Mplus Short Courses
Day 1

An Overview Of Factor Analysis, Structural Equation Modeling, And Growth Modeling

Linda K. Muthén
Bengt Muthén

www.statmodel.com
12/06/2010

Table of Contents

Mplus Background

- Inefficient dissemination of statistical methods:
  - Many good methods contributions from biostatistics, psychometrics, etc are underutilized in practice
- Fragmented presentation of methods:
  - Technical descriptions in many different journals
  - Many different pieces of limited software
- Mplus: Integration of methods in one framework
  - Easy to use: Simple, non-technical language, graphics
  - Powerful: General modeling capabilities

Mplus Background

- Mplus versions
  - V1: November 1998
  - V2: February 2001
  - V3: March 2004
  - V4: February 2006
  - V5: November 2007
  - V5.21: May 2009
  - V6: April, 2010

- Mplus team: Linda & Bengt Muthén, Thuy Nguyen, Tihomir Asparouhov, Michelle Conn, Jean Maninger

Statistical Analysis With Latent Variables
A General Modeling Framework

Statistical Concepts Captured By Latent Variables

Continuous Latent Variables

- Measurement errors
- Factors
- Random effects
- Frailties, liabilities
- Variance components
- Missing data

Categorical Latent Variables

- Latent classes
- Clusters
- Finite mixtures
- Missing data

Statistical Analysis With Latent Variables
A General Modeling Framework (Continued)

Models That Use Latent Variables

Continuous Latent Variables

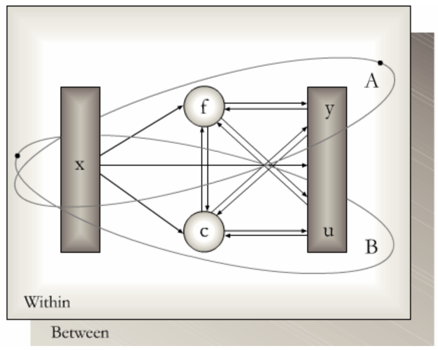
- Factor analysis models
- Structural equation models
- Growth curve models
- Multilevel models

Categorical Latent Variables

- Latent class models
- Mixture models
- Discrete-time survival models
- Missing data models

Mplus integrates the statistical concepts captured by latent variables into a general modeling framework that includes not only all of the models listed above but also combinations and extensions of these models.

General Latent Variable Modeling Framework



- Observed variables
  - x background variables (no model structure)
  - y continuous and censored outcome variables
  - u categorical (dichotomous, ordinal, nominal) and count outcome variables
- Latent variables
  - f continuous variables
    - – interactions among f's
  - c categorical variables
    - – multiple c's

Mplus

Several programs in one

- Exploratory factor analysis
- Structural equation modeling
- Item response theory analysis
- Latent class analysis
- Latent transition analysis
- Survival analysis
- Growth modeling
- Multilevel analysis
- Complex survey data analysis
- Monte Carlo simulation

Fully integrated in the general latent variable framework

Factor Analysis

Factor Analysis

Factor analysis is a statistical method used to study the dimensionality of a set of variables. In factor analysis, latent variables represent unobserved constructs and are referred to as factors or dimensions.

- Exploratory Factor Analysis (EFA)

  Used to explore the dimensionality of a measurement instrument by finding the smallest number of interpretable factors needed to explain the correlations among a set of variables – exploratory in the sense that it places no structure on the linear relationships between the observed variables and on the linear relationships between the observed variables and the factors but only specifies the number of latent variables

Factor Analysis (Continued)

- Confirmatory Factor Analysis (CFA)
  Used to study how well a hypothesized factor model fits a new sample from the same population or a sample from a different population – characterized by allowing restrictions on the parameters of the model

Applications Of Factor Analysis

- Personality and cognition in psychology
  - Child Behavior Checklist (CBCL)
  - MMPI
- Attitudes in sociology, political science, etc.
- Achievement in education
- Diagnostic criteria in mental health

Holzinger-Swineford Data

The data are taken from the classic 1939 study by Karl J. Holzinger and Frances Swineford. Twenty-six tests intended to measure a general factor and five specific factors were administered to seventh and eighth grade students in two schools, the Grant-White School (n = 145) and Pasteur School (n = 156). Students from the Grant-White School came from homes where the parents were American-born. Students from the Pasteur School came from the homes of workers in factories who were foreign-born.

