sense is that the BIC should be avoided. AIC is better if the goal is to choose a model with good predictive accuracy.

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

Model Modification and Statistical Power

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The most common reasons for model rejection are
  - violations of underlying assumptions,
  - incorrect restrictions placed on the model, and
  - sample size sensitivity.
- Methods of model modification usually involve relaxing restrictions in the model by freeing parameters that were fixed in the initial specification.
- The decision to free such parameters is often guided by the size of the LM (modification index) statistic.

Model Modification and Statistical Power

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- Three problems with model modification
  - First, model modification is post-hoc.
  - Second, searching for specification errors via the LM test does not always result in locating the specification errors imposed on a “true” model - that is, the model that generated the covariance matrix.
  - Third, there is an increase in the probability of Type II errors resulting from the goal of not rejecting the null hypothesis that the model fits the data
  - The way to mitigate the problem of Type II errors is to free paths that have maximum power.

Model Modification and Statistical Power

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- Satorra (1989) recognized that the LM test could be used to approximate the noncentrality parameter for each restriction in the model.
- In practical terms this means that for each restriction in the model one can assess whether the test is powerful enough to reject the null hypothesis that the parameter in question is zero.

Model Modification and Statistical Power

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Sample size sensitivity is only relevant when the null hypothesis is false. Look again at

$$N \times F_{ML} \tag{37}$$

- A method of gauging the influence of sample size and model misfit in the context of power analysis is through the use of the expected parameter change (EPC) statistic.

- The EPC is a point estimate of the alternative hypothesis for the parameter in question.

Model Modification and Statistical Power

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Model Modification suggestions based on Saris, Satorra, & Van der Veld (2009) *SEM*. Provided in "lavaan"

1. "**m**": Possible presence of a misspecification. Power is low but the MI is significant.
2. "**nm**": No misspecification. Power is high but the MI is non-significant.
3. "**EPC:m**": Possible serious misspecification. Power of the test is high and the MI is significant. Should also look at EPC to determine substantive importance. If EPC is large, then it is a serious misspecification.
4. "**EPC:nm**": EPC value is deemed low.
5. "**I**": Inconclusive.

Model Modification and Statistical Power

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Four Possible Scenarios (Saris, Satorra, Van der Veld (2009) *SEM*

	High Power	Low Power
Sign. MI	Inspect EPC (EPC)	Misspecification present (m)
Non-sign. MI	No misspecification (nm)	Inconclusive (I)

- Remember that these decisions are often made post-hoc!!

Model Modification and Statistical Power

Introduction

Path Analysis

Model specification

Model identification

Parameter estimation

Testing

Decomposition of effects

Model evaluation and modification

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

- Factors Influencing Model Modification and Power (Kaplan, 1988, 1989a,b; Kaplan & Wenger, 1993)
  - Propagation of specification errors has implications for model modification and power
  - Power has been found to differ in different parts of a model for the same size specification error

Introduction
Path Analysis
Model specification
Model identification
Parameter estimation
Testing
Decomposition of effects
Model evaluation and modification
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

See file: ”lavaan.PathAnalysis.html”

Factor Analysis

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- We have dealt so far with path analysis among observed variables
- As with regression, this assumes that the measures are perfectly reliable.
- Often, we are interested in underlying constructs which are "error free".
- Factor analysis concerns the identification and testing of underlying constructs.

Motivating Problem

- The example that will be used throughout this section will explore the factor structure of student perceptions of school climate.
- This problem has important implications for the input-process-output model not only because student perceptions are important education indicators in their own right, but they may be predictive of achievement.
- A researcher may postulate that there are several important dimensions to student perceptions.
- The question is whether a set of measurements that ask students to rate their agreement to statements about the climate of the school correlate in such a way as to suggest the existence of the factors in question.

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

The Linear Factor Model

- The linear factor analysis model and can be written as

Linear Factor Model

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta}, \tag{38}$$

where $\mathbf{y}$ is a vector of responses, $\boldsymbol{\Lambda}$ is a matrix of factor loadings, $\boldsymbol{\eta}$ is a vector of factors, and $\boldsymbol{\delta}$ is a vector of unique variables (uniquenesses).

- The covariance matrix of the observed data can be written in the form of the fundamental factor analytic equation,

$$\begin{aligned} \boldsymbol{\Sigma} &= Cov(\mathbf{y}\mathbf{y}') = \boldsymbol{\Lambda} E(\boldsymbol{\eta}\boldsymbol{\eta}')\boldsymbol{\Lambda}' + E(\boldsymbol{\delta}\boldsymbol{\delta}') && (39) \\ &= \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_{\delta}. && (40) \end{aligned}$$

where $\boldsymbol{\Phi}$ is the factor covariance matrix and $\boldsymbol{\Theta}$ is a diagonal matrix of the uniquenesses.

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

The Nature of Unique Variables

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The unique variables $\boldsymbol{\delta}$ do not contain only measurement error.
- Consider the model for a vector of true scores $\mathbf{t}$ rather than the observed scores $\mathbf{y}$.
- From classical true score theory, the vector of true scores is defined as

$$\mathbf{t} = \mathbf{y} - \mathbf{e}, \tag{41}$$

where $\mathbf{e}$ are true measurement errors.

- The factor model for the true scores then is

$$\mathbf{t} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \mathbf{s} \tag{42}$$

where $\mathbf{s}$ contains specific variances, defined as the variances in the true scores that are due to the particular selection of variables.

The Nature of Unique Variables

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- We see that the uniqueness term is

$$\boldsymbol{\delta} = \mathbf{s} + \mathbf{e} \tag{43}$$

- Despite the fact that the unique variance is composed of specific variance and error variance we typically assume that specific variances are small relative to measurement error variance.

Principal Components Analysis (PCA)

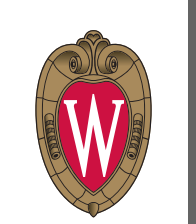
Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- PCA **is not** factor analysis.
- However, PCA provides results that are often quite similar to factor analysis and is included in many statistical packages as the default.
- A major difference between PCA and other forms of unrestricted factor analysis lies in the assumptions made about the existence of measurement error.
- PCA does not assume that the variables are measured with error. PCA simply transforms the original set of measurements into orthogonal components retaining the original amount of variance in the data.

Principle Components Analysis (PCA)

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Factor analysis, by contrast, specifically models measurement error, and extracts factors that account for maximum covariation in the observed variables.
- Often, PCA is used as a factor analysis model and a decision is made to retain fewer principal components for future rotation and interpretation.



Principal Axis Factoring

Introduction

Path Analysis

Factor Analysis

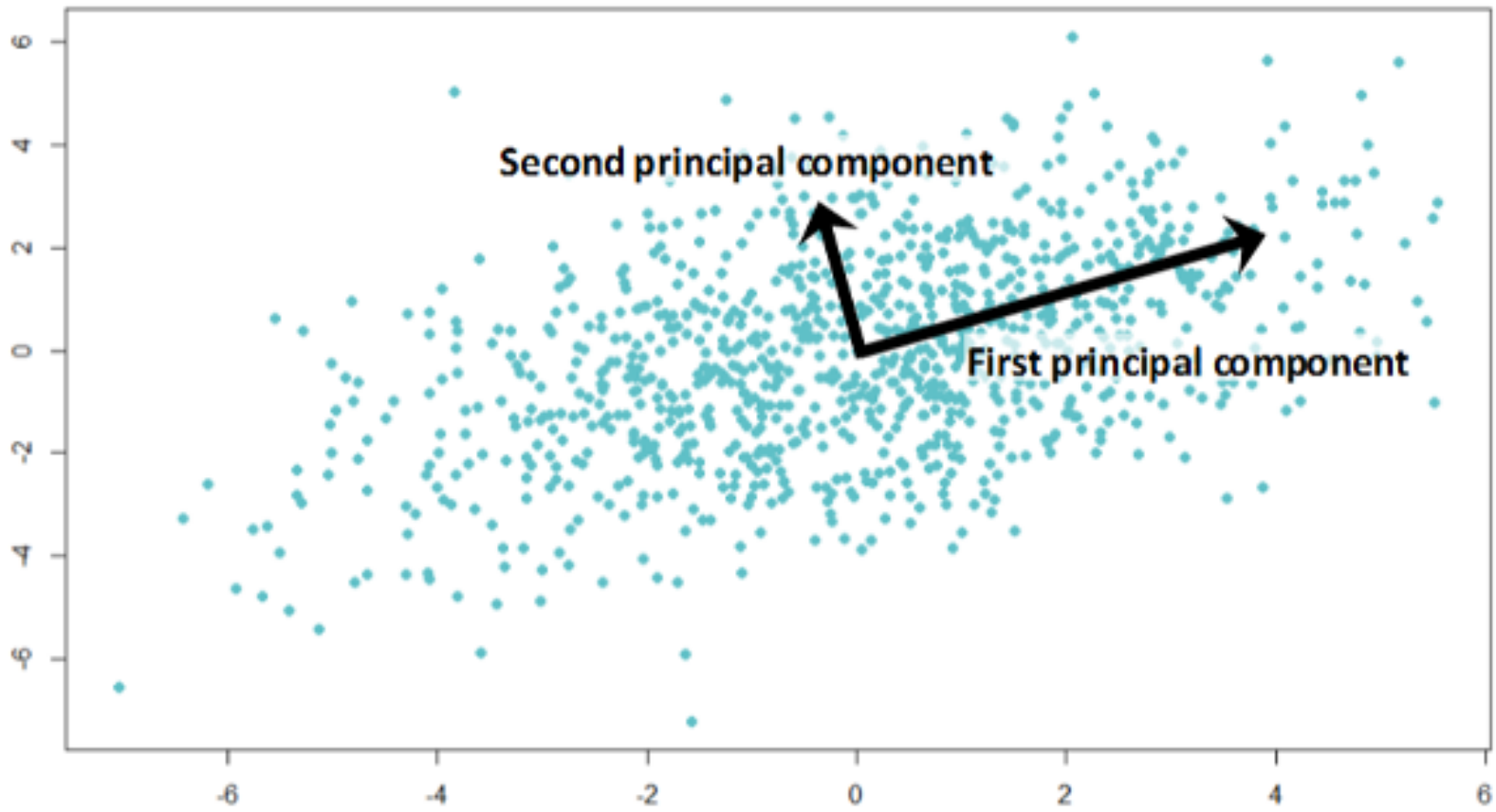
- Linear factor model
- Principal Components Analysis
- PAF
- Estimation of unrestricted model
- Rotation
- CFA
- Factor scores

Statistical Assump-tions

SEM

Multiple Group Modeling

Multilevel SEM



PAF

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- There are many ways that factors can be extracted.
- The method of PAF might be the most common in canned programs.
- PAF rests on a "variance maximizing transformation"

$$(\mathbf{\Sigma} - \lambda \mathbf{I})\mathbf{u} = \mathbf{0} \tag{44}$$

where λ are the eigenvalues and $\mathbf{u}$ are the eigenvectors of $\mathbf{\Sigma}$

- A trivial solution to the problem is $\mathbf{u} = \mathbf{0}$. A non-trivial solution would be to solve the determinant

$$|\mathbf{\Sigma} - \lambda \mathbf{I}| = \mathbf{0} \tag{45}$$

- The solution to this problem yields the identity

$$\mathbf{\Sigma} = \mathbf{u}\mathbf{D}\mathbf{u}' \tag{46}$$

showing that the the covariance matrix can be decomposed into the eigenvectors and eigenvalues.

A Simple Example

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Let Σ be

$$\begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$$

- Then the non-trival solution is

$$\begin{vmatrix} 1-\lambda & 0.5 \\ 0.5 & 1-\lambda \end{vmatrix} = 0$$

- This becomes the quadratic equation $\lambda^2 - 2\lambda + .75 = 0$.
- The roots are 1.5 and 0.5.
- So, eigenvalues are roots to a quadratic equation.
- Note that the first eigenvalue is 1.5, and this divided by the sum (2) is $1.5/2$ = .75. The second is $0.5/2$ = .25 The total is 1.0, which is the total amount of variance that can be explained.

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- What about the eigenvectors $\mathbf{u}$?
- There will be one eigenvector for each eigenvalue.
- This is obtained by solving the linear system $\mathbf{\Sigma u} = \lambda \mathbf{u}$.
- The (normalized) solution for 1.5 and for 0.5 are

$$\mathbf{u} = \begin{pmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{pmatrix}$$

- Note that the eigenvectors are orthogonal: $\mathbf{u}'\mathbf{u} = \mathbf{I}$.

Obtaining Principal Components

- Consider the construction of a new set of variables

$$\mathbf{z} = \mathbf{u}'\mathbf{y}. \tag{47}$$

- Then we can show that

$$\begin{aligned} V(\mathbf{y}) &= V(\mathbf{u}'\mathbf{y}) && (48)\\ &= \mathbf{u}'V(\mathbf{y})\mathbf{u} && (49)\\ &= \mathbf{u}'\Sigma\mathbf{u} && (50) \end{aligned}$$

- Making use of the fact that $\mathbf{u}$ is orthogonal

$$\begin{aligned} V(\mathbf{z}) &= \mathbf{u}'\mathbf{u}\mathbf{D}\mathbf{u}'\mathbf{u} && (51)\\ &= \mathbf{I}\mathbf{D}\mathbf{I} && (52)\\ &= \mathbf{D} && (53) \end{aligned}$$

- The principal components are orthogonal to each and the variance of the principal components are the eigenvalues of Σ

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Why run PCA?
  1. Your advisor told you to and doesn't know the difference between PCA and FA.
  2. To orthogonalize a large set of predictors.
  3. To cut down the number of predictors to a smaller number while retaining most of the variance. This can be useful in some settings.
- But keep in mind that PCA assumes perfectly measured variables and is simply a transformation of data. It is a mathematical and not statistical procedure and it is atheoretical!

The Common Factor Model

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The **Common Factor Model** (aka exploratory factor analysis) makes a distinction between the common variance (shared among variables) and unique variance.
- What is required is some kind of measure of common variance.
- The usual starting point is the squared multiple correlation of a variable with the other $p - 1$ variables.
- The idea is to compute the squared multiple correlations and insert them into the diagonal of the sample correlation matrix $\mathbf{R}$ and then subject the correlation matrix to principal axis factoring.

Statistical Estimation in the Unrestricted Model

- PCA and the common factor model do not assume statistical distributions of the data.
- ML and GLS are statistical approaches that assume an underlying distribution.
- With maximum likelihood estimation of the common factor model one can test the hypothesis that there are k common factors that underlie the data.
- Recall that the large sample distribution of the likelihood ratio test is chi-square with degrees-of-freedom given by the difference in the number non-redundant elements in $\boldsymbol{\Sigma}$ and the number of parameters in the common factor model that need to be estimated - the so-called“free” parameters.

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

Rotation in the Unrestricted Model

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- A fundamental problem of EFA is that we can rotate the factor model an infinite number of ways and come up with the same values for the covariance matrix.
- This is known as the problem of “identification” or “rotational indeterminacy”.
- Indeterminacies are removed for identification purposes. Without dealing with the identification problem, no solution would be possible.
- However, because the initial set of estimates are completely arbitrary, another set of k^2 restrictions can be chosen by rotating the initially unrotated solution.

Rotation in the Unrestricted Model

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- For this, we use an orthogonal transformation matrix $\mathbf{T}$. For the $k = 2$ case, this matrix is of the form

$$\mathbf{T} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{54}$$

- If we have two factors, and we want to keep the factors uncorrelated (orthogonal), this means $\mathbf{\Phi} = \mathbf{I}$ leaving $k(k-1)/2$ indeterminacies to be resolved.

- Choosing $k(k-1)/2 = 1$ angle of rotation solves the problem. That angle is $90°$ which keeps the factors uncorrelated.

Rotation in the Unrestricted Model

Introduction

Path Analysis

Factor Analysis

Linear factor model

Principal Components Analysis

PAF

Estimation of unrestricted model

Rotation

CFA

Factor scores

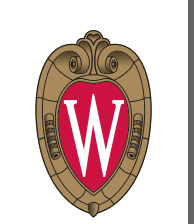
Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

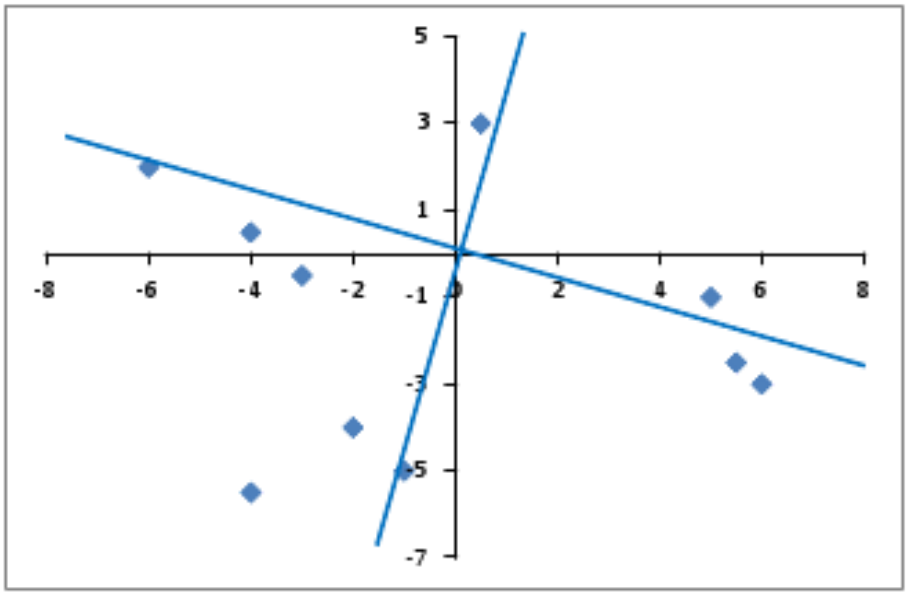
- The decision to rotate the solution usually rests on a desire to achieve a simple structure representation of the factors (Thurstone, 1935; Thurstone, 1947a).
- Simple structure criteria are designed to reduce the complexity of the variables - that is the number of factors that the variable is loaded on.



Orthogonal rotation: Varimax

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM

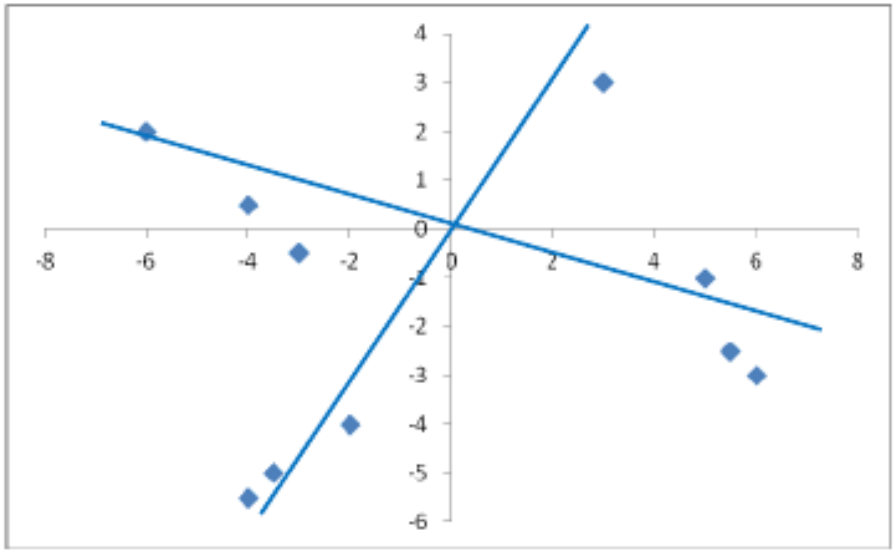
- The basic idea behind the varimax criterion is that after rotation, the resulting loadings on a factor should be either large or small relative to the original loadings.



Oblique rotation: Promax

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- In reality, factors tend to be correlated.
- One approach to the problem would be to first orthogonally rotate to a set of loadings, say $\mathbf{\Lambda_1}$ using varimax.
- Then, find a new set of loadings, say $\mathbf{\Lambda_2}$, corresponding to a new $\mathbf{\Phi}$, say $\mathbf{\Phi_2}$, such that $\mathbf{\Phi_2}$ has unit diagonal elements and non-zero off-diagonal elements.
- The result is an oblique solution yielding correlations among the factors contained in $\mathbf{\Phi_2}$. This is **PROMAX**.



Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

Orthogonal or Oblique Rotation?

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Orthogonal or Oblique Rotation?
  - Is orthogonal rotation ever preferred?
  - Recall that promax starts with an orthogonal rotation.
  - If factors are truly uncorrelated, then the factor correlation will be zero. Otherwise not.
  - Orthogonal rotation force the factors to be uncorrelated.
  - But zero is a valid correlation coefficient value.
  - So, oblique rotation will tell you if your factors are uncorrelated, but orthogonal rotation will not tell you if your factors are correlated.

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

See file: “FactorAnalysis.html

Confirmatory Factor Analysis

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The term “confirmatory factor analysis” reflects a difference in the number and position of the restrictions imposed on the factor space (Joreskog, 1969).
- In the unrestricted solution we saw that identification is achieved by imposing k^2 restrictions on the model.
- Because those restrictions are arbitrary, the solution can also be rotated to achieve simple structure.
- Regardless of the rotation, the factor model will yield the same fit to the observed covariance matrix.

The nature of the restrictions

- In a restricted solution, by contrast, usually more than k^2 restrictions are imposed.
- These restrictions are imposed on the elements of $\mathbf{\Lambda}$ in a manner that reflects an a priori hypothesis of simple structure.
- As a result, it is not possible to rotate the restricted model because doing so would destroy the positioning of the restrictions and hence the hypothesis under study.
- In addition to the k^2 restrictions imposed on the model as described in previous sections, it is also necessary to fix the metric of the latent variables.

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- This can be accomplished in two ways.
  1. Fix the variances of the latent variables to 1.0.
  2. Set the scale of each factor to the scale of one of its indicators

$$\lambda_{ij} = 1.0 \tag{55}$$

- Once the scale of the latent variables is determined, the next step is to decide on the pattern of fixed and freed loadings in the model.
- The fixed loadings are those that are usually fixed to zero (except, as noted before, when the need is to set the scale).
- The fixed loadings represent a priori hypotheses regarding the simple structure underlying the model.

Identification in CFA

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM

- Two-indicators/one factor: Not identified
- Three indicators/one factor: Just identified
- Two indicators/two factors: Over-identified with 1 df

Testing the CFA Model

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The difference between the unrestricted model and the restricted model lies with issues of model testing.
- Maximum likelihood estimation of the parameters in the restricted model proceeds in much the same way as estimation in the unrestricted case.
- Unlike the unrestricted case where we test the null hypothesis that there exists k common factors underlying the data, here the likelihood ratio chi-square is used to test the null hypothesis that the specified pattern of fixed and free loadings holds in the population.

Testing the CFA Model

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- This hypothesis implies that there are not only *k* common factors, but that a particular simple structure describes the relationship between the variables and the factors.
- The additional restrictions beyond the k^2 restrictions necessary to obtain a unique solution will result in greater degrees-offreedom compared to the unrestricted model.
- The degrees-of-freedom are obtained by subtracting the total number of estimated parameters, say *t*, from the $1/2q(q+1)$ distinct elements in $\mathbf{\Sigma}$.
- In addition to a global test of whether the restricted model holds in the population, one can also test hypotheses regarding the individual fixed and freed parameters (loadings, error variances, and factor variances and covariances) in the model.

Factor Scores

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The goal of factor analysis is typically to determine the underlying structure of a set of variables or items.
- But the factor model specifies a set of factor scores that serve as predictors.

$$\mathbf{y} = \mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\delta}, \tag{56}$$

- What happened to the factor scores $\boldsymbol{\eta}$?

Factor scores

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- Generally, factor scores are not considered directly in the estimation. The goal is usually to test a structure under a presumed number of factors.
- What if want to know what the scores of a sample would be if we could have measured them on factors directly?

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- To obtain factor scores, typically some kind of regression model is used.
- Note that the joint distribution of $\mathbf{y}$ and η is normal with the mean of $\mathbf{y} = \mathbf{0}$ and the mean of $\boldsymbol{\eta} = \mathbf{0}$, and

$$\boldsymbol{\Sigma} = Cov(\mathbf{y}\mathbf{y}') = \boldsymbol{\Lambda} E(\boldsymbol{\eta}\boldsymbol{\eta}')\boldsymbol{\Lambda}' + E(\boldsymbol{\delta}\boldsymbol{\delta}') \tag{57}$$

$$= \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_\delta. \tag{58}$$

- It can be shown that the regression of $\boldsymbol{\eta}$ on $\mathbf{y}$ is

$$\boldsymbol{\eta} = \boldsymbol{\Phi}\boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_\delta)^{-1}\mathbf{y} \tag{59}$$

$$= \boldsymbol{\Phi}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{60}$$

Factor scores

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- With parameter estimates in hand (from ML or GLS), the factor scores can be estimated as

$$\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Phi}}\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x} \tag{61}$$

or

$$\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Theta}}_{\delta}^{-1}\hat{\boldsymbol{\Lambda}})^{-1}\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Theta}}_{\delta}^{-1}\mathbf{x} \tag{62}$$

depending on the method of estimation.

Factor scores

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

- The difficulty is that the parameters are treated as fixed at their estimated values and so the sampling variability is not taken into account.
- Also, $\hat{\eta}$ is a non-linear function of the parameter estimates and hence the distribution of $\hat{\eta}$ could be quite complex.
- Factor scores are not rotationally invariant. Different rotations might give different factor scores (though CFA addresses this because there is no rotation). **TREAT FACTOR SCORES CAUTIOUSLY**
- Bayesian approaches that treat η as missing data and use a form of imputation can solve this problem.

CFA CAUTION

- Use CFA with caution!
  - CFA places a lot of restrictions (zero loadings) on a model, thus increasing degrees-of-freedom.
  - This is a problem because EFAs clearly show that cross loadings are not zero.
  - Models with lots of degrees of freedom are much less likely to be consistent with data and hence often leads to reporting misfitting models especially in large sample sizes (remember $N * F_{ML}$
  - This can be a problem, especially when EFAs show very nice simple structure.
  - This problem will come back to haunt us in SEMs.

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

Introduction
Path Analysis
Factor Analysis
Linear factor model
Principal Components Analysis
PAF
Estimation of unrestricted model
Rotation
CFA
Factor scores
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM

See file “cfa.lavaan.html”

Statistical Assumptions

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Structural equation modeling requires that certain underlying assumptions be satisfied in order to ensure accurate inferences.
- We will consider the following assumptions
  - Multivariate normality,
  - Completely random missing data,
  - Sufficiently large sample size,
  - Correct model specification.
  - Simple random sampling

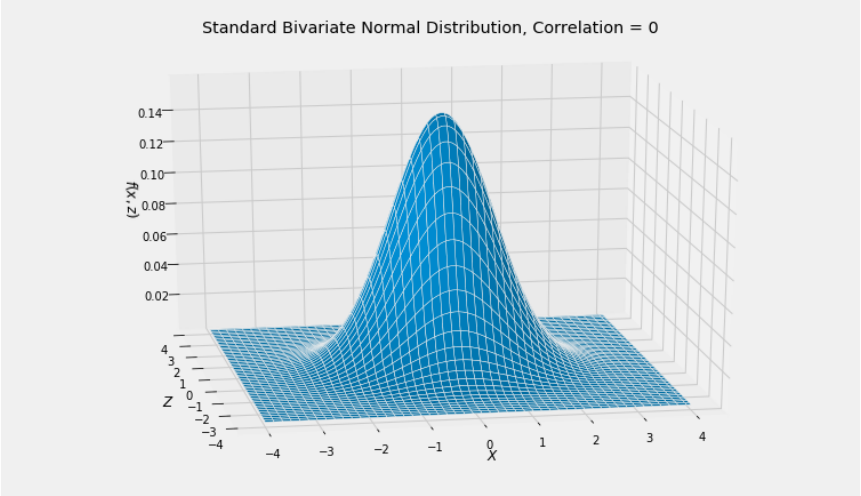
The Normality Assumption

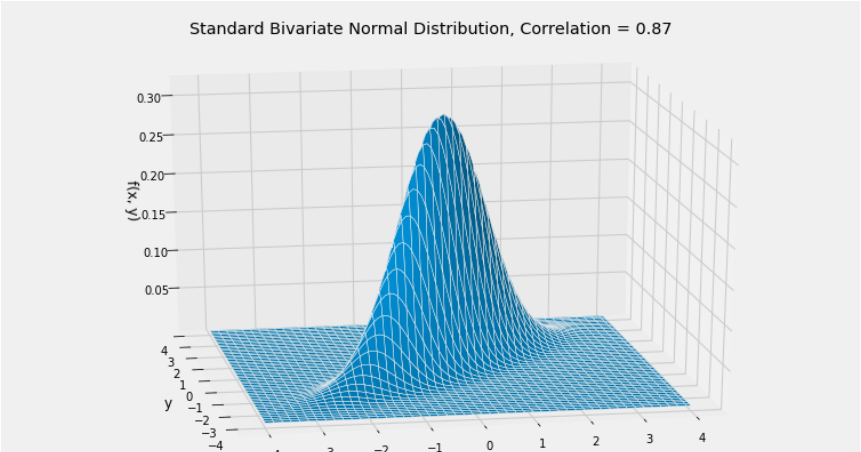
Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

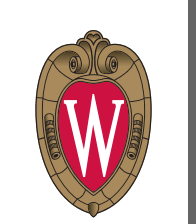
- A basic assumption underlying the standard use of structural equation modeling is that the observations are drawn from a continuous and multivariate normal population.
- This assumption is particularly important for maximum likelihood estimation because, as we saw the ML estimator is derived directly from the expression for the multivariate normal distribution.
- The Effects of Non-Normality
  - No effect on parameter estimates
  - Standard errors are downward bias (leading to increased Type I errors)
  - Goodness of fit test is inflated (dependent in part on the df of the model)

Intuition

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical







Intuition

Introduction

Path Analysis

Factor Analysis

Statistical Assumptions

The Normality Assumption

Sampling assumptions

Missing Data

Specification error

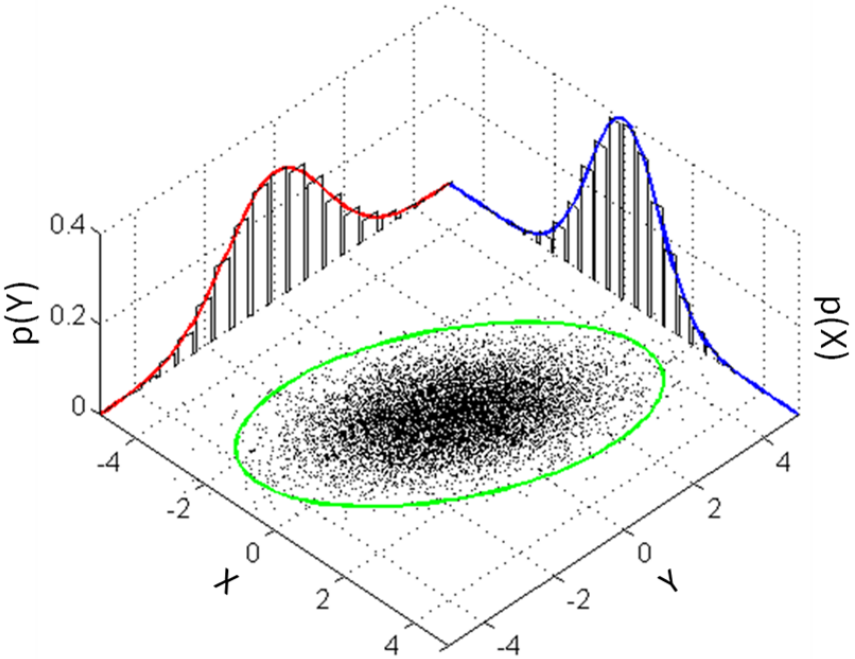
SEM

Multiple Group Modeling

Multilevel SEM

Growth Curve Modeling

Categorical



Estimation with Continuous Non-Normal Data

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- In both the continuous and categorical cases, the historical approach to estimation under non-normality utilized a class of discrepancy functions based on generalized least squares.
- We will refer to this class of estimators as weighted least squares (WLS).
- The WLS discrepancy function can be written as

$$F_{WLS} = (\mathbf{s} - \boldsymbol{\sigma})' \mathbf{W}^{-1} (\boldsymbol{s} - \sigma), \tag{63}$$

- For the function in to be a proper discrepancy function, $\mathbf{W}$ must be invertible.

Estimation with Continuous Non-Normal Data

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- The key characteristic of WLS estimation is the choice of an appropriate weight matrix $\mathbf{W}$.
- Consider an element of the vector $\mathbf{s}$ denoted as s_{ij}, the covariance of variable i and variable j.
- The expected value of the sample covariance element can be written as

$$E(s_{ij}) = \sigma_{ij} \tag{64}$$

- The general form of the asymptotic covariance matrix of the covariances is

$$(N-1)acov(s_{ij}, s_{gh}) = \sigma_{ig}\sigma_{jh} + \sigma_{ih}\sigma jg + \frac{N-1}{N}\kappa_{ijgh} \tag{65}$$

- The weight matrix formed of these elements is the ADF estimator

Estimation with Continuous Non-Normal Data

- The results of early studies of the ADF estimator were somewhat mixed.
- Browne (1982) found ADF estimates to be biased.
- Muthén and Kaplan (1985; 1992), on the other hand, found very little bias in ADF estimates. In all cases, the ADF chi-square was smaller than ML chi-square when applied to continuous non-normal data.
- When ADF was applied to categorical data, Muthén and Kaplan (1992) found that the ADF chi-square was markedly sensitive and that this sensitivity increased as the size of the model increased.
- ADF standard errors were noticeably downward biased, becoming worse as the model size increased.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Estimation with Continuous Non-Normal Data

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- A problem with ADF was the computational difficulties encountered for models of moderate size. Specifically, with p variables there are $u = 1/2p(p+1)$ elements in the sample covariance matrix $\mathbf{S}$. The weight matrix $\mathbf{W}$ is of order $u \times u$. Therefore, the size of the weight matrix grows rapidly with the number of variables.

- So, if a model were estimated with 20 variables, the weight matrix would contain 22,155 distinct elements.

Estimators for Categorical Variables

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Consider the following categorical variable

$$y_i = \begin{cases} C_i = 1 & \text{if } \nu_{i,C-1} < y_i^*; \\ C_i = 2 & \text{if } \nu_{i,C-2} < y_i^* \leq \nu_{i,C-1}; \\ \vdots & \\ 1 & \text{if } \nu_{i,1} < y_i^* \leq \nu_{i,2}; \\ 0 & \text{if } y_i^* \leq \nu_{i,1}. \end{cases} \tag{66}$$

- If it is reasonable to assume that continuous and normally distributed y^* variables underlie the categorical y variables, then from classic psychometric theory a variety of latent correlations can be specified.

Estimators for Categorical Variables

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Observed versus Latent Correlations

x-Variable Scale	y-Variable Scale	Observed Correlation	Latent Correlation
Continuous	Continuous	Pearson	Pearson
Continuous	Categorical	Pearson	Polyserial
Continuous	Dichotomous	Point-Biserial	Biserial
Categorical	Categorical	Pearson	Polychoric
Dichotomous	Dichotomous	Phi	Tetrachoric

Assumption of an underlying continuous variable for categorical or dichotomous variables

Estimators for Categorical Variables

- The first step in Muthén's approach is to estimate the thresholds for the categorical variables using maximum likelihood.
- In the second step, the latent correlations (e.g. tetrachoric correlations) are estimated.
- Finally, in the third step, a consistent estimator of the asymptotic covariance matrix of the latent correlations is obtained and implemented in a weighted least squares estimator.
- Muthén's approach is quite flexible insofar as any combination of categorical and continuous observed variables can be present in the data.
- The only requirement is the assumption that the categorical variables are associated with continuous normally distributed latent response variables.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Robust Methods

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Robust Maximum Likelihood (MLR)
  - Assumes continuous data but works well with categorical data
  - Computes standard errors using so-called ”sandwich approach”
  - Scales the chi-square test with a correction factor to help with non-normality
  - Considered best practice and available in lavaan.

Robust Methods

- Diagonally weighted least squares (WLSMV)
  - Does not assume continuous variables but assumes normal $y*$ variables
  - Does not require large sample sizes
  - WLSMV first obtains polychoric correlations using ML. The parameter estimates are then obtained from the estimated asymptotic variances of the polychoric correlation and threshold estimates used in a diagonal weight matrix.
  - Standard errors are obtained as in MLR.
  - Chi-square is scaled so as to make it closer to the reference chi-square distribution with proper df.
  - Available in lavaan and considered best practice for categorical data in SEM.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

See file: “Assumptions Examples/EP960.non-normality.html”

Sampling Assumptions

- For purposes of proper estimation and inference, a very important question concerns the sampling mechanism that yielded the sample data.
- In the absence of explicit information to the contrary, estimation methods such as maximum likelihood assume that data are generated according to simple random sampling.
- More often than not however, structural equation models are applied to data that have been obtained through a mechanism other than simple random sampling, e.g. multi-stage sampling.
- Properly accounting for multi-stage sampling designs requires the application of sampling weights at each level of analysis as well as special methods for obtaining correct variance estimates.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

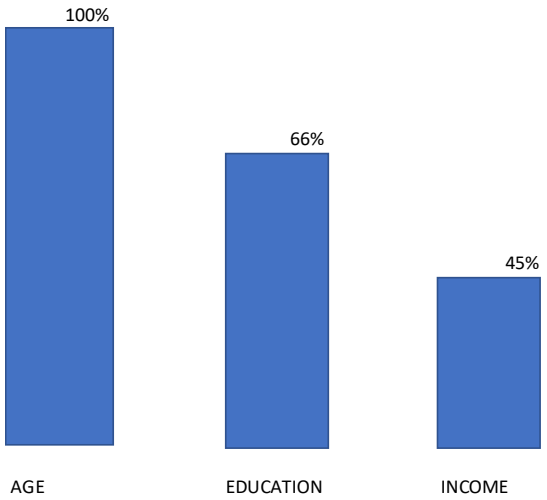
Missing Data

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Missing data is a very common in applied research.
- We will focus first on how missing data has been conceptualized – relying heavily on the seminal work of Little and Rubin (2002).
- We will then overview some general methods for handling missing data.
- Finally, we will examine Bayesian approaches to the analysis of missing data.

Missing Data

- Consider responses to three items on a parent questionnaire.
- These items ask for the parent's age, education level, and household income.
- For a sample of parents, we might observe the following frequency of responses on these three items.



Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Missing Data

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Little and Rubin (2002) considered three possible scenarios that would give rise to missing (and observed) data in this example.
- Let M be a missing data indicator, taking the value of 1 if the data are observed, and 0 if the data are missing.
- Let y be the complete data, y_{obs} represent observed data and y_{miss} represent missing data.
- Let ϕ be the scalar or vector-valued parameter describing the process that generates the missing data.

Missing Data

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- The following three forms of missing data are possible.

$$\text{MCAR: } f(M|y) = f(M|\phi),$$

$$\text{MAR: } f(M|y) = f(M|y_{obs}, \phi),$$

$$\text{NMAR: } f(M|y) = f(M|y_{obs}, y_{miss}, \phi),$$

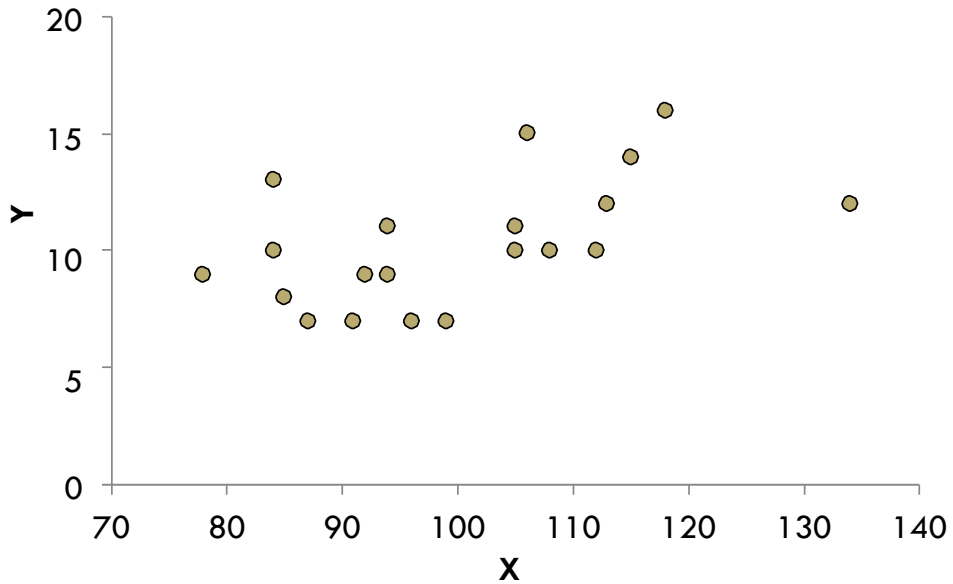
Ad hoc deletion methods

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- We consider the following ad hoc approaches to missing data:
  1. Listwise deletion
  2. Pairwise deletion
  3. Mean imputation
  4. Regression imputation
  5. Stochastic regression imputation
  6. Hot deck imputation

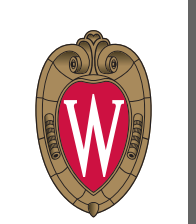
Complete Data

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical



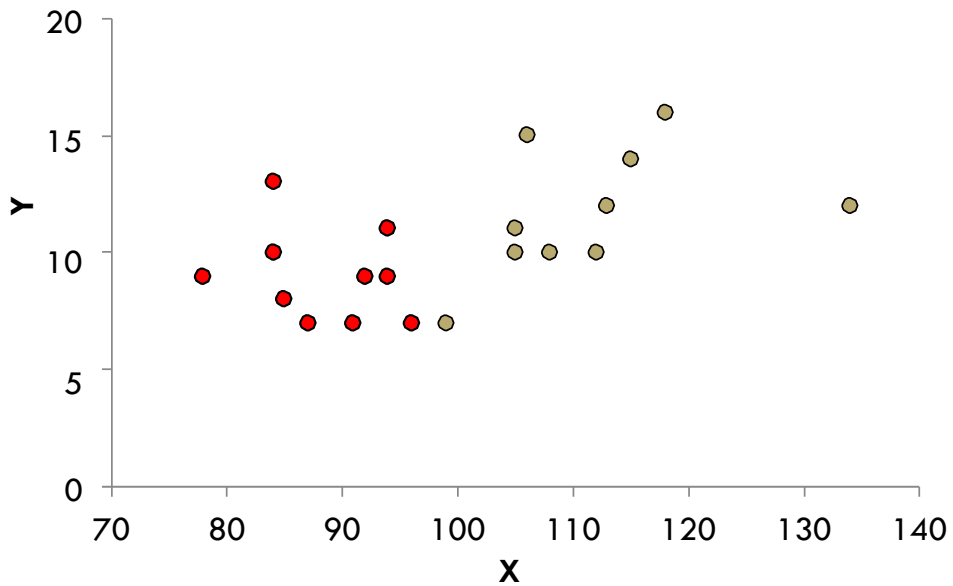
5

Figure: Complete Data



Missing values

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical



6

Figure: Missing values in red

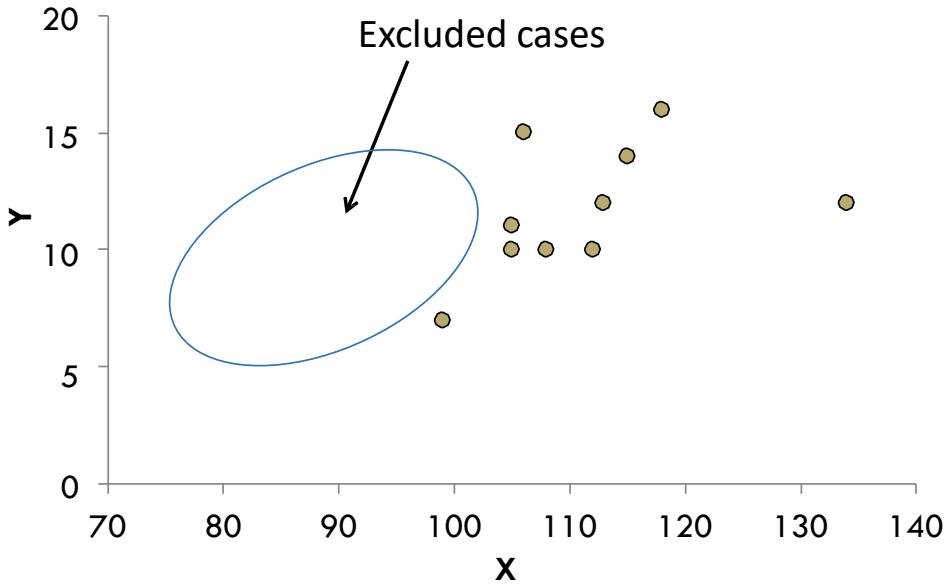
Listwise deletion

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Listwise deletion involves omitting any respondent for which there is any missing data on any variable to be used in an analysis.
- Listwise deletion is used as a convenient method for handling missing data, but suffers from two critical flaws.
  1. From the frequentist framework, listwise deletion can result in an unacceptable loss of data, hence reducing statistical power.
  2. Listwise deletion assumes that the missing data are MCAR.

Listwise Deletion

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical



8

Figure: Listwise deletion

Pairwise deletion

- Pairwise deletion is the use of listwise deletion when statistical methods require using pairs of variables in the estimation.
- The simplest example is the calculation of a covariance matrix among a set of variables.
- There are three fundamental flaws with the use of pairwise deletion.
  1. Pairwise deletion rests on the unrealistic assumption that the missing data are MCAR.
  2. Pairwise deletion can result in a loss of power due to the decrease in sample size.
  3. Sample size is not uniform across covariance calculations and thus the covariance matrix no longer will follow a Wishart distribution and, in some cases, might not be positive definite.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Single imputation methods

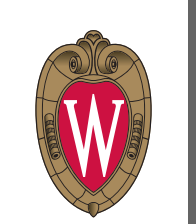
Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- It might be preferable to impute values for the missing data. Two types of imputations are possible.
  1. Imputing one single value for each missing data point. Methods to be discussed next are *mean imputation*, *regression imputation*, and *stochastic regression imputation*.
  2. Multiple imputation which will set the stage for Bayesian methods.

Mean imputation

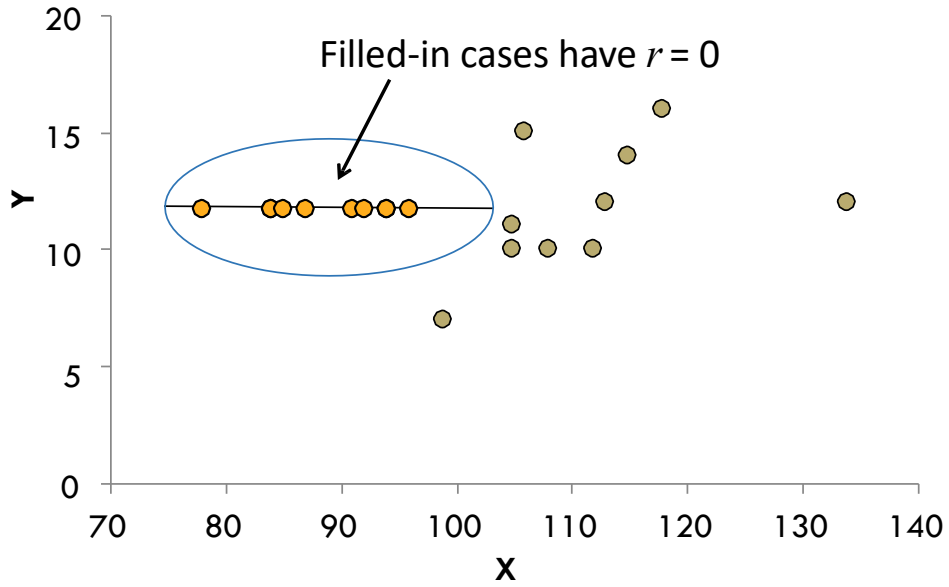
Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Mean imputation for a given variable y_k, $(k = 1, 2, \ldots, K)$, requires that we calculate the mean $\bar{y}_k$ and insert that value for all cases that are missing on y_k.
- Mean imputation suffers from two problems.
  1. Mean imputation assumes that the missing data are MCAR.
  2. A constant $\bar{y}_k$ is imputed for all occurrences of missing data on y_k. In the situation with a large amount of missing data, mean imputation will result in a loss of variation on that variable.



Mean Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical



9

Figure: Mean Imputation

Regression Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Regression imputation begins by specifying a regression equation to predict the missing data on each variable.
- Using listwise deletion to start, a regression equation is formed for the missing data on income and the missing data on education.

$$educ_i = \beta_0 + \beta_1(age_i) + \beta_2(income_i) + e_{i,educ}, \tag{67}$$

and

$$income_i = \beta_0 + \beta_1(age_i) + \beta_2(educ_i) + e_{i,income}. \tag{68}$$

Regression Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- From here, predicted values of education and income are obtained as

$$\widehat{educ_i} = \hat{\beta}_0 + \hat{\beta}_1(age_i) + \hat{\beta}_2(income_i), \tag{69}$$

and

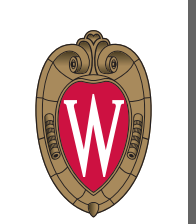
$$\widehat{income_i} = \hat{\beta}_0 + \hat{\beta}_1(age_i) + \hat{\beta}_2(educ_i), \tag{70}$$

and these predicted values are imputed for the corresponding missing data point.

Regression Imputation

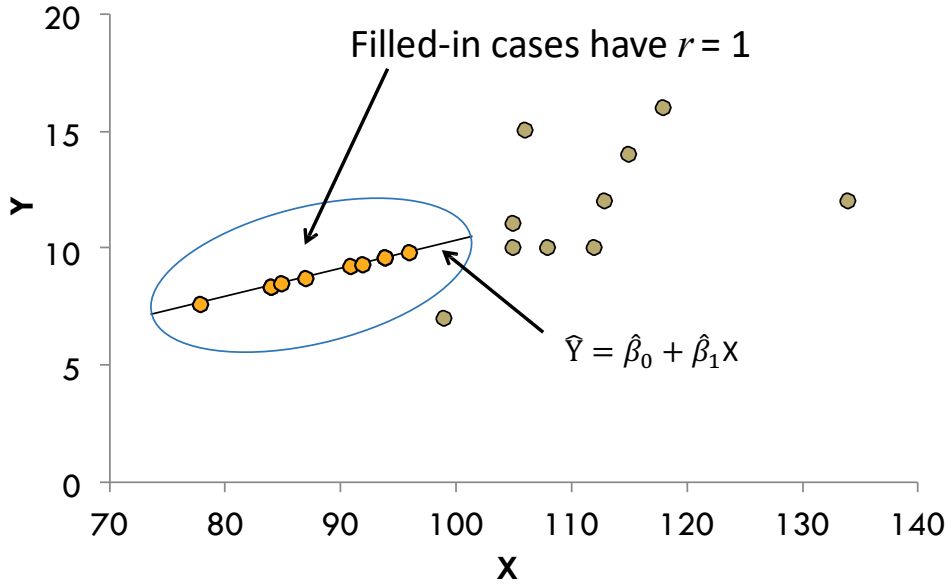
Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Regression imputation suffers from two major drawbacks.
  1. Predicted values of the missing data based on the regression in equations (69) and (70) will lie exactly on the regression line. Correlations among the variables of interest will be 1.0, and overall R^2 value will be overestimated.
  2. It is presumed that the imputed values would be the ones observed had there been no missing data. For this to be true, the regression model would have to been correctly specified.



Regression Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical



10

Figure: Regression Imputation

Stochastic Regression Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- To prevent the problems with regression imputation, we can add a random term to equations (69) and (70) that serve to *perturb* the imputed values from the predicted line.

$$\widehat{educ}_i = \hat{\beta}_0 + \hat{\beta}_1(age_i) + \hat{\beta}_2(income_i) + r_{i,educ}, \tag{71}$$

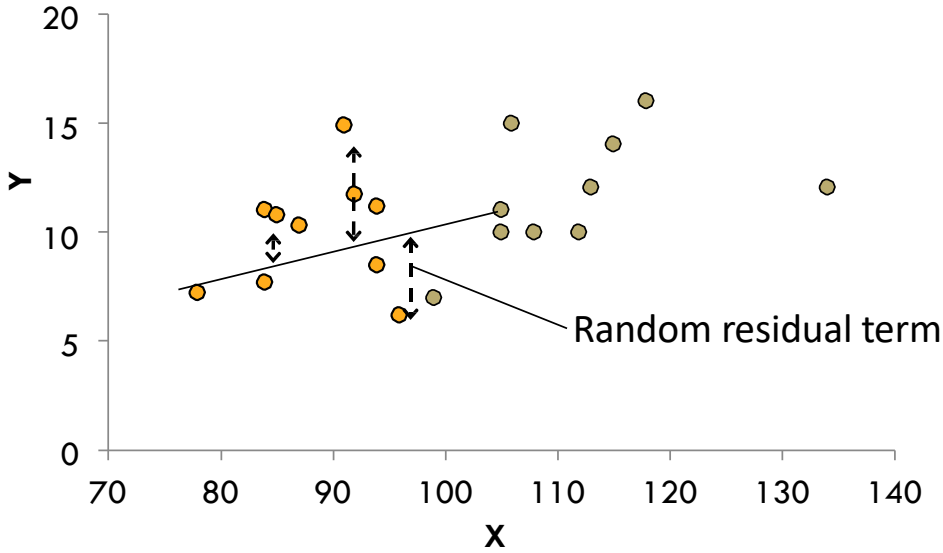
and

$$\widehat{income}_i = \hat{\beta}_0 + \hat{\beta}_1(age_i) + \hat{\beta}_2(educ_i) + r_{i,inc}, \tag{72}$$

where $r_{i,educ}$ and $r_{i,inc}$ are values drawn from a normal distribution with mean 0 and variance equal to the residual variance obtained estimation of equation (67).

Stochastic Regression Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical



23

Figure: Stochastic Regression Imputation

Stochastic Regression Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Stochastic regression imputation preserves the variability in the data that is lost with mean imputation and regression imputation.
- It has been shown that stochastic regression imputation yields unbiased estimates under the somewhat more realistic assumption of MAR.
- However, it should be noted that stochastic regression imputation only applies one draw from the normal distribution to be used for residual terms in equations (71) and (72).

Matching methods: Hot deck imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Hot deck imputation involves replacing the missing value on a variable from some individual by a value from a “similar” individual who, in fact, did respond to the variable.
- In the simplest case, a sample of respondents would be classified across a variety of selected demographic characteristics.
- Then, a missing value would be replaced by a draw from the subsample of individuals who are similar as possible on the selected demographic characteristics as the individual who omitted the response.

Hot deck imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Hot deck is not limited to categorical variables such as most background demographic characteristics.
- Continuous variable can also be used for hot deck matching, and algorithms such as *nearest neighbor, maximum deviation, or Mahalanobis distance* matching can be employed.
- Let, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})'$ be a set of K covariates $k = 1, 2, \ldots, K$ measured on individual i, and let y_i by a variable that is missing for individual i. Then, an indicator variable $d(i, j)$ can be formed as

$$d(i,j) = \begin{cases} 0, & i,j \text{ in same cell} \\ 1, & i,j \text{ in different cells} \end{cases} \tag{73}$$

Matching methods: Predictive Mean Matching

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Regression imputation and hot deck matching sets the ground work for *predictive mean matching*.
- The essential idea is that a missing value is imputed by matching its predicted value based on regression imputation to the predicted values of the observed data on the basis a predictive mean metric.

$$d(i,j) = [\hat{y}(\mathbf{x}_i) - \hat{y}(\mathbf{x}_j)]^2. \tag{74}$$

- Once a match is found, the procedure uses the actual observed value for the imputation.

Multiple Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- For all of the imputation methods, it is still the case that a single value is being imputed and treated as though that is the value that would have been obtained if the missing data point had, in fact, been observed.
- The imputed missing data point based on a single imputation method does not account for uncertainty about the missing data point itself.
- Thus, rather than imputing a single value, it may be theoretically better justified to draw multiple plausible missing data values.
- This idea forms the basis of *multiple imputation*, which rests on Bayesian theory.

Multiple Imputation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- The reason for adopting a Bayesian perspective to missing data problems is that by viewing parameters probabilistically rather than fixed and specifying a prior distribution on the parameters of interest, the imputation method is *Bayesianly proper*.
- Imputations reflect uncertainty about the missing data as well as uncertainty about the unknown model parameters.
- Although the method of stochastic regression imputation described above has a Bayesian flavor, it is not Bayesianly proper insofar as it does not account for parameter uncertainty, but rather only uncertainty in the predicted missing data values.

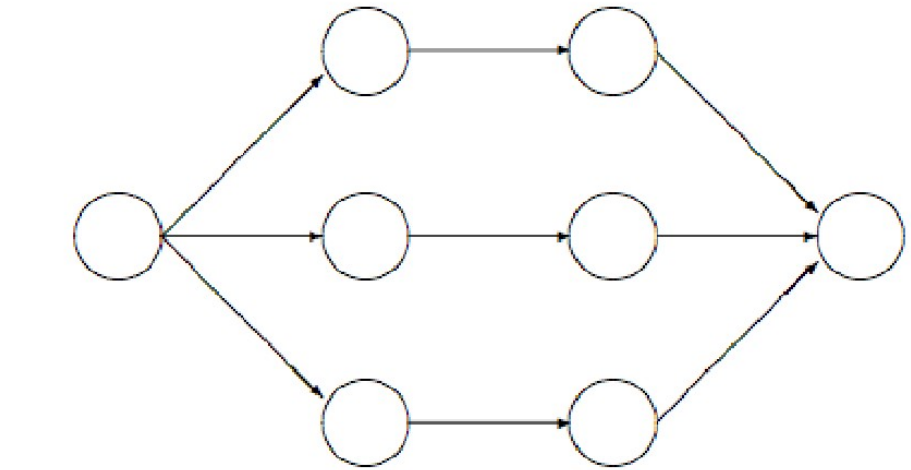
Phases of Multiple Imputation

1. Imputation Phase
   - Create multiple copies of the data, each with different imputed values for the missing data.
2. Analysis Phase
   - Perform statistical analysis on each data set.
3. Pooling Phase
   - Combine the collection of estimates and standard errors into a single set of results

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Multiple Imputation Phases

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical



15

Figure: Phases of multiple imputation

Imputation under joint normality: The EM algorithm

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

1. Initialize the parameters of the joint normal distribution. This includes the means, variances, and covariances.

2. E-step (Expectation Step):
   - For each observation with missing data, compute the expected values of the missing data given the observed data and the current parameter estimates.
   - Use the conditional distribution of the missing data given the observed data and parameters.

3. M-step (Maximization Step)
   - Update the parameters of the joint normal distribution using both the observed and imputed data.
   - Update the mean vector variances, and covariances, based on the observed and imputed data.

EM algorithm

4. Log-Likelihood Calculation
   - Compute the log-likelihood of the observed data and the imputed data based on the current parameter estimates.
5. Convergence and Iteration
   - If convergence has not been reached, repeat steps 2-4 until convergence is achieved.
6. Final Parameter Estimates
   - The final parameter estimates represent the maximum likelihood estimates of the parameters under the assumption of joint normality.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Data augmentation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Data augmentation is an algorithm used for multiple imputation.
- Under the assumption of multivariate normality of the data and ignorable missing data, the posterior distribution of the model parameters, given observed data can be written as

$$p(\theta|y_{obs}) \propto p(y_{obs}|\theta)p(\theta), \tag{75}$$

where $p(\theta)$ is the so-called *prior distribution* of the model parameter.

- DA algorithm is composed of two steps: The I(mputation) step and the P(osterior) step.

Data augmentation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- The I-step begins with an initial value of θ, denoted as θ^s. Then

$$\text{I -step:} \qquad \text{Draw } y_{miss}^{(s+1)} \text{ from } p(y_{miss}|y_{obs}, \theta^s).$$

- In other words, we use the current value $\theta^{(s)}$ and the observed data y_{obs} to generate a value for the missing data from the predictive distribution of the missing data $p(y_{miss}|y_{obs}, \theta^s)$.

Data augmentation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- The I-step is followed by the P-step which draws a new value of θ, namely $\theta^{(s+1)}$ from the posterior distribution of θ given the observed data y_{obs} and simulated missing data from the previous step, y_{miss}^{s+1}. Formally,

$$\text{P-step:} \qquad \text{Draw } \theta^{(s+1)} \text{ from } p(\theta|y_{obs}, y_{miss}^{(s+1)}).$$

- As the number of iterations goes to infinity, the DA algorithm converges to a draw from the posterior distribution.

- m separate *chains* of iterations are conducted simultaneously using different start values for θ to create m data sets. Typically, m should be greater than 10 and less than 20.

Chained Equations

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- The method of chained equations uses a series of single univariate imputations along with diagnostic checking rather than assuming a multivariate normal model for imputation that might be sensitive to specification issues.
- Univariate regression models consistent with the scale of the variable with missing data are used to provide predicted values of the missing data given the observed data.
- Once a variable of interest is “filled-in”, that variable, along with the variables for which there is complete data, is used in the sequence to fill in another variable.
- The order of the sequence is determined by the amount of missing data, where the variable with least amount of missing data is imputed first, and so on.

Chained Equations

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Once the sequence is completed for all variables with missing data, the posterior distribution of the regression parameters are obtained and the process is started again.
- Specifically, the filled-in data from the previous cycle, along with complete data are used for the second and subsequent cycles.
- The algorithm (called the Gibbs sampler) is used to generate the sequence of iterations.
- Running the Gibbs sampler with m chains provides m imputed data sets.
- This methods is implemented in the R program “mice”.

An Aside: The EM algorithm

- EM stands for *expectation-maximization* and is an algorithm that is widely used to obtain maximum likelihood estimates of model parameters in the context of missing data problems.
- First using arbitrary start values, the E-step creates the sufficient statistics necessary to obtain regression equations that yield the predictions of the missing data given the observed data and the initial set of model parameters.
- The M-step use the "filled-in" data to obtain new estimates of model parameters which is simply the use of straightforward equations to obtain new estimates of the vector of means and the covariance matrix of the data.
- The algorithm cycles between the E-step and the M-step until a convergence criterion has been met, at which point the maximum likelihood estimates have been obtained.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Full Information ML

- FIML requires that missing values to be at least MAR (i.e., either MAR or MCAR are ok).
- The process works by estimating a likelihood function for each individual based on the variables that are present so that all the available data are used.
- Individuals will have vectors of data of different length.
- Model fit information is derived from a summation across fit functions for individual cases.
- Under multivariate normality, FIML has been shown to produce unbiased parameter estimates and standard errors under MAR and MCAR (Enders & Bandalos, 2001).
- Less flexible than MI because of the strong normality assumption.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

See file: “Assumptions Examples/EP960.MissingData.html”

Specification Error

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Specification error is defined to be the omission of relevant variables in any equation of the system of equations defined by the structural equation model.
- To what extent do omitted variables affect inferences?
- In simple linear regression, specification errors in the form of omitted variables induces a correlation between the errors and exogenous variables in the model.
- The ordinary least squares estimators will no longer be unbiased, and the bias will depend directly on the size of the correlation between the errors and exogenous variables.

Studies of Specification Error

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

- Specification errors in the form of omitted variables can result in substantial parameter estimate bias (Kaplan (1988).
- For sampling variability, specification errors have been found to be relatively robust to small specification errors (Kaplan, 1989c).
- However, the z-test associated with free parameters in the model is affected in such a way that specification error in one parameter can propagate to affect the power of the ztest in another parameter in the model.
- Sample size, as expected theoretically, interacts with the size and type of the specification error.

Studies of Specification Error

- Specification error in one part of a model can propagate to other parts of the model.
- This error propagation was first noticed by Kaplan (1988) and studied more closely in a paper by Kaplan and Wenger (1993). We discovered that the propagation was dependent on the covariance matrix of the estimates, which, in itself, was model dependent.
- The form of the covariance matrix of the estimates is determined by the initial specification of the model.
- Each addition (or deletion) of parameters results in a change in the form of the covariance matrix of the estimates, and hence in the ways that specification errors will manifest themselves through the model.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
The Normality Assumption
Sampling assumptions
Missing Data
Specification error
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical

SEM: Putting It All Together

- SEM links restricted (confirmatory) factor analysis with path analysis
- The advantage is that the path coefficients are adjusted (disattenuated) for possible measurement error in the observed variables.
- The basic structural model

Structural Model

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{76}$$

$$\mathbf{y} = \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon} \tag{77}$$

$$\mathbf{x} = \boldsymbol{\Lambda}_{\boldsymbol{x}}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{78}$$

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Identification in SEM
Testing in SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian

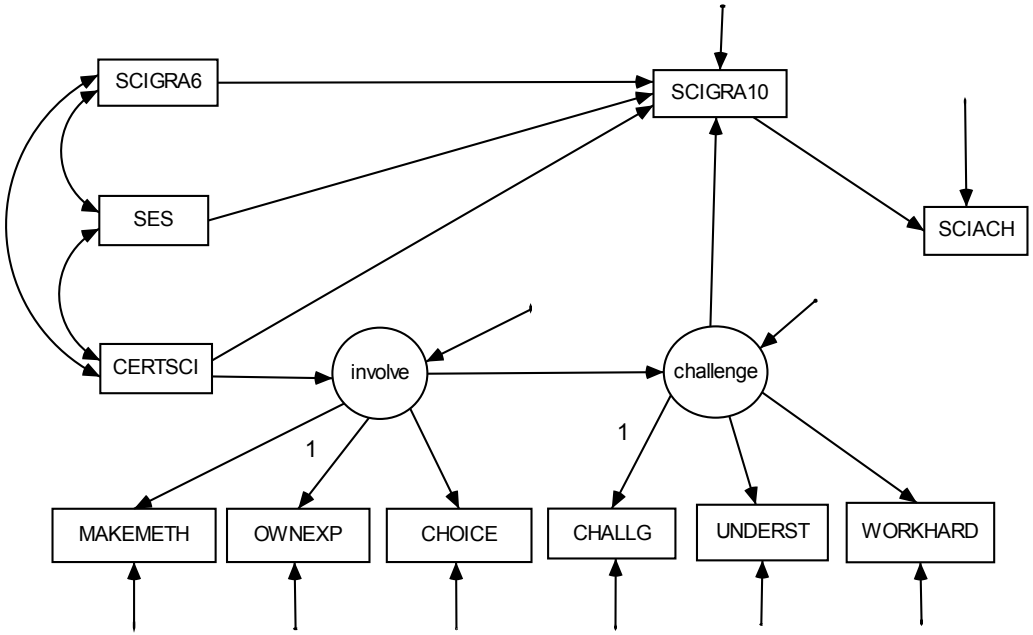


Figure 4.1 Expanded science achievement model.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Identification in SEM
Testing in SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Identification in SEM
Testing in SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian

- The covariance matrix therefore is

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_y & sym. \\ \mathbf{\Sigma}_{xy} & \mathbf{\Sigma}_x \end{bmatrix}$$

and in terms of model parameters

Covariance Structure

$$\mathbf{\Sigma(\Omega)} = \begin{bmatrix} \mathbf{\Lambda_y (I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi)(I - B)^{-1} \Lambda_y + \Theta} & \mathbf{\Lambda (I - B)^{-1} \Gamma \Phi \Lambda'} \\ \mathbf{\Lambda_x \Phi (I - B)^{-1} \Lambda'_y} & \mathbf{\Lambda_x \Phi \Lambda'_x + \Theta} \end{bmatrix}$$

Identification in SEM

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Identification in SEM
Testing in SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian

- When combining the measurement and structural models together into a single analytic framework, a set of new identification conditions can be added to those that have already been presented.
- The metric of the latent variables must be set as usual
- The counting rule must hold for the entire model.
- As an identification trick, treat the entire model as a CFA with restrictions on the factor covariance matrix.

Testing and Interpretation of Structural Equation Models

- The test of the fit of the model is now going to be based on many more degrees-of-freedom than usual.
- It is possible to partition the total degrees-of-freedom into those based on restrictions in the measurement part of the model, and those based on restrictions in the structural part of the model.
- The degrees-of-freedom from the measurement part of the model are usually greater than those from the structural part of the model.
- However, it is the structural part of the model that is typically the focus of substantive inquiry.
- It is possible to reject a relatively well-fitting structural model because of a poorly developed measurement model.

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Identification in SEM
Testing in SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Identification in SEM
Testing in SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian

See file: “SEM.lavaan.html”

Multiple Group Modeling: No Mean Structure

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- Suppose that an investigator wishes to understand the differences between public and private schools in terms of student perceptions of school climate.
- First, the investigator might wish to determine whether the measurement structure of the student perception items is the same way across school types.
- Second, the investigator may be interested in knowing if the means of the factors of school climate are different between public and private school students.

Multiple Group Modeling: No Mean Structure

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- Multiple Group Specification and Testing (after Jöreskog, 1971)
- Comparability of factor structure across groups

$$\mathbf{y}_g = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta}_g + \boldsymbol{\delta}_g. \tag{79}$$

- We assume that the samples are independent of one another, and also assume that the values of the variables are drawn from a multivariate normal population.

Multiple Group Modeling: No Mean Structure

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- First we test the equality of the number of factors without regard to the pattern of fixed and free loadings – *configural invariance*.

$$H_{0k} : k_1 = k_2 = \ldots = k_G \tag{80}$$

- This test is conducted as separate unrestricted factor analysis models.

- Next we test equality of factor loadings – *metric invariance*

$$H_{0\Lambda} : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \ldots = \mathbf{\Lambda}_G \tag{81}$$

- The test of factorial invariance is obtained by setting equality constraints across groups for common elements in the factor loading matrix $\mathbf{\Lambda}$ and allowing the remaining parameters to be free across groups.

Multiple Group Modeling: No Mean Structure

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- Next we test the equality of both the factor loadings and intercepts – *scalar invariance*

$$H_{0\Lambda\tau} : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \ldots = \mathbf{\Lambda}_G \tag{82}$$

$$\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \ldots = \boldsymbol{\tau}_G \tag{83}$$

- We can test other hypotheses, but scalar invariance is required for what comes next – assessment of factor mean differences across groups.

Bringing in the Means

- The next step in a program of research looking at track differences in self-concept and locus-of-control might be to consider if there are mean differences across tracks on the latent variables.
- To estimate the factor means, we now need to bring in the regression intercepts

$$\mathbf{y}_g = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta}_g + \boldsymbol{\delta}_g. \tag{84}$$

- We add the assumption

$$E(\mathbf{y}_g) = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g E(\boldsymbol{\eta}_g) + E(\boldsymbol{\delta}_g). \tag{85}$$
$$= \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\kappa}_g. \tag{86}$$

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

Bringing in the Means

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- The mean structure model is not identified.
- It is not possible to estimate the intercepts and all factor means at the same time. With 5 indicators and one factor, you have 5 intercepts and 1 factor mean but only 5 data means.
- Constraints must be imposed.
- An approach to the problem would be to set the factor mean of one of the groups to zero (for the one factor case).
- The remaining factor means estimates are interpreted as differences relative to group g.

Multiple Group Path Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- Multiple group modeling can be extended to the path analysis framework.
- By setting up a two group path model, one can explore the moderating effects of a grouping variable on the path coefficients in a structural model.
- Typically one sets all regression coefficients to be equal, and examines modification indices to see what needs to be relaxed.
- The path(s) that need to be relaxed indicate where moderation might be occurring.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

See files:
“mgcfa.lavaan.html”
“TwoGroup.PathAnalysis.html”

MIMIC Modeling

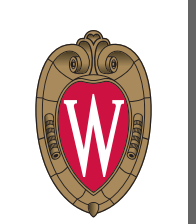
Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- MIMIC stands for the Multiple Indicators and Multiple Causes model and was proposed by Jöreskog and Goldberger (1975).
- Denote by $\mathbf{x}$ a vector that contains dummy codes representing group membership (e.g. 1 = public school; 0 = private school). Then, the MIMIC model can be written as

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta}, \tag{87}$$

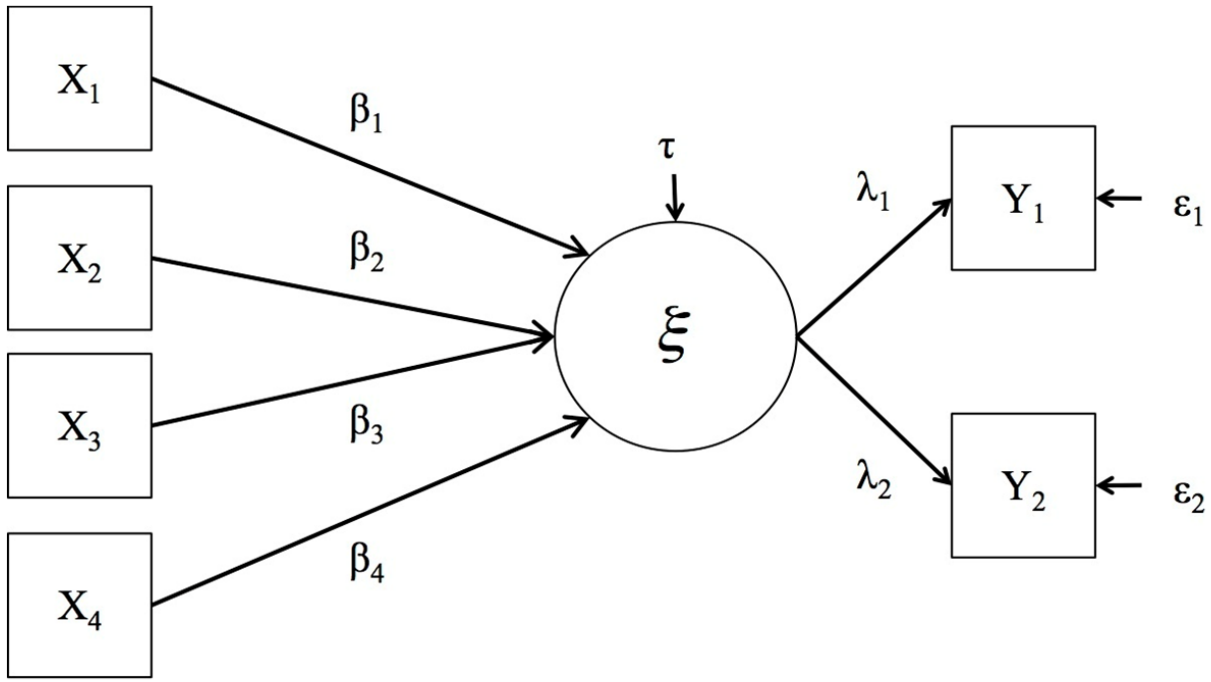
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}. \tag{88}$$

- Note that the final equation means that $\boldsymbol{\Lambda} = \mathbf{I}$ and $\boldsymbol{\epsilon} = \mathbf{0}$. This means that $\mathbf{x}$ has no measurement structure (but it could - and that would be a full SEM).



MIMIC model

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM



Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
MIMIC modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

See file: “MIMIC.lavaan.html”

Multilevel SEM

- Conventional SEM assumes simple random sampling and independent observations - namely

$$p(y_i|y_j) = p(y_i) \tag{89}$$

- This assumption is violated under more complex forms of sampling such as multistage clustered sampling.

- Three choices for addressing the problem

1. **Disaggregation of the data**: Ignores clustering; Leads to non-independent errors and biased regression coefficients.
2. **Aggregation**: Inflates relationships because of reduction of variance.
3. **Modeling the sampling structure** Appropriate for the actual structure; Leads to correct estimates and standard errors; HLM!!

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

Brief introduction to HLM

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- Also known as mixed-effects, random-effects models, random coefficient models, and covariance components models.
- The intercepts and slopes as outcomes model

$$y_{ig} = \beta_{0g} + \beta_{1g} x_{ig} + r_{ig} \tag{90}$$

where β_{0g} is the mean achievement score for group g and β_{1g} is the SES-achievement slope for school g.

- The interpretation of β_{0g} is dependent on the centering of x_{ig}.

Brief introduction to HLM

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- Modeling the intercepts and slopes as a function of level–2 variables

$$\beta_{0g} = \gamma_{00} + \gamma_{01} W_g + u_{0g}, \tag{91}$$
$$\beta_{1g} = \gamma_{01} + \gamma_{11} W_g + u_{1g}, \tag{92}$$

  where W_g represent the percent of students in school g on subsidized lunch programs.

- The mixed-effects specification of this model is

$$y_{ig} = \gamma_{00} + \gamma_{01} W_g + \gamma_{10} x_{ig} + \gamma_{11} W_g x_g + u_{0g} + u_{1g} x_{ig} + r_{ig}. \tag{93}$$

Multilevel SEM

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- What does this have to do with SEM?
- In recent years attempts have been made to integrate multilevel modeling with structural equation modeling in order to provide a general methodology that can account for issues of measurement error and simultaneity as well as multi-stage sampling.
- This can now be handled in a relatively straightforward fashion.

Multilevel SEM

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- Consider a model that decomposes a p-dimensional response vector $\mathbf{y}_{ig}$ for student i in school g into the sum of a grand mean $\boldsymbol{\mu}$, a between groups part $\boldsymbol{\nu}_g$ and a within groups part $\mathbf{u}_{ig}$. That is,

$$\mathbf{y}_{ig} = \boldsymbol{\mu} + \boldsymbol{\nu}_g + \mathbf{u}_{ig}, \tag{94}$$

  The total sample covariance matrix for the response vector can be written as $\mathbf{y}_{ig}$

$$\boldsymbol{\Sigma}_T = \boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_w, \tag{95}$$

  where $\boldsymbol{\Sigma}_T$ is the population total sample covariance matrix, $\boldsymbol{\Sigma}_b$ is the population between groups covariance matrix, and $\boldsymbol{\Sigma}_w$ is the population within groups covariance matrix.

Multilevel SEM

- Sample quantities can be defined as

$$\bar{\mathbf{y}}_{.g} = \frac{1}{n_g} \sum_{i=1}^{n_g} \bar{\mathbf{y}}_{ig} \tag{96}$$

$$\bar{\mathbf{y}} = \frac{1}{N} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \bar{\mathbf{y}}_{ig} \tag{97}$$

$$\mathbf{S}_w = \frac{1}{N-G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} (\mathbf{y}_{ig} - \bar{\mathbf{y}}_{.g})(\mathbf{y}_{ig} - \bar{\mathbf{y}}_{.g})' \tag{98}$$

$$\mathbf{S}_b = \frac{1}{G-1} \sum_{g=1}^{G} n_g (\bar{\mathbf{y}}_{.g} - \bar{\mathbf{y}})(\bar{\mathbf{y}}_{.g} - \bar{\mathbf{y}})', \tag{99}$$

where $\bar{\mathbf{y}}_{.g}$ is the sample mean for group g, $\bar{\mathbf{y}}$ is the grand mean, $\mathbf{S}_w$ is the sample pooled within group covariance matrix, and $\mathbf{S}_b$ is the between groups covariance matrix.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

Multilevel Factor Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- Let's assume that the vector of student responses can be expressed in terms of the multilevel linear factor model as

Multilevel Linear Factor Model

$$\mathbf{y}_{ig} = \boldsymbol{\nu} + \boldsymbol{\Lambda}_w \boldsymbol{\eta}_{w_{ig}} + \boldsymbol{\Lambda}_b \boldsymbol{\eta}_{b_g} + \boldsymbol{\epsilon}_{w_{ig}} + \boldsymbol{\epsilon}_{b_g}, \tag{100}$$

where $\mathbf{y}_{ig}$ was defined earlier, $\boldsymbol{\nu}$ is the grand mean, $\boldsymbol{\Lambda}_w$ is factor loading matrix for the within group variables, $\boldsymbol{\eta}_{w_{ig}}$ is a factor that varies randomly across units within groups, $\boldsymbol{\Lambda}_b$ is the between groups factor loading matrix, $\boldsymbol{\eta}_{b_g}$ is a factor that varies randomly across groups, $\boldsymbol{\epsilon}_{w_{ig}}$ and $\boldsymbol{\epsilon}_{b_g}$ are within and between group uniquenesses.

Multilevel Factor Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- Under the assumptions of linear factor analysis, here extended to the multilevel case, the total sample covariance matrix defined in Equation (95) can be expressed in terms of factor model parameters as

Total Sample Covariance Matrix

$$\boldsymbol{\Sigma}_T = \boldsymbol{\Lambda}_w \boldsymbol{\Phi}_w \boldsymbol{\Lambda}_w' + \boldsymbol{\Theta}_w + \boldsymbol{\Lambda}_b \boldsymbol{\Phi}_b \boldsymbol{\Lambda}_b' + \boldsymbol{\Theta}_b, \tag{101}$$

where $\boldsymbol{\Phi}_w$ and $\boldsymbol{\Phi}_b$ are the factor covariance matrices for the within group and between group parts and $\boldsymbol{\Theta}_w$ and $\boldsymbol{\Theta}_b$ are diagonal matrices of unique variances for the within group and between groups part.

Multilevel Factor Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- It is usually straightforward to specify a factor structure for the within school variables.
- It is also straightforward to allow for within school variables to vary between schools.
- Conceptual difficulties sometimes arise in warranting a factor structure to explain variation between groups.
- The fact that it is sometimes difficult to conceptualize a factor structure for the between groups covariance matrix does not diminish the importance of taking the between group variability into account when conducting a factor analysis on multilevel structured data.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

See file: “MultilevelCFA.html”

Multilevel Path Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- Conventional multilevel regression models may not be suited for capturing the structural complexity within and between organizational levels.
- For example, it may be of interest to determine if school level variation in student science achievement can be accounted for by school level variables.
- Moreover, one might hypothesize and wish to test direct and indirect effects of school level exogenous variables on that portion of student level achievement that varies over schools.
- These questions are important for a fuller understanding of educational systems and such questions can be addressed via multilevel structural equation modeling.

Multilevel Path Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- The model that we will consider allows for varying intercepts and varying structural regression coefficients.
- Earlier work on multilevel path analysis by Kaplan & Elliott (1997a) building on the work of Muthen & Satorra (1989) specified a structural model for varying intercepts only.
- This “intercepts as outcomes” model was applied to a specific educational problem in Kaplan & Elliott (1997b) and Kaplan & Kreisman (2000).
- Recent developments now allow for modeling structural slopes.

Multilevel Path Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- The within-school (level–1) full path model as

Multilevel Path Model

$$\mathbf{y}_{ig} = \boldsymbol{\alpha}_g + \mathbf{B}_g\mathbf{y}_{ig} + \boldsymbol{\Gamma}_g\mathbf{x}_{ig} + \boldsymbol{r}_{ig}, g = 1, 2, \ldots, G \tag{102}$$

- The level–2 model as

Level–2

$$\boldsymbol{\alpha}_g = \boldsymbol{\alpha}_{00} + \boldsymbol{\alpha}_{01}\mathbf{z}_g + \boldsymbol{\alpha}_{02}\mathbf{w}_g + \boldsymbol{\epsilon}_g, \tag{103}$$

$$\mathbf{B}_g = \mathbf{B}_{00} + \mathbf{B}_{01}\mathbf{z}_g + \mathbf{B}_{02}\mathbf{w}_g + \boldsymbol{\zeta}_g, \tag{104}$$

$$\boldsymbol{\Gamma}_g = \boldsymbol{\Gamma}_{00} + \boldsymbol{\Gamma}_{01}\mathbf{z}_g + \boldsymbol{\Gamma}_{02}\mathbf{w}_g + \boldsymbol{\theta}_g. \tag{105}$$

Multilevel Path Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- The full multilevel path model allows for a set of structural relationships among between school endogenous and exogenous variables, $\mathbf{z}_g$ and $\mathbf{w}_g$, which we can write as

$$\mathbf{z}_g = \boldsymbol{\tau} + \boldsymbol{\Delta}\mathbf{z}_g + \boldsymbol{\Omega}\mathbf{w}_g + \boldsymbol{\delta}_g, \tag{106}$$

where $\boldsymbol{\tau}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\Omega}$ are the fixed structural effects.

Multilevel Path Analysis

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

- We can substitute these expressions into each other and express $\mathbf{y}_{ig}$ as a function of a grand mean, the main effect of within-school variables, the main effect of between–school variables and the cross level moderator effects of between and within school variables.
- These reduced form effects contain the structural relations as specified in Equations (102) through (106).
- $\mathbf{w}$ could consist of variables that could be manipulated in the context of an experiment

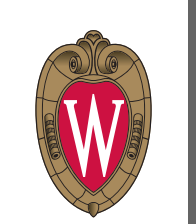
Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Multilevel Factor Analysis
Multilevel path analysis
Growth Curve Modeling
Categorical Latent Variables

See file: “MultilevelMIMIC.html”

Growth Curve Modeling

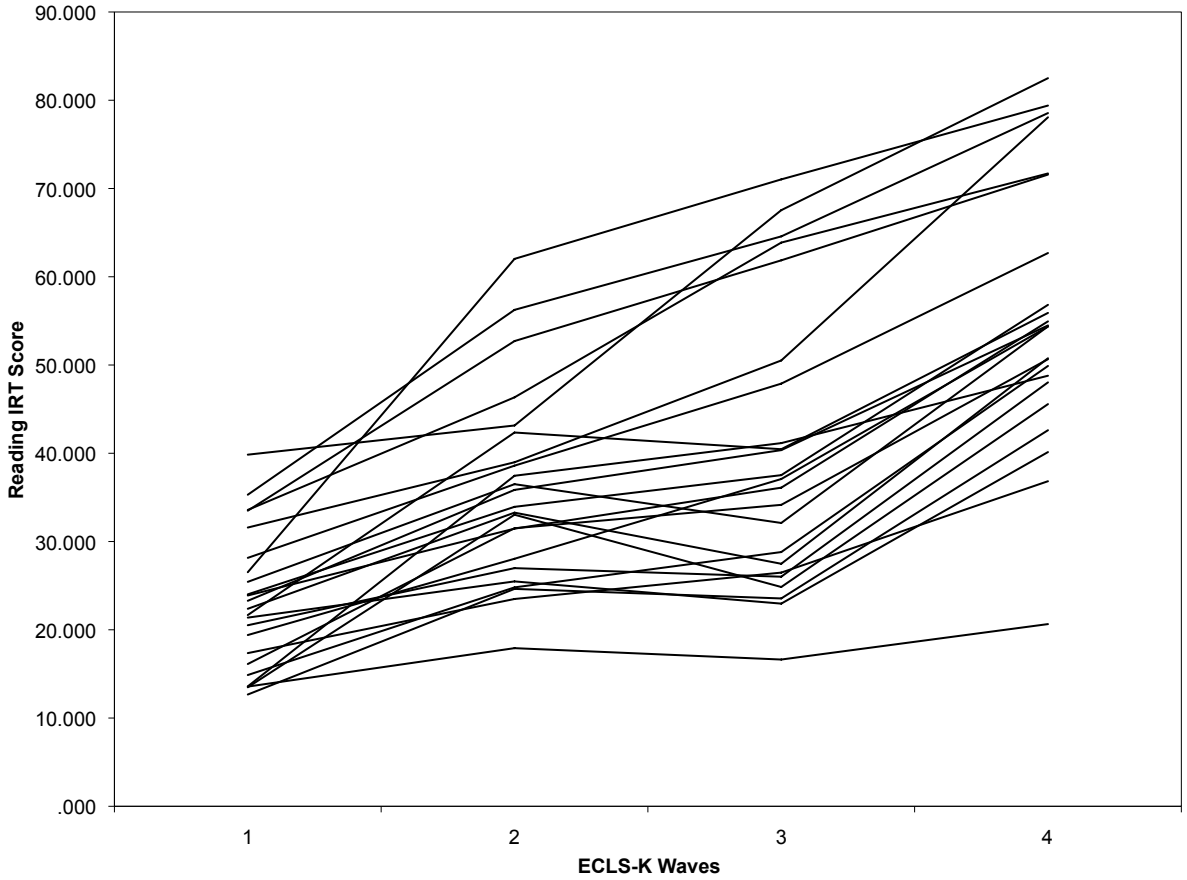
Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

- Many social and behavioral processes are dynamic - i.e. changing over time.
- Easier access to longitudinal data allows researchers to address an important class of substantive questions - namely the growth and development of social and behavioral outcomes over time.
- Growth curve modeling has been advocated for many years by researchers such as Bryk and Raudenbush (1992), Rogosa, Brandt, & Zimowski, (1982), and Willett (1988) (see also Willett & Sayer, 1994).



Change in Continuous Variables over Discrete Time

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables



GCM from the HLM Perspective

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

- Level – 1 (intra-individual differences)

Level–1 GCM

$$y_{it} = \pi_i + \pi_{1i} a_{it} + \epsilon_{it}, \tag{107}$$

- Quadratic growth can be incorporated into the model by extending the specification as

Quadratic GCM

$$y_{it} = \pi_i + \pi_{1i} a_{it} + \pi_{2i} a_{it}^2 + \epsilon_{it}. \tag{108}$$

GCM from the HLM Perspective

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

- The model can be further extended to handle predictors of individual differences in the initial status and growth trajectory parameters.
- Individuals are considered Level - 2.
- A single time-invariant predictor of initial status and growth for person i, denoted as x_i.

Time-invariant Level–2

$$\pi_{0i} = \mu_0 + \gamma_0 x_i + \zeta_{0i} \quad (109)$$

$$\pi_{1i} = \mu_1 + \gamma_1 x_i + \zeta_{1i} \quad (110)$$

GCM from an SEM Perspective

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

- Research by Muthén (1991) and Willett and Sayer (1994) have shown how the general growth model described in the previous section can also be incorporated into a structural equation modeling framework.
- We need a measurement model and a structural model.
- The measurement model is

$$\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon} \tag{111}$$

- The form of the measurement model for an example of 4 time points is.

$$\boldsymbol{\tau} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\pi}_0 \\ \boldsymbol{\pi}_1 \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

- The structural model takes the form of

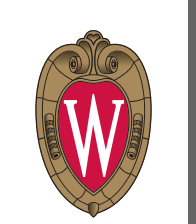
$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}. \tag{112}$$

- In the 4 wave example, the form is

$$\boldsymbol{\eta} = \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}, \boldsymbol{\alpha} = \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}, \mathbf{B} = \mathbf{0}, \boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

- The latent variable growth model can, however, be extended to include exogenous predictors of initial status and growth.

$$\mathbf{x} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \tag{113}$$



Growth Curve Model Diagrams

Introduction

Path Analysis

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

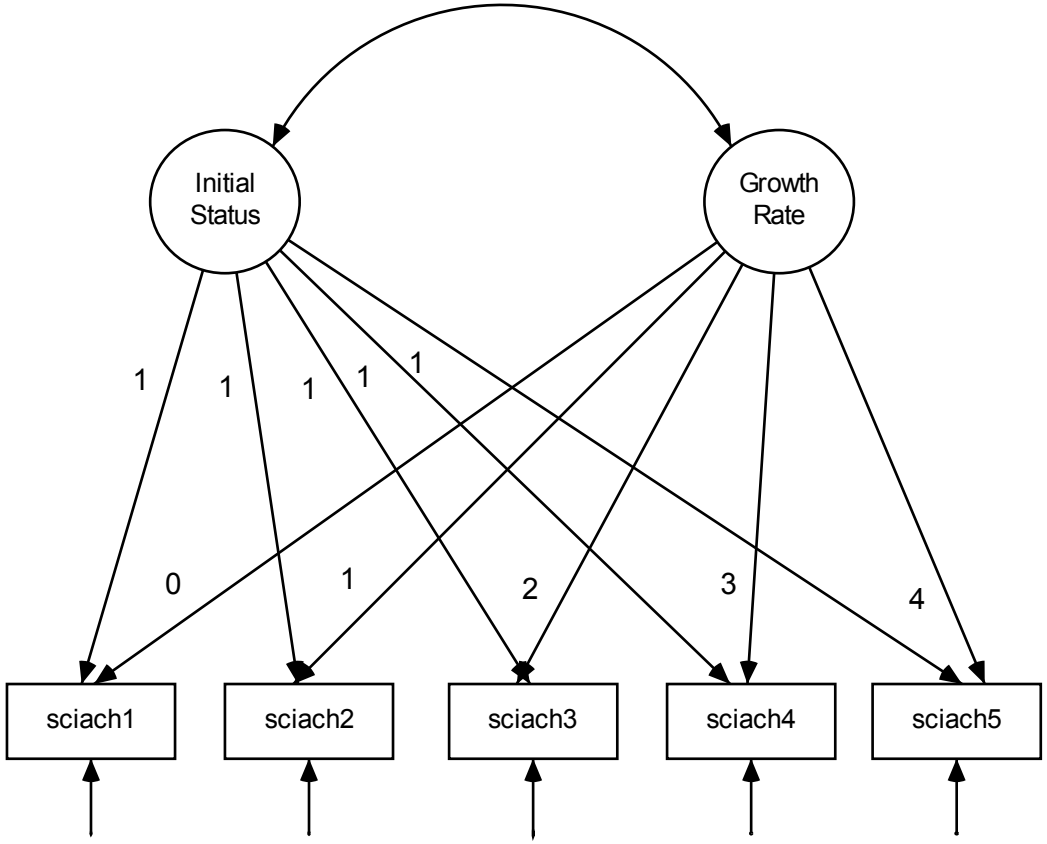
Growth Curve Modeling

GCM as HLM

GCM from SEM

Extensions of GCM

Categorical Latent Variables



Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

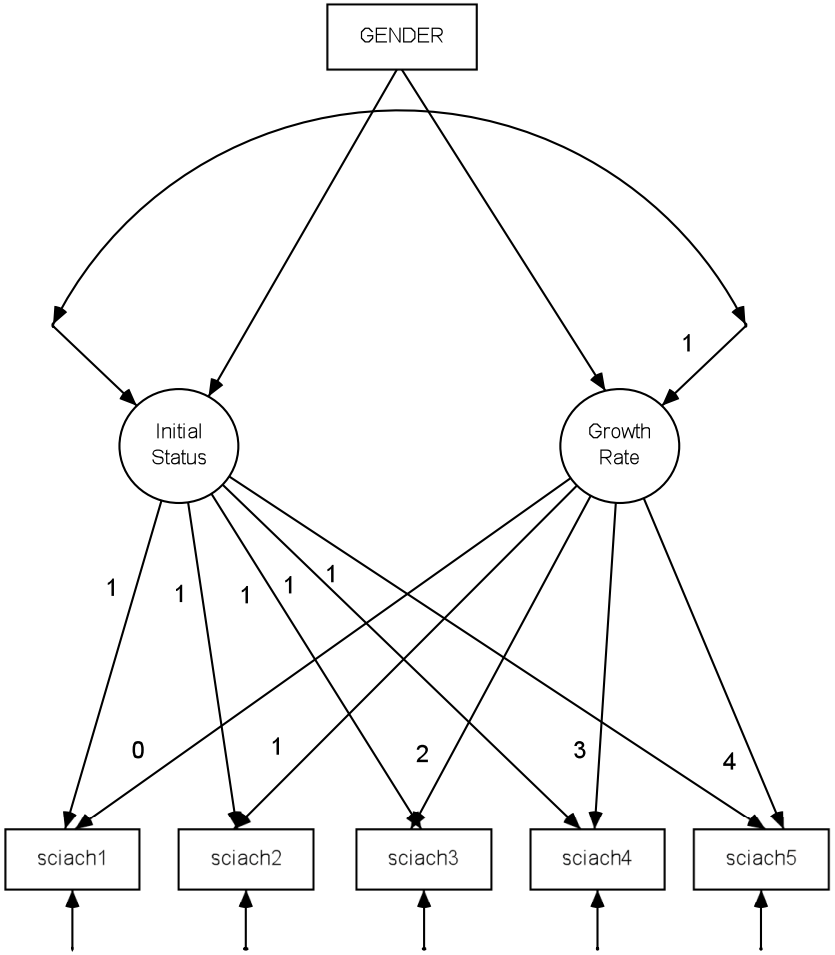
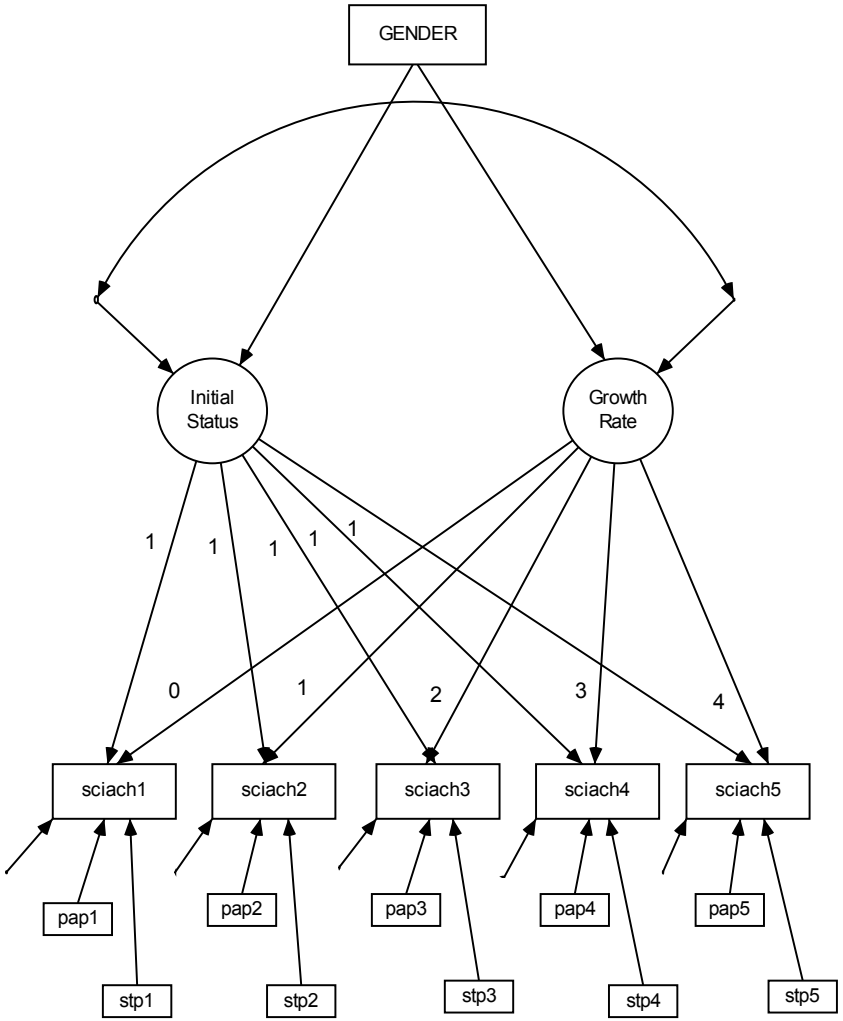


Figure 8.4 Growth curve model of science achievement with time invariant predictor



Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

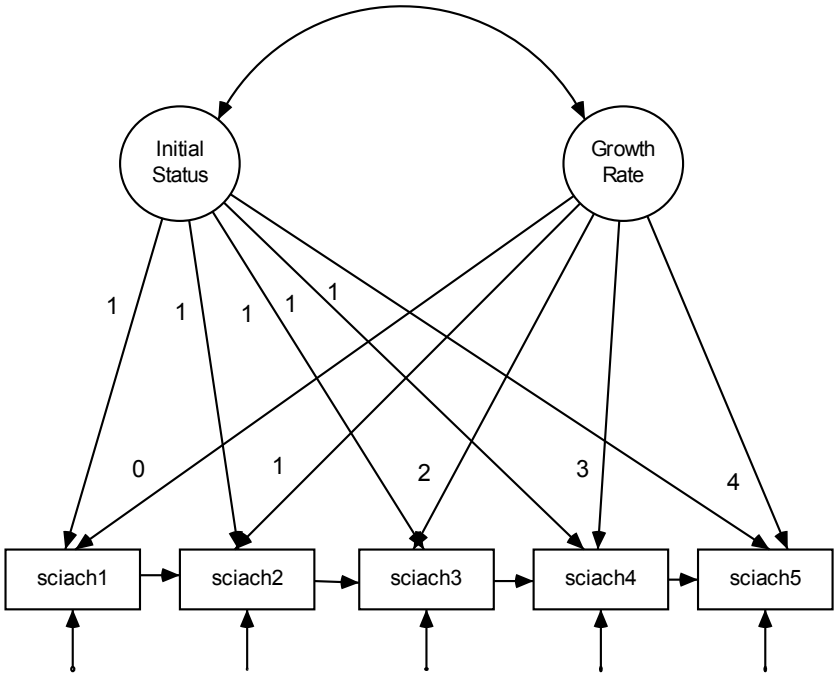


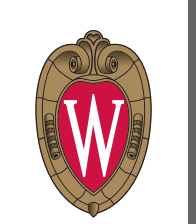
Figure 8.6 ALT(1) model of science achievement.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

Extensions of Growth Curve Modeling

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

- Binary Outcome Measures
- Multivariate Growth Curve Models
  1. Parallel growth processes
  2. Sequential growth processes
  3. Proximal and Distal Outcomes



Parallel growth process: Jordan et al. (2008) *Dev. Science*

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

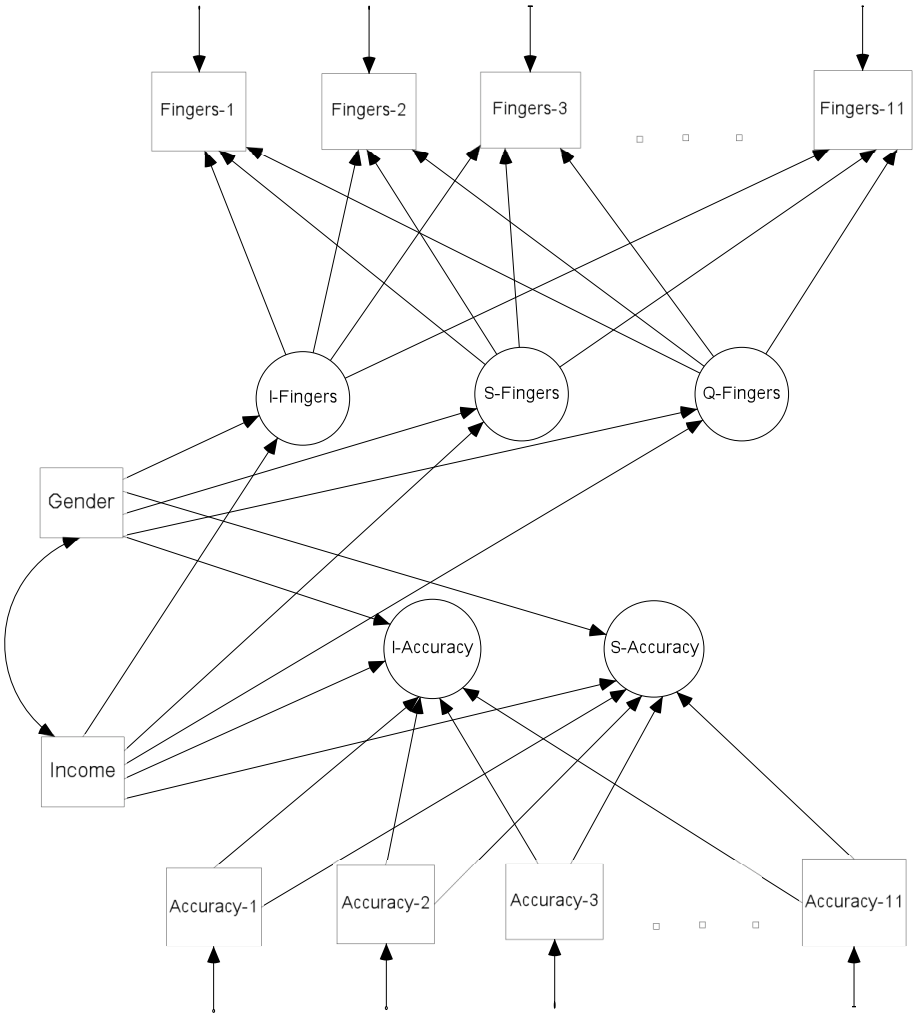
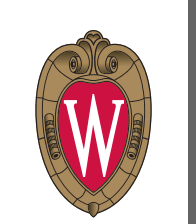
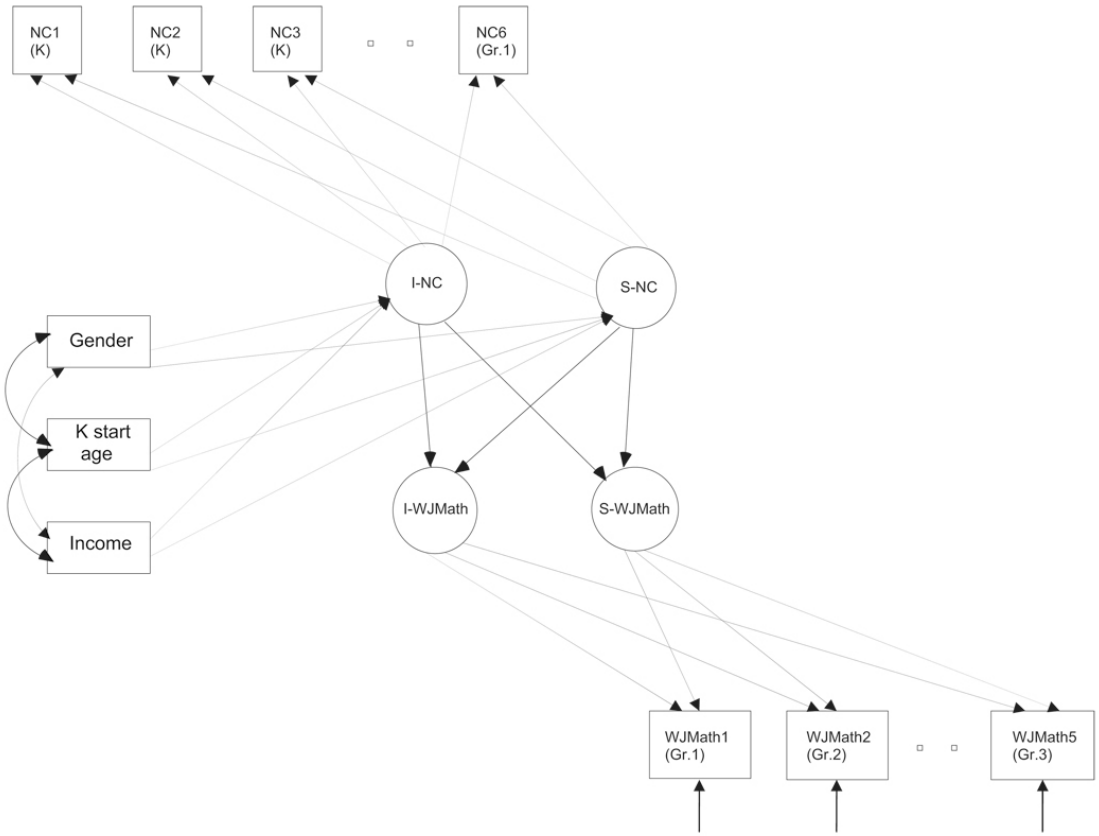


Figure 1. Path diagram of sequential growth curve model for finger use and number combinations. "Fingers" stands for frquencey of finger use and"Accuracy" stands for performance on number combinations.Note that for ease of reading, correlations among growth parameters are not presented but are estimated in the analysis.



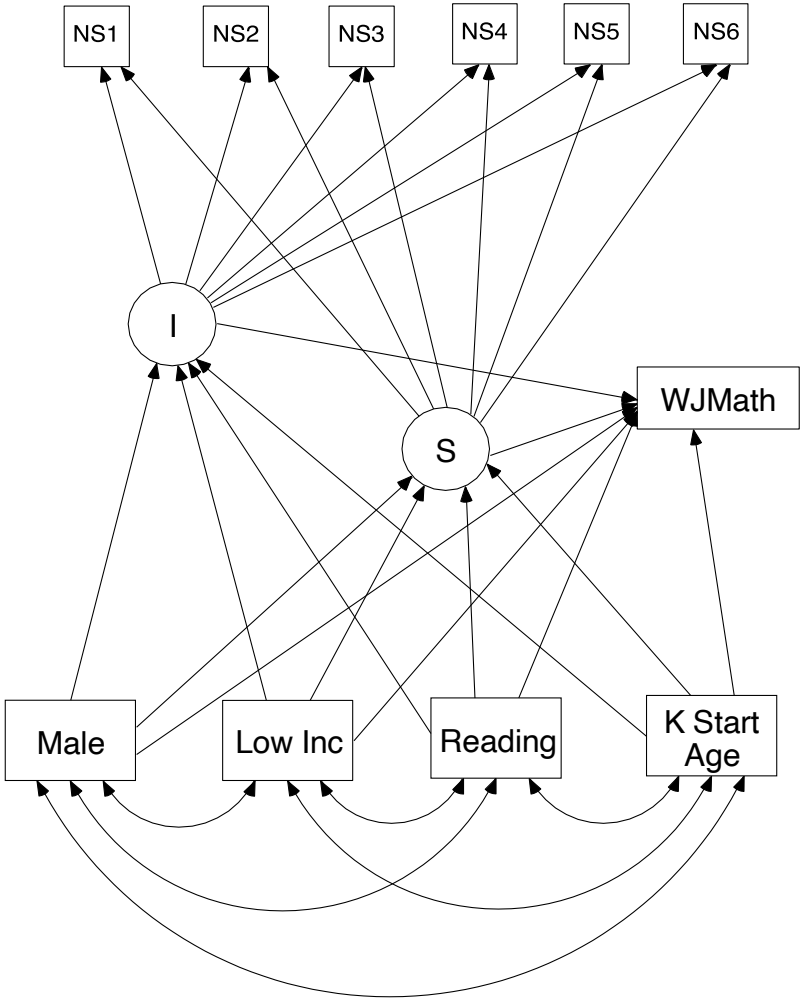
Sequential growth process: Jordan et al. (2008) *Dev. Psych.*

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables



Proximal outcome growth process: Jordan et al. (2007) *LDRP*

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables



Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

- Flexible modeling of non-linear growth
  - Estimating the variance of the quadratic term is sometimes difficult.
  - A better approach is to use the data to estimate the degree of non-linearity in the curve.
  - This is referred to as “latent-basis” growth curve modeling.
  - The trick is in the specification of the loading matrix.

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & * \\ 1 & * \end{bmatrix}$$

  - The *’s are free parameters. The first two fixed parameters identify the initial status and shape and then the last two provide estimates of the shape. If they are not equidistant, then it suggests non-linearity.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

- Flexible modeling of non-linear growth (cont'd)
  - Estimating percent change over time.
  - The trick again is in the specification of the loading matrix.

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & * \\ 1 & * \\ 1 & 1 \end{bmatrix}$$

  - Now, the estimates of the free parameters represent the percent change in the outcome over time.
- Keep in mind that latent-basis models are partially a data-based model.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
GCM as HLM
GCM from SEM
Extensions of GCM
Categorical Latent Variables

See file: “lavaan.GCM.html”

Latent Class Analysis

- In a traditional latent class analysis, it is assumed that an individual belongs to one and only one latent class, and that given the individual's latent class membership, the observed responses are independent of one another, referred to as the local independence assumption.
- The latent classes are, in essence, categorical factors arising from the pattern of response frequencies to categorical items, where the response frequencies play a role similar to that of the correlation matrix in factor analysis.
- The analog of factor loadings are parameters that estimate the probability of a particular response on the manifest indicator given membership in the latent class.
- Unlike continuous latent variables (i.e. factors), categorical latent variables (latent classes) divide individuals into mutually independent groups.

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

The latent class model

- The latent class model can be written as

Latent Class Model

$$P_{ijkl} = \sum_{c=1}^{C} \delta_c \rho_{i|c} \rho_{j|c} \rho_{k|c} \rho_{l|c} \tag{114}$$

where δ_c is the proportion of observations in latent class c. The parameters $\rho_{i|c}, \rho_{j|c}, \rho_{k|c}, \rho_{l|c}$ are the response probabilities for the items $ijkl$ conditional on class membership c.

- For example, $ijkl$ are the cells of a, say, 2 x 2 x 2 x 2 frequency table with i = 1 to 2. More generally, $i = 1, 2, \ldots, I$, and so on.

- The ρ's are the response frequencies, and δ_c is the proportion of observations in latent class c.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

LCA model estimation

- A method used to estimate the parameters of the latent class model is the EM algorithm.
- First, consider that we have data in the form of response frequencies and parameters under an assumed model (response probabilities conditional on the latent classes and the latent class proportions).
- Second, we consider the latent classes as missing data.
- The likelihood function for the presumed distribution of the data is said to be “incomplete” insofar as the latent classes are “missing”.
- The complete (log)-likelihood function involves both the likelihood of the response model and the latent classes.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

LCA model estimation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- The EM algorithm proceeds as follows:
  1. Begin with arbitrary starting values for the model parameters
  2. Find the expected value of the complete log-likelihood function given the data and the initial values of the parameters. This is the E-Step.
  3. Maximize the expected value found in the E-Step with respect to the model parameters. This is the M-Step.
  4. Insert the parameter values obtained in the M-Step back into the E-Step and repeat until a convergence criterion is obtained.

LCA model fit

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- Model fit is assessed by comparing the observed response proportions against the response proportions predicted by the model.
- A likelihood ratio statistic is obtained that is asymptotically distributed as chi-square.
- The degrees of freedom are calculated by subtracting the number of free parameters to be estimated from the total number of response patterns that are possible.

LCA model fit

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- Other methods of model assessment are available.
- The Akaike information criterion (AIC).
  - Allows for comparisons of different models specifying a different number of latent classes.
  - In terms of model comparison, the model with the lower AIC among a set of competing models is preferred from a predictive point of view.

LCA Model Evaluation

- Comparison of estimated and predicted class proportions
  - Each individual has a probability of being in a latent class.
  - The sum of the probabilities across individuals within each latent class gives the predicted class membership.
  - The same data can be used to uniquely assign an individual to a latent class based on the highest probability (then coded 1, the other classes coded zero for that individual).
  - The proportion of 1's in a latent class provides the estimated class population shares.
  - The predicted class memberships and estimated class population shares should be close, indicating a good solution.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

LCA Model Evaluation (cont'd)

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- Another method of LCA model evaluation is *entropy*.
- Entropy is a measure of class separation and is obtained as

$$H = -\sum_{i=1}^{n}\sum_{c=1}^{C} p_{ic}\log(p_{ic}), \tag{115}$$

  where n is the number of individuals in the sample, C is the number of latent classes, p_{ic} is the estimated probability of individual i belonging to class c.
- Lower entropy denotes better class separation. Note that an over-fit model will have low entropy so this should not be used for model selection.

LCA analysis steps

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- A reasonable approach to building LCA models is as follows:
  1. Test the model of complete independence among cells (log-linear model of independence).
  2. Build models with different numbers of latent classes.
  3. Compare models based on AIC, entropy, and substantive judgement.
  4. Add predictors to the final model and check on whether latent class structure changes

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

See file: “LCA.html”

Latent Markov Modeling

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- The question we are concerned with is how latent class membership might change over time.
- When considering changes in latent class membership, we consider latent classes as latent states.
- Examples:
  - Piaget's stages of cognitive development
  - Kohlberg's stages of moral development
- Change is conceptualized as movement from one state to another over discrete time.

LMM Brief Historical Background

- Change over time in qualitative states is based on the original work of A. A. Markov.
- It was Markov's work on the probability theory of chains involving discrete random variables that is of specific relevance to methods for studying latent transition analysis.
- A stochastic model specifying several states that an individual can occupy over time is called a "Markov process".
- A Markov process is a stochastic process in which the transition probabilities can be specified by knowing only the present state, and not the past history of the individual (Coleman, 1964, pg 23).
- The outcomes of interest are estimates of the transition probabilities of moving from one state to another.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

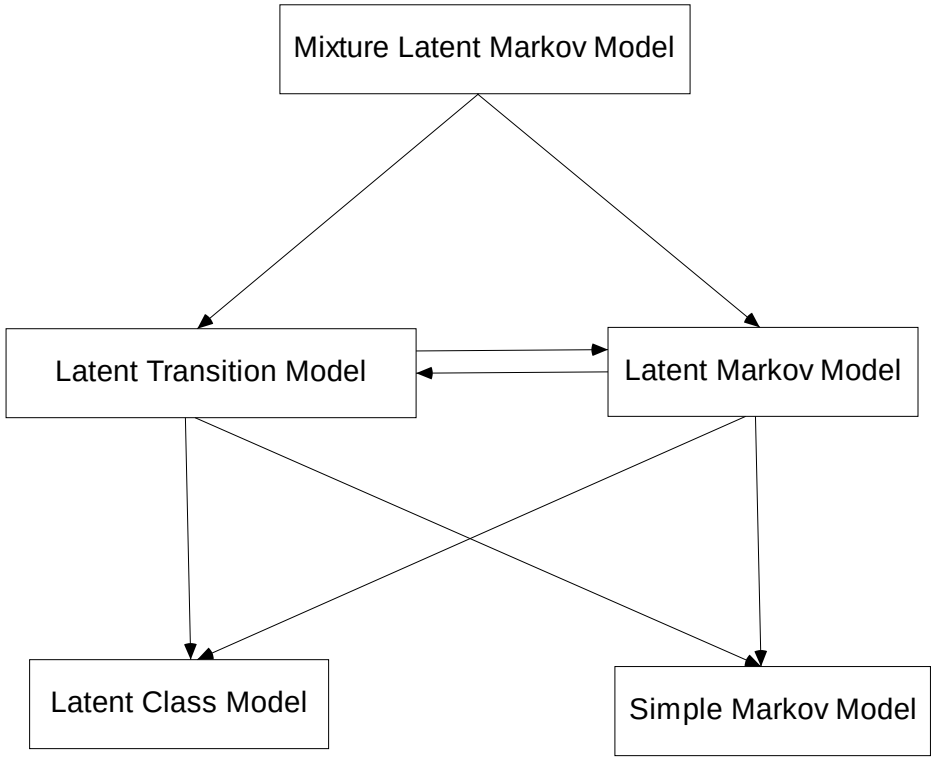
Types of Markov Models

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- There are a variety of Markov models that can be specified, ranging from very simple to very complex.
- A simple Markov model consists of a single chain where predicting the current state of an individual only requires data from the immediate past occasion.
- In a simple Markov model, it is assumed that the transition probabilities are the same for all individuals – an assumption that the population is homogenous.
- If it is assumed that the probabilities are the same over time, this is referred to as a “time-invariant” Markov model.
- Time varying Markov models can be defined as well.

Hierarchy of Markov Models

Markov chain models for developmental processes 40



Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

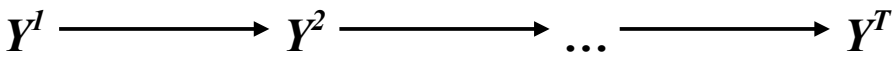


Figure 2. Basic Markov Model.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

Measurement Error in Markov Models

- It is possible to have measurement error in the categorical variable.
- This would be considered mis-classification.

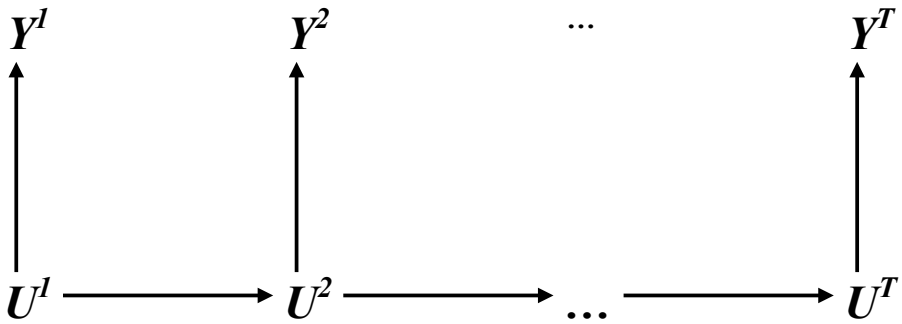


Figure 3. Basic Latent Markov Model.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- ECLS-K provides measures that can be used to test hypotheses regarding stage sequential change in reading competencies.
- Prior studies of stage sequential studies reading
  - Chall (1995)
  - Juel (1988)
- Neither used advanced models for estimation.
- Kaplan & Walpole (2005; using ECLS-K). First to use Markov modeling in reading.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

See file: “LTA.letterrec.k2.html”
See file: “LTA.3T.K3.html”

Extensions

Introduction

Path Analysis

Factor Analysis

Statistical Assump-tions

SEM

Multiple Group Modeling

Multilevel SEM

Growth Curve Modeling

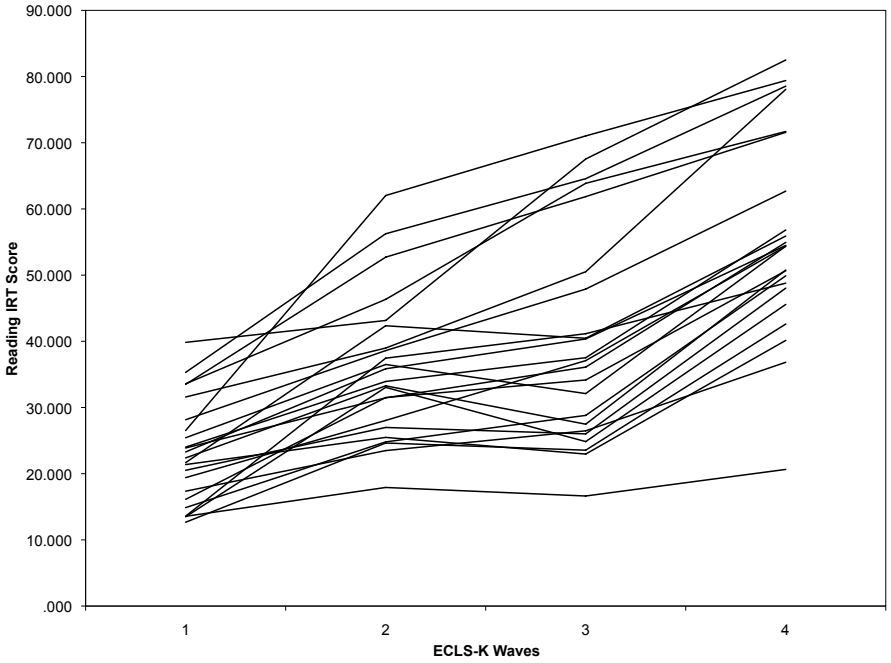
Categorical Latent Variables

Latent class analysis

LCA model estimation

LCA model fit

- Let's consider the growth curves for the ECLS-K one more time.



- Notice that there are clusters of different curve shapes

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- Our concern with conventional growth curve modeling is that is presumes a “one-size-fits-all” model for the growth trajectories.
- In other words, we are assuming that the growth trajectories can be described by a single average curve.
- What if this is not true? What if there are distinct curve shapes?
- The implications for interventions are important. Knowing if there are sub-populations with distinct curve shapes permits tailored interventions.
- We need a method to discern if there are distinct growth trajectories, how many of them exist in the data, and how many individuals can be classified into having these shapes.
- This is the goal of growth mixture modeling.

Growth Mixture Modeling

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- Growth mixture modeling is a type of general mixture model.
- The idea is that in any given population, there might be mixtures of sub-populations (latent classes) of varying number and size.
- For growth mixture modeling, the sub-populations are characterized by qualitatively (or substantively important) growth curve shapes.
- The model parameters here represent the shapes of these trajectories within sub-populations.

From Kaplan, D. (2002) *Peabody Journal of Education*

Introduction

Path Analysis

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

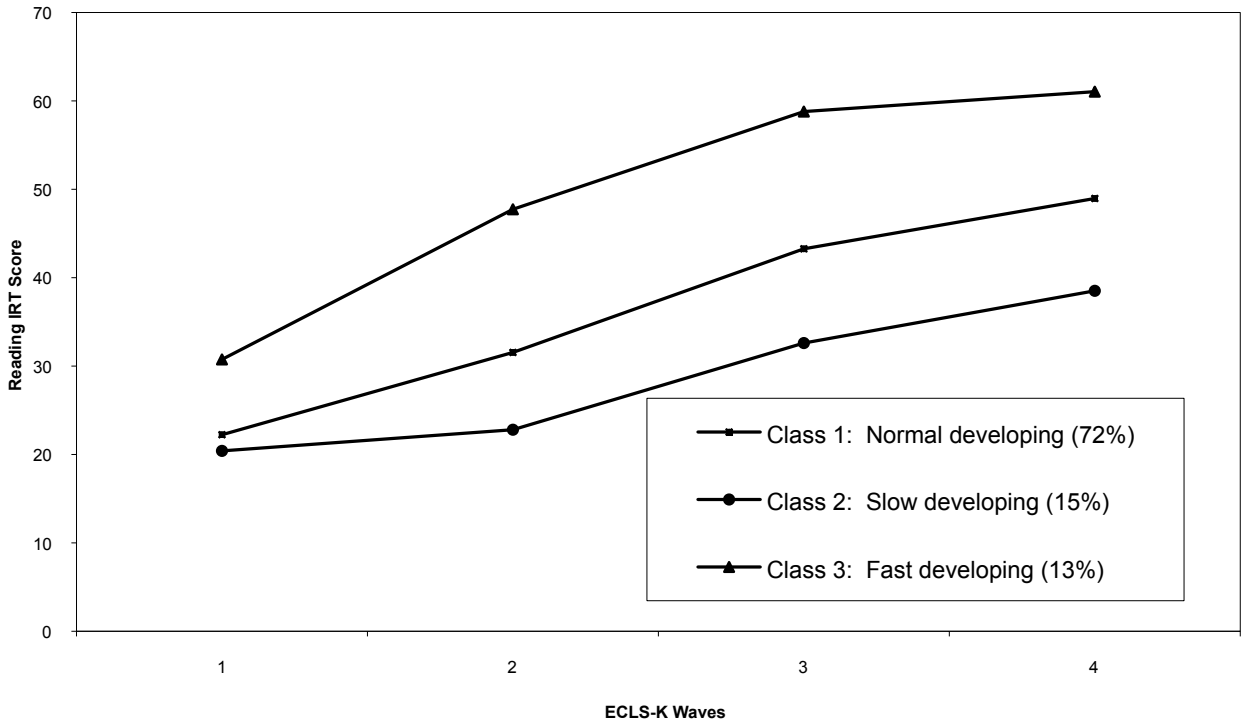
Growth Curve Modeling

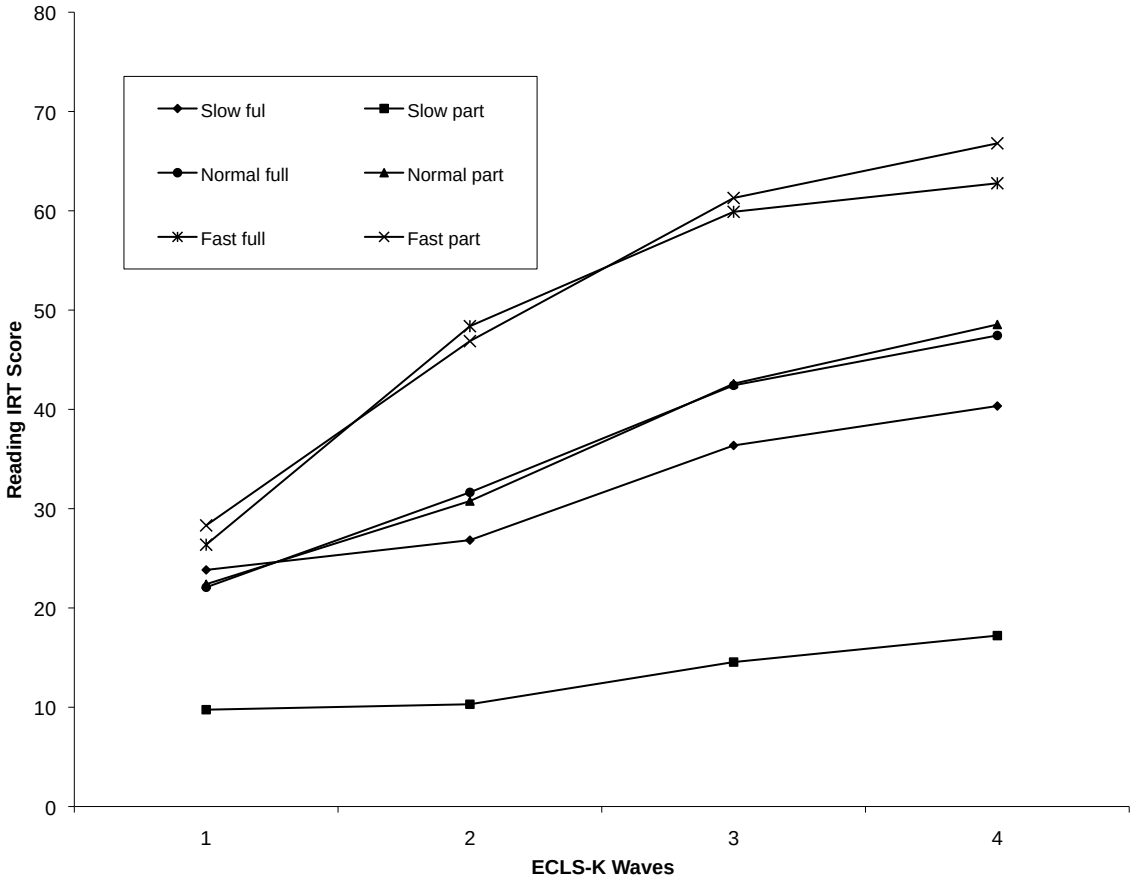
Categorical Latent Variables

Latent class analysis

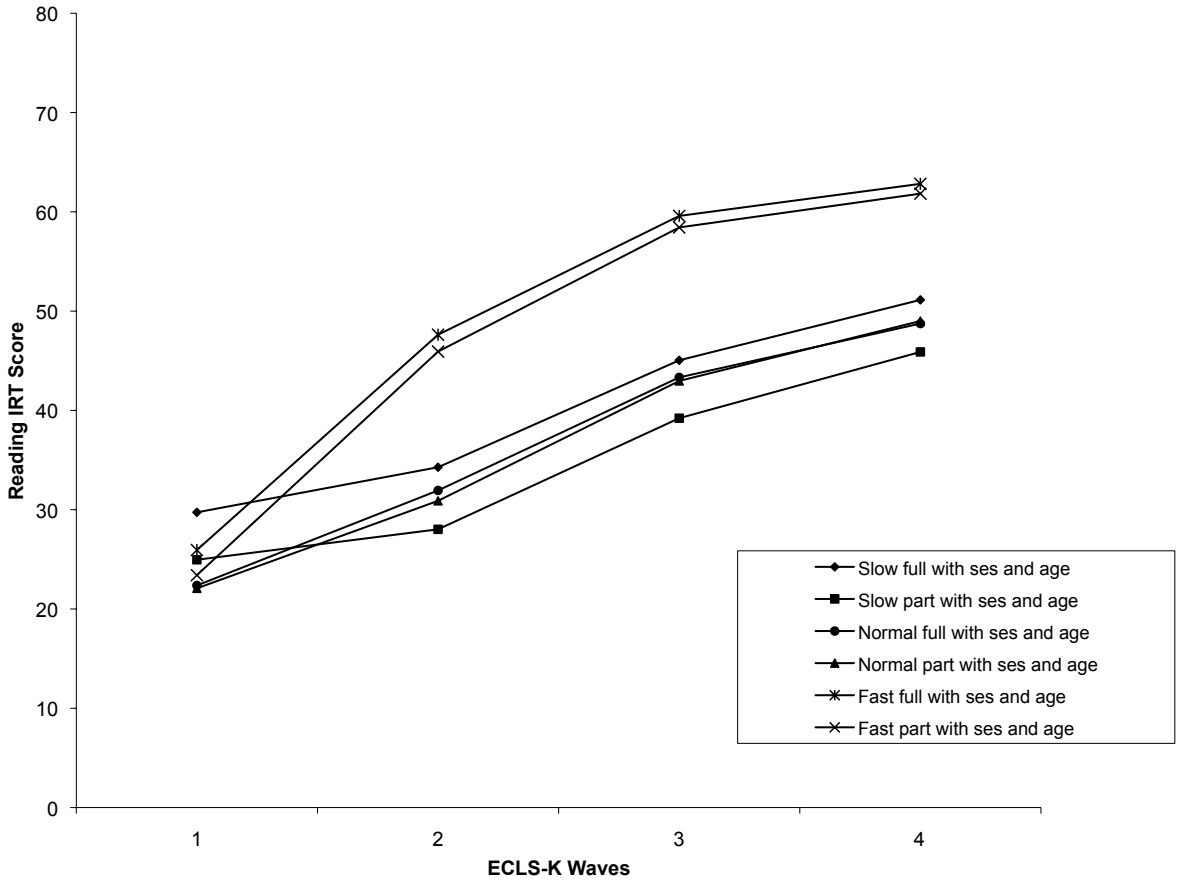
LCA model estimation

LCA model fit





Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit



Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

Causal inference

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Latent class analysis
LCA model estimation
LCA model fit

- Mixture models in experimental designs
  - Complier average causal effects (CACE)
  - Handles non-compliance through the specification of a unobserved (latent) complier class.
  - Comparison of treatment compliers with those in control group who would have complied if they had the treatment.
- Growth mixture models and mixture models for experimental designs can be estimated in Mplus.

Bayesian SEM

- Bayesian statistics has long been overlooked in the quantitative methods training of social scientists.
- Typically, the only introduction that a student might have to Bayesian ideas is a brief overview of Bayes' theorem while studying probability in an introductory statistics class.
  1. Until recently, it was not feasible to conduct statistical modeling from a Bayesian perspective owing to its complexity and lack of available software.
  2. Bayesian statistics represents a powerful alternative to frequentist (classical) statistics, and is therefore, controversial.
- Bayesian statistical methods are now more popular owing to the development of powerful statistical software tools that make the estimation of complex models feasible from a Bayesian perspective.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Paradigm Differences

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- For frequentists, the basic idea is that probability is represented by the model of **long run frequency**.
- The frequentist formulation rests on the idea of equally probable and stochastically independent events
- The physical representation is the coin toss, which relates to the idea of a very large (actually infinite) number of repeated experiments.

Paradigm Differences

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

- The entire structure of Neyman - Pearson hypothesis testing and Fisherian statistics (together referred to as the **frequentist school**) is based on frequentist probability.
- Our conclusions regarding null and alternative hypotheses presuppose the idea that we could conduct the same experiment an infinite number of times.
- Our interpretation of confidence intervals also assumes a fixed parameter and CIs that vary over an infinitely large number of identical experiments.

Paradigm Differences

- But there is another view of probability as **subjective belief**.
- We refer to this as *epistemic probability*.
- From Definetti

  > *"... only subjective probabilities exist – i.e. the degree of belief in the occurrence of an event attributed by a given person at a given instant with a given set of information." (pp 4–5, italics de finetti's)*

- This is the view of a radical subjectivist. More modern views see the problem as one of quantifying prior empirical evidence/assumptions about effects of interest.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

Paradigm Differences

- The physical model for epistemic probability is that of the “bet”.
- Here, probability is not based on an infinite number of repeatable and stochastically independent events, but rather on how much knowledge/information you have and how much you are willing to bet.
- Epistemic probability allows one to address questions such as “what is the probability that my team will win the World Series?”
- Relative frequency supplies information, but it is not the same as probability and can be quite different.
- This notion of epistemic probability underlies Bayesian statistics.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM

Probability Concepts

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Most students in the social sciences were introduced to the axioms of probability by studying the properties of the coin toss or the dice roll. These studies address questions such as
  1. What is the probability that the flip of a fair coin will return heads?;
  2. What is the probability that the roll of two fair die will return a value of seven?
- To answer these questions requires enumerating the possible outcomes and then counting the number of times the event could occur.
- The probabilities of interest are obtained by dividing the number of times the event occurred by the number of possible outcomes.

Bayes' Theorem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Consider the joint probability of two events, Y and X, for example observing lung cancer and smoking jointly.
- The joint probability can be written as

$$p(cancer, smoking) = p(cancer|smoking)p(smoking) \quad (116)$$

- Similarly

$$p(smoking, cancer) = p(smoking|cancer)p(cancer) \quad (117)$$

Bayes' Theorem

- Because these are symmetric, we can set them equal to each other to obtain the following

$$p(cancer|smoking)p(smoking) = p(smoking|cancer)p(cancer) \tag{118}$$

$$p(cancer|smoking) = \frac{p(smoking|cancer)p(cancer)}{p(smoking)} \tag{119}$$

- The inverse probability theorem (Bayes' theorem) states

$$p(smoking|cancer) = \frac{p(cancer|smoking)p(smoking)}{p(cancer)} \tag{120}$$

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Bayes' Theorem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Why do we care?
- Because this is how we go from the probability of a patient having cancer given that the patient smokes, to the probability that the patient smokes given that he/she has cancer.
- We simply need the marginal probability of smoking and the marginal probability of cancer ("base rates" or what we will call **prior probabilities**).

Statistical Elements of Bayes' Theorem

- What is the role of Bayes' theorem for statistical inference?
- Denote by Y a random variable that takes on a realized value y.
- For example, a person's socio-economic status could be considered a random variable taking on a very large set of possible values.
- Once the person identifies his/her socioeconomic status, the random variable Y is now realized as y.
- Because Y is unobserved and random, we need to specify a probability model to explain how we obtained the actual data values y.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Statistical Elements of Bayes' Theorem

- Next, let θ be a parameter that we believe characterizes the probability model of interest.
- The parameter θ can be a scalar, such as the mean or the variance of a distribution, or it can be vector-valued, such as a set of regression coefficients in regression analysis or factor loadings in factor analysis.
- We are concerned with determining the probability of observing y given the unknown parameters θ, which we write as $p(y|\theta)$.
- In statistical inference, the goal is to obtain estimates of the unknown parameters given the data.
- This is expressed as the likelihood of the parameters given the data, often denoted as $L(\theta|y)$.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Statistical Elements of Bayes' Theorem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- A key difference between Bayesian statistical inference and frequentist statistical inference concerns the nature of the unknown parameters θ.
- In the frequentist tradition, the assumption is that θ is unknown and has a fixed value that we wish to estimate. Our uncertainty about θ is not taken into account in the frequentist tradition – only sampling variability. This is what the standard error tells you.
- In Bayesian statistical inference, θ is also considered unknown requiring a probability distribution that reflects our uncertainty about the true value of θ.
- Because both the observed data y and the parameters θ are assumed random, we can model the joint probability of the parameters and the data.

Statistical Elements of Bayes' Theorem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- More formally,

$$p(\theta, y) = p(y|\theta)p(\theta). \tag{121}$$

where $p(\theta, y)$ is the joint distribution of the parameters and the data. Following Bayes' theorem described earlier, we obtain

Bayes' Theorem

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{122}$$

where $p(\theta|y)$ is referred to as the *posterior distribution* of the parameters θ given the observed data y.

Statistical Elements of Bayes' Theorem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Notice that $p(y)$ does not involve model parameters, so we can omit the term and obtain the *unnormalized posterior distribution*

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{123}$$

- This can also be expressed in terms of the likelihood

$$p(\theta|y) \propto L(\theta|y)p(\theta). \tag{124}$$

Statistical Elements of Bayes' Theorem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Equations (123) and (124) represents the core of Bayesian statistical inference and is what separates Bayesian statistics from frequentist statistics.
- Equation (124) states that our uncertainty regarding the parameters of our model, as expressed by the prior distribution $p(\theta)$, is *weighted* by the actual data $p(y|\theta)$ (or equivalently, $L(\theta|y)$), yielding an updated estimate of our uncertainty, as expressed in the posterior density $p(\theta|y)$.
- We learn by updating our prior beliefs/information with data.

Prior Distributions

- Why do we specify a prior distribution on the parameters?
  - The ”orthodox” view is one of ”Bayesian updating”, that we learn by updating our prior beliefs with data.
  - Information from previous research is almost always incorporated into our choice of designs, variables to be measured, or conceptual diagrams to be drawn.
  - Bayesian statistical inference requires that our prior beliefs be made explicit, but then moderates our prior beliefs by the actual data in hand.
  - Moderation of our prior beliefs by the data in hand is the key meaning behind equation (123).

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Prior Distributions

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- This ”orthodox” view is inductive in nature - that Bayesian updating gets us closer to the truth.
- A deductive ”neo-Popperian” view considers prior distributions as assumptions about parameters coming from empirical evidence that help stabilize and regularize estimation (Gelman & Shalizi, 2012).
- These assumptions may be based on prior knowledge, but must be made transparent.
- The probability calculus that leads from the prior to the posterior is a deductive theorem, and the posterior distribution can be subjected to probes in an attempt to locate errors in the model (e.g. Mayo & Spanos, 2004).

Cromwell's Rule

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Named by Lindley, the reference is to Oliver Cromwell, who wrote to the General Assembly of the Church of Scotland on 3 August 1650, including a phrase that has become well known and frequently quoted

  > *I beseech you, in the bowels of Christ, think it possible that you may be mistaken.*

- Prior probabilities should be greater than zero and less than one. Why?

$$p(\theta|y) = \frac{p(y|\theta) * 0}{p(y)} = 0 \tag{125}$$

- No amount of evidence to the contrary would change your mind.

Cromwell's Rule

- What about $p(\theta) = 1$?
- Consider that the denominator $p(y)$ is a marginal distribution across ALL possible values of θ.
- Then, if $p(\theta) = 1$, the denominator $p(y)$ collapses to only your hypothesis $p(y|\theta)$ and .

$$p(\theta|y) = \frac{p(y|\theta)}{p(y|\theta)} = 1 \tag{126}$$

- Again, in other words, no amount of evidence to the contrary would change your mind. The probability of your hypothesis is 1.0. You're sticking to your hypothesis, no matter what.
- Cromwell's rule states that unless the statements are deductions of logic, then one should leave some doubt (however small) in probability assessments.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Non-informative Priors

- In some cases we may not be in possession of enough prior information to aid in drawing posterior inferences.
- From a Bayesian perspective, this lack of information is still important to consider and incorporate into our statistical specifications.
- In other words, it is equally important to quantify our ignorance as it is to quantify our cumulative understanding of a problem at hand.
- The standard approach to quantifying our ignorance is to incorporate a non-informative prior into our specification.
- Non-informative priors are also referred to as *vague* or *diffuse* priors.

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Non-informative Priors

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- There are a large class of non-informative priors, but perhaps the most common non-informative prior distribution is the uniform distribution $U(\alpha, \beta)$ over some sensible range of values from α to β.
- The uniform distribution indicates that we believe that the value of our parameter of interest lies in range $\beta - \alpha$ and that all values have equal probability.
- Care must be taken in the choice of the range of values over the uniform distribution. A $U[-\infty, \infty]$ is an *improper* prior distribution because it does not integrate to 1.0 as required of probability distributions.

Other Non-informative Priors

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- The vague normal prior: $\theta \sim N(0, 10^6)$
- Jeffreys' prior
- Maximum entropy prior
- Zellner's g-prior

Weakly Informative Priors

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- We may also simply want to specify reasonable bounds on model parameters.
- For example, certain parameters just simply are never (or can't be) negative.
- *Weakly informative priors*

  "...let the data speak while being strong enough to exclude various "unphysical" possibilities which, if not blocked, can take over a posterior distribution in settings with sparse data..." (Gelman, 2009, pg. 176).

Informative Priors

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- It may be the case that some information can be brought to bear on a problem and be systematically incorporated into the prior distribution.
- Such “subjective” priors are called *informative*.
- One type of informative prior is based on the notion of a *conjugate distribution*.
- A conjugate prior distribution is one that, when combined with the likelihood function yields a posterior that is in the same distributional family as the prior distribution.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

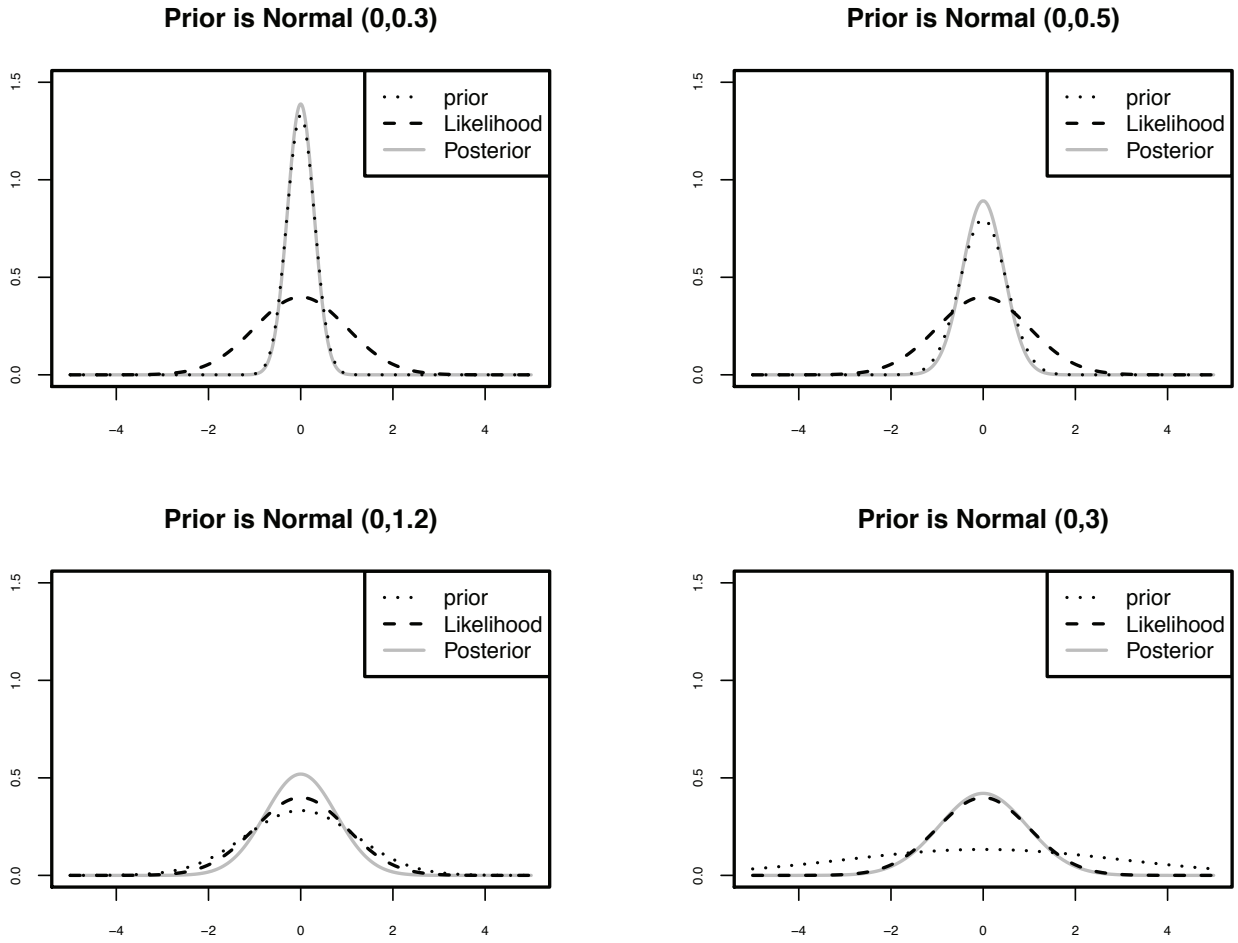


Figure: Normal distribution, mean unknown/variance known with varying conjugate priors

Bayesian Hypothesis Testing

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- A critically important component of applied statistics is hypothesis testing.
- The approach most widely used in the social and behavioral sciences is the Neyman-Pearson approach.
- An interesting aspect of the Neyman-Pearson approach to hypothesis testing is that students (as well as many seasoned researchers) appear to have a very difficult time grasping its principles.
- Gigerenzer (2004) argued that much of the difficulty in grasping frequentist hypothesis testing lies in the conflation of Fisherian hypothesis testing and the Neyman-Pearson approach to hypothesis testing.

Fisher

- Fisher's early approach to hypothesis testing required specifying only the null hypothesis.
- A conventional significance level is chosen (usually the 5% level).
- Once the test is conducted, the result is either significant ($p < .05$) or it is not ($p > .05$).
- If the resulting test is significant, then the null hypothesis is rejected. However, if the resulting test is not significant, then no conclusion can be drawn.
- Fisher's approach was based on looking at how data could inform evidence for a hypothesis.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Neyman and Pearson

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- The Neyman and Pearson approach requires that two hypotheses be specified – the null and alternative hypothesis – and is designed to inform specific sets of actions. It's about making a choice, not about evidence against a hypothesis.
- By specifying two hypotheses, one can compute a desired tradeoff between two types of errors: Type I errors (α) and Type II errors (β)
- The goal is not to assess the evidence against a hypothesis (or model) taken as true. Rather, it is whether the data provide support for taking one of two competing sets of actions.
- In fact, prior belief as well as interest in "the truth" of the hypothesis is irrelevant – only a decision that has a low probability of being the wrong decision is relevant.

Neyman and Pearson

- The conflation of Fisherian and Neyman-Pearson hypothesis testing lies in the use and interpretation of the p-*value*.
- In Fisher's paradigm, the p-value is a matter of convention with the resulting outcome being based on the data.
- In the Neyman-Pearson paradigm, α and β are determined prior to the experiment being conducted and refer to a consideration of the cost of making one or the other error.
- Indeed, in the Neyman-Pearson approach, the problem is one of finding a balance between α, power, and sample size.
- The Neyman-Pearson approach is best suited for experimental planning. Sometimes, it is used this way, followed by the Fisherian approach for judging evidence. But, these two ideas may be incompatible (Royall, 1997).

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- However, this balance is virtually always ignored and α = 0.05 is used.
- The point is that the p-value and α are not the same thing.
- This is made worse by the fact that statistical software packages often report a number of p-values that a researcher can choose from after having conducted the analysis (e.g., .001, .01, .05).
- This can lead a researcher to set α ahead of time (perhaps according to an experimental design), but then communicate a different level of "significance" after running the test.
- The conventional practice is even worse than described, as evidenced by nonsensical phrases such as results ...

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- “trending toward significance”,

or

- “approaching significance”,

or

- “nearly significant”

See also Still Not Significant

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- “trending toward significance”,

or

- “approaching significance”,

or

- “nearly significant”

See also Still Not Significant

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- “trending toward significance”,

or

- “approaching significance”,

or

- “nearly significant”

See also Still Not Significant

NHST - the p-value problem

Introduction

Path Analysis

Factor Analysis

Statistical Assumptions

SEM

Multiple Group Modeling

Multilevel SEM

Growth Curve Modeling

Categorical Latent Variables

Bayesian SEM

Paradigm Differences

- In NHST, one proceeds by specifying a test statistic $t()$ that summarizes the data x in a single number, $t(x)$. For example, a t-statistic.
- The p-value is obtained by considering $Pr[t(x)|H_0]$, as well as that part of the sampling distribution $\boldsymbol{t}(\boldsymbol{x}^{rep}|H_0)$ more extreme than the observed $t(x)$.
- A p value is obtained from the distribution of a test statistic over hypothetical replications (i.e., the sampling distribution). The p value is the sum or integral over values of the test statistic that are at least as extreme as the one that is actually observed.
- In other words, the p-value is the probability of observing data **at least as extreme** as the data that was actually observed, computed under the assumption that the null hypothesis is true.

NHST - the p-value problem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- NHST violates the "Likelihood Principle" – a foundational principle in statistics (e.g. Birnbaum, 1962).

The Likelihood Principle. *In making inference or decisions about a parameter after data are observed, all relevant observational information is contained in the likelihood function for the observed data. Furthermore, two likelihood functions contain the same information about the parameter if they are proportional to each other (as functions of the parameter).*

NHST - the p-value problem

- The calculation of the p-value is based on data that were never observed.
- If we observe an effect, say $x = 5$, then the significance calculations involve not just $x = 5$ but also more extreme values, $x > 5$.
- But $x > 5$ was not observed and it might not be possible to observe it in reality!
- To quote Jeffreys (1961, pg. 385)

> "I have always considered the arguments for the use of P absurd. They amount to saying that a hypothesis that may or may not be true is rejected because a greater departure from the trial value was improbable; that is, that it has not predicted something that has not happened...This seems a remarkable procedure."

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

NHST - the p-value problem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- In addition, the null hypothesis is never true in reality, and moreover, it is typically not even the research question of interest.
- This leads to the sample size sensitivity of all test statistics.
- All test statistics are of the form $\hat{\theta}/[SE(\hat{\theta})] = \hat{\theta}/[n^{-1/2}(SD)]$
- Holding everything else constant, the test statistic can be made arbitrarily large by increasing the sample size - leading to rejection of H_0.
- Effect sizes are equally arbitrary.

NHST - the p-value problem

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Misunderstanding the Fisher or Neyman-Pearson framework to hypothesis testing and/or poor methodological practice is not a criticism of the approach per se.
- It is more than a misunderstanding. It seems to be fundamentally flawed.
- What can a Bayesian alternative provide?

Point Summaries of the Posterior Distribution

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Bayesian hypothesis testing begins first by obtaining summaries of relevant distributions.
- The difference between Bayesian and frequentist statistics is that with Bayesian statistics we wish to obtain summaries of the posterior distribution.
- The expressions for the mean (EAP), variance, and mode (MAP) of the posterior distribution come from expressions for the mean and variance of conditional distributions generally.

Interval Summaries

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Posterior Probability Intervals
  - In addition to point summary measures, it may also be desirable to provide interval summaries of the posterior distribution.
  - Recall that the frequentist confidence interval requires that we imagine an infinite number of repeated samples from the population characterized by μ.
  - For any given sample, we can obtain the sample mean $\bar{x}$ and then form a $100(1-\alpha)$% confidence interval.
  - The correct frequentist interpretation is that $100(1-\alpha)$% of the confidence intervals formed this way capture the true parameter μ under the null hypothesis. Notice that the probability that the parameter is in the interval is either zero or one.

Interval Summaries

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Posterior Probability Intervals (cont'd)
  - In contrast, the Bayesian framework assumes that a parameter has a probability distribution.
  - Sampling from the posterior distribution of the model parameters, we can obtain its quantiles. From the quantiles, we can directly obtain the probability that a parameter lies within a particular interval.
  - So, a 95% posterior probability interval would mean that the probability that the parameter lies in the interval is 0.95.
  - Notice that this is entirely different from the frequentist interpretation, and arguably aligns with common sense.

Interval Summaries

- It is very important to realize that any interval of interest can be assessed.
- For example, one may be interested in whether the parameter of interest lies between two substantively important values.
- With the mean and standard deviation of the posterior distribution in hand, it is extremely easy to calculate the probability that the estimated value lies in the interval, using just the z-score formula.
- So, even if zero is in the interval, and thus a plausible value for the parameter, the obtained value might have a very large probability of being different from zero.
- We no longer have to dichotomize the world into significant/non-significant.

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Bayesian Computation

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- The key reason for the increased popularity of Bayesian methods in the social and behavioral sciences has been the (re)-discovery of numerical algorithms for estimating the posterior distribution of the model parameters given the data.
- Prior to these developments, it was virtually impossible to analytically derive summary measures of the posterior distribution, particularly for complex models with many parameters.
- Rather than attempting the impossible task of analytically solving for estimates of a complex posterior distribution, we can instead draw samples from $p(\theta|y)$ and summarize the distribution formed by those samples. This is referred to as *Monte Carlo integration*.
- The two most popular methods of MCMC are the Gibbs sampler and the Metropolis-Hastings algorithm.

MCMC

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Bayesian inference focuses on estimating features of the posterior distribution, such as point estimates (e.g. the EAP or MAP) and posterior probability intervals.
- For complex, high-dimensional problems involving multiple integrals, solving expectations can be virtually impossible.
- Rather than attempting the impossible task of analytically solving high dimensional problems, we can instead draw samples from the posterior distribution and summarize the distribution formed by those samples.
- This is referred to as *Monte Carlo integration*.

MCMC

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- MCMC is based on first drawing T samples of the parameters of interest $\{\theta_t, t = 1, \ldots, T\}$ from the posterior distribution $p(\theta|y)$ and approximating the expectation by

$$E[p(\theta|y)] \approx \frac{1}{T} \sum_{t=1}^{T} p(\theta_t|y). \tag{127}$$

- Assuming the samples are independent of one another, the law of large numbers ensures that the approximation in equation (127) will be increasingly accurate as T increases.

- Samples do not have to be drawn independently. All that is required, is that the sequence $\{\theta_t, t = 1, \ldots, T\}$ results in samples throughout the support of the distribution

- The “Markov Chain” allows sampling throughout the support of a distribution while also relaxing the assumption of independent sampling.

MCMC

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- The Markov chain has a number of very important properties.
- Over a long sequence, the chain will "forget" its initial state θ^0 and converge to its stationary distribution $p(\theta|y)$, which does not depend either on the number of samples T or on the initial state θ^0.
- The number of iterations prior to the stability of the distribution is referred to as the *burn-in* samples.
- Letting m represent the initial number of burn-in samples, we can obtain an *ergodic average* of the posterior distribution $p(\theta|y)$ as

$$\bar{p}(\theta|y) = \frac{1}{T-m} \sum_{t=m+1}^{T} p(\theta_t|y). \tag{128}$$

MCMC algorithms

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- At present, there are three dominant MCMC algorithms
  1. Metropolis-Hastings
  2. Gibbs Sampling
  3. Hamiltonian Monte Carlo
- Hamiltonian Monte Carlo is considered the most stable, requiring the fewest iterations, and thus is more efficient than MH or Gibbs.
- HMC is available in the program Stan and its R interface rstan.

More on Chains

- A decision must be made regarding the number of Markov chains to be generated, as well as the number of iterations of the sampler.
- Each chain samples from another location of the posterior distribution based on starting values.
- With multiple chains it may be the case that fewer iterations are required.
- In some cases, the same result can be obtained from one chain, although often requiring a considerably larger number of iterations.
- Once the chain has stabilized, the burn-in samples are discarded. Summary statistics, including the posterior mean, mode, standard deviation, and posterior probability intervals, are calculated on the post-burn-in iterations.

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Convergence Diagnostics

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- It is crucial to assess the convergence of MCMC algorithms.
- Results **cannot** be interpreted unless the chains associated with **all** parameters have demonstrated convergence.
- The difficulty of assessing convergence stems from the very nature of MCMC in that the MCMC algorithm is designed to converge in distribution rather than to a point estimate.
- This means that convergence diagnostics are used most often to ensure that the distribution is properly “explored”.
- It is common to inspect several different diagnostics that examine varying aspects of convergence.

Convergence Diagnostics

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

1. Trace/history plots
   - Assesses the mixing or the “stickiness” of the chain(s).
2. Density plots
   - Assesses any bi-modality in the posterior densities.
3. Autocorrelation function (ACF) plots
   - Assesses the mixing and “stickiness” of the chains.
4. Gelman-Rubin-Brooks diagnostic (also called rhat)
   - Assesses mixing of multiple chains after burn-in.

Bayesian Model Building

- Another difference between the Bayesian and frequentist goals of model building relates to the justification for choosing a particular model among a set of competing models.
- Model building and model choice in the frequentist domain is based primarily on choosing the model that best fits the data.
- This has certainly been the key motivation for model building, respecification, and model choice in the context of structural equation modeling (Kaplan 2009).
- In the Bayesian domain, the choice among a set of competing models is based on which model provides the best posterior predictions.
- That is, the choice among a set of competing models should be based on which model will best predict what actually happened.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Posterior Predictive Checking

- A very natural way of evaluating the quality of a model is to examine how well the model fits the actual data.
- In the context of Bayesian statistics, the approach to examining how well a model predicts the data is based on the notion of *posterior predictive checks*, and the accompanying *posterior predictive p-value*.
- The general idea behind posterior predictive checking is that there should be little, if any, discrepancy between data generated by the model, and the actual data itself.
- Posterior predictive checking is a method for assessing the specification quality of the model. Any deviation between the data generated from the model and the actual data implies model misspecification.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Posterior Predictive Checking

- In the Bayesian context, the approach to examining model fit and specification utilizes the posterior predictive distribution of replicated data.
- Let y^{rep} be data replicated from our current model.

Posterior Predictive Distribution

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)d\theta$$
$$= \int p(y^{rep}|\theta)p(y|\theta)p(\theta)d\theta \quad (129)$$

- Equation (129) states that the distribution of future observations given the present data, $p(y^{rep}|y)$, is equal to the probability distribution of the future observations given the parameters, $p(y^{rep}|\theta)$, weighted by the posterior distribution of the model parameters.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Posterior Predictive Checking

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- To assess model fit, posterior predictive checking implies that the replicated data should match the observed data quite closely if we are to conclude that the model fits the data.
- One approach to model fit in the context of posterior predictive checking is based on Bayesian p-values.
- Denote by $T(y)$ a test statistic based on the data, and let $T(y^{rep})$ be the same test statistics for the replicated data (based on MCMC). Then, the Bayesian p-value is defined to be

$$p\text{-}value = pr[T(y^{rep}, \theta) \geq T(y, \theta)|y]. \quad (130)$$

- The p-value is the proportion of replicated test values that equal or exceed the observed test value. High (or low if signs are reversed) values indicate poor model fit.

Posterior Predictive Checking

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Choose a discrepancy function that measures some part of the distribution for which the model might not fit the data.
- For example, if one is concerned with outliers, choose a discrepancy function that measures whether residuals are more extreme than a certain value.
- If one is concerned with overall lack of fit, then choose a discrepancy function that measures overall fit, such as χ^2.
- Posterior predictive checking is consider the standard approach to evaluating the quality of models and is extremely flexible.

Bayes Factors

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- A very simple and intuitive approach to model comparison and model selection uses so-called *Bayes factors* (Kass & Raftery, 1995)
- In essence, the Bayes factor provides a way to quantify the odds that the data favor one hypothesis over another. A key benefit of Bayes factors is that models do not have to be nested.
- Consider two competing models, denoted as M_1 and M_2, that could be nested within a larger space of alternative models. Let θ_1 and θ_2 be the two parameter vectors associated with these two models.
- These could be two regression models with a different number of variables.

Bayes Factors

- The goal is to develop a quantity that expresses the extent to which the data support M_2 over M_1. One quantity could be the posterior odds of M_2 over M_1, expressed as

Bayes Factors

$$\frac{p(M_2|y)}{p(M_1|y)} = \frac{p(y|M_2)}{p(y|M_1)} \times \left[\frac{p(M_2)}{p(M_1)}\right]. \tag{131}$$

- Notice that the first term on the right hand side of equation (131) is the ratio of two integrated likelihoods.
- This ratio is referred to as the *Bayes factor* for M_2 over M_1, denoted here as B_{21}.
- Our prior belief regarding the odds of M_2 over M_1, given by $p(M_2)/p(M_1)$ is weighted by our consideration of the data, given by $p(y|M_2)/p(y|M_1)$.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Bayes Factors

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- This weighting gives rise to our updated view of evidence provided by the data for either hypothesis, denoted as $p(M_2|y)/p(M_1|y)$.
- An inspection of equation (131) also suggests that the Bayes factor is the ratio of the posterior odds to the prior odds.
- In practice, there may be no prior preference for one model over the other. In this case, the prior odds are neutral and $p(M_1) = p(M_2) = 1/2$.
- When the prior odds ratio equals 1, then the posterior odds is equal to the Bayes factor.

Criticisms of the Bayes Factor

- Two issues from Gelman & Rubin (1995)

1. It is not possible to select a model for some purpose without stating what that purpose might be.
   - The Bayes factor only provides a selection of a model that has more evidence in its favor (as measured by its likelihood) compared to some other model.
   - The chosen model could not only fit the data rather badly, say in terms of predictive performance, but could easily be supplanted by another model that fits the data quite well from a predictive point of view.
2. Posterior model probabilities are blind regarding the intended purposes of the model.
   - This deprives the analyst from using other Bayesian methods to deeply examine the performance of a model in light of its purpose - advocating instead the use of posterior predictive distributions through posterior predictive checking.

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Bayesian Information Criterion

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- A popular measure for model selection used in both frequentist and Bayesian applications is based on an approximation of the Bayes factor and is referred to as the *Bayesian information criterion* (BIC), also referred to as the Schwarz criterion.

- Consider again two models. Under conditions where there is little prior information, the BIC can be written as

$$BIC = -2\log(\hat{\theta}|y) + p\log(n) \tag{132}$$

where $-2\log\hat{\theta}|y$ describes model fit while $p\log(n)$ is a penalty for model complexity, where *p* represents the number of variables in the model and n is the sample size.

Bayesian Information Criterion

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- The difference between two BIC measures comparing, say M_2 to M_1 can be written as

$$\begin{aligned}\Delta(BIC_{21}) &= BIC_{(M_2)} - BIC_{(M_1)} \\ &= \log(\hat{\theta}_2|y) - \log(\hat{\theta}_1|y) - \frac{1}{2}(p_2 - p_1)\log(n)\end{aligned} \tag{133}$$

Bayesian Information Criterion

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Rules of thumb have been developed to assess the quality of the evidence favoring one hypothesis over another using Bayes factors and the comparison of BIC values from two competing models.

Table: Grades of Evidence Corresponding to Values of the Bayes Factor for M_1 Against M_2, and the BIC Difference (Jeffreys, 1961; modified by Raftery, 1995)

BIC Difference	Bayes Factor	$p(M_1\|y)$	Evidence for M_1
0 to 2	1 to 3	0.50 – 0.75	Weak
2 to 6	3 to 20	0.75 – 0.95	Positive
6 to 10	20 to 150	0.95 – 0.99	Strong
> 10	> 150	$>$ .99	Decisive

Criticisms of the BIC

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- The BIC is, arguably, neither Bayesian nor an information criterion.
  1. The BIC does not involve draws from a posterior distribution, but rather requires the evaluation of a likelihood.
  2. Almost all information criteria are derived from the Kullback-Leibler divergence and have their roots in information theory. The BIC is not derived from the Kullback-Leibler divergence.

Criticisms of the BIC

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Aside from the misnomer, there are more important criticisms of the BIC when used for model selection.
- The Bayes factor requires that the prior odds of, say, model 1 against model 2, that is $p(M_1)/p(M_2)$, be specified, though in practice and by default this ratio is set to 1.0, which itself might not be what a researcher truly believes about the prior odds of the models.
- If, however, we wish to account for differences in our prior evidence about these models, then each model would likely also have different prior distributions on the model parameters.

Criticisms of the BIC

- The BIC implies specific (and identical) distributions for each model's set of parameters and these priors might not be what a research truly believes about the two models taken separately.
- In other words, there is only one Bayes factor implied by the use of the BIC, and this might not accurately describe a researcher's prior belief.
- At the very least, researchers who use of the BIC for model selection should interpret their findings with caution, but the BIC probably should be avoided altogether.

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Leave-One-Out Cross-Validation (LOOCV)

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- A recent and important approach to evaluating a model is based on so-called k-fold cross-validation (k-fold CV)
- In k-fold CV, a sample is split into k groups (folds) and each fold is taken to be the validation set with the remaining $k-1$ folds serving as the training set.
- A special case of k-fold CV is *leave-one-out cross validation* (LOOCV), where $k = n$.
- For LOOCV, each observation serves as the validation set with the remaining $n-1$ observations serving as the training set.

LOOCV

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

- Given $p(\theta)$, we can obtain the posterior distribution, $p(\theta|y)$ and the posterior predictive distribution of predicted values $p(\tilde{y}|y) = \int p(\tilde{y}_i|\theta)p(\theta|y)d\theta$.
- The Bayesian LOOCV rests on the derivation of the expected log point-wise predictive density (elpd), defined as

$$\text{elpd} = \sum_{i=1}^{n} \int p_t(\tilde{y}_i) \log p(\tilde{y}_i|y) d\tilde{y}_i, \tag{134}$$

  where $p_t(\tilde{y}_i)$ represents the distribution of the true data-generating process for the predicted values $\tilde{y}_i$ which are approximated by cross-validation procedures.
- The ELPD provides a measure of predictive accuracy for the n data points taken one at a time

LOOCV

- The Bayesian LOO estimate can be written as

$$\text{elpd}_{loo} = \sum_{i=1}^{n} \log p(y_i|y_{-i}), \tag{135}$$

where

$$p(y_i|y_{-i}) = \int p(y_i|\theta)p(\theta|y_{-i})d\theta, \tag{136}$$

- An information criterion based on LOO (*LOOIC*) can be can be derived as

$$\text{LOOIC} = -2\,\widehat{\text{elpd}}_{loo} \tag{137}$$

- Among a set of competing models, the one with the smallest LOOIC is considered best from an out-of-sample point-wise predictive point of view.

Introduction
Path Analysis
Factor Analysis
Statistical Assump-tions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Introduction
Path Analysis
Factor Analysis
Statistical Assumptions
SEM
Multiple Group Modeling
Multilevel SEM
Growth Curve Modeling
Categorical Latent Variables
Bayesian SEM
Paradigm Differences

Install and load “blavaan”
See files: BayesianPathAnalysis.blavaan.html