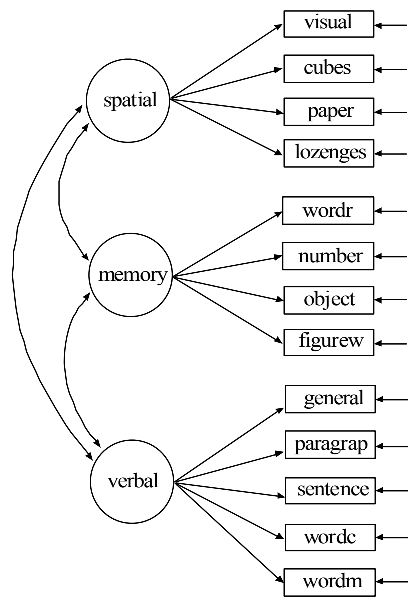
Data for the analysis include nineteen test intended to measure four domains: spatial ability, verbal ability, speed, and memory. Data from the 145 students from the Grant-White School are used.

Holzinger-Swineford Variables

- SPATIAL TESTS
  - Visual perception test
  - Cubes
  - Paper form board
  - Lozenges
- VERBAL TESTS
  - General information
  - Paragraph comprehension
  - Sentence completion
  - Word classification
  - Word meaning

Holzinger-Swineford Variables (Continued)

- SPEED TESTS
  - Add
  - Code
  - Counting groups of dots
  - Straight and curved capitals
- MEMORY
  - Word recognition
  - Number recognition
  - Figure recognition
  - Object-number
  - Number-figure
  - Figure-word



Input Excerpts For Holzinger-Swineford Simple Structure CFA Using 13 Variables

```
MODEL:     spatial BY visual-lozenges;
           memory BY wordr-figurew;
           verbal BY general-wordm;

OUTPUT:    STANDARDIZED MODINDICES(3.84) SAMPSTAT FSDETERMINACY;
```

Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                              56.254
             Degrees of Freedom                     62
             P-Value                            0.6817
CFI/TLI
             CFI                                 1.000
             TLI                                 1.012
RMSEA (Root Mean Square Error Of Approximation)
             Estimate                            0.000
             90 Percent C.I.                     0.000  0.041
             Probability RMSEA <= .05            0.983
SRMR (Standardized Root Mean Square Residual)
             Value                               0.046
```


Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
SPATIAL BY					
VISUAL	1.000	.000	.000	4.539	.659
CUBES	.481	.102	4.691	2.182	.492
PAPER	.329	.066	4.975	1.491	.530
LOZENGES	1.303	.219	5.941	5.915	.714
MEMORY BY					
WORDR	1.000	.000	.000	6.527	.605
NUMBERR	.642	.142	4.534	4.191	.557
OBJECT	.435	.091	4.776	2.840	.624
FIGUREW	.247	.063	3.937	1.613	.450
VERBAL BY					
GENERAL	1.000	.000	.000	9.363	.806
PARAGRAP	.295	.027	11.077	2.766	.822
SENTENCE	.413	.037	11.294	3.866	.834
WORDC	.394	.044	8.857	3.688	.691
WORDM	.716	.062	11.513	6.707	.847

Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables (Continued)

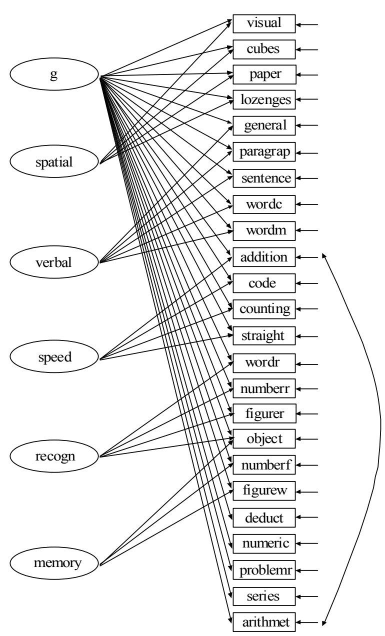
	Estimates	S.E.	Est./S.E.	Std	StdYX
VERBAL WITH					
SPATIAL	25.118	5.700	4.407	.591	.591
MEMORY WITH					
SPATIAL	13.323	4.329	3.077	.450	.450
VERBAL	31.883	8.340	3.823	.522	.522
Variances					
SPATIAL	20.597	5.450	3.779	1.000	1.000
VERBAL	87.646	15.363	5.705	1.000	1.000
MEMORY	42.606	13.205	3.226	1.000	1.000

Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables (Continued)

R-Square

VISUAL	0.434
CUBES	0.243
PAPER	0.281
LOZENGES	0.509
GENERAL	0.650
PARAGRAP	0.676
SENTENCE	0.696
WORDC	0.477
WORDM	0.717
WORDR	0.366
NUMBERR	0.311
OBJECT	0.389
FIGUREW	0.203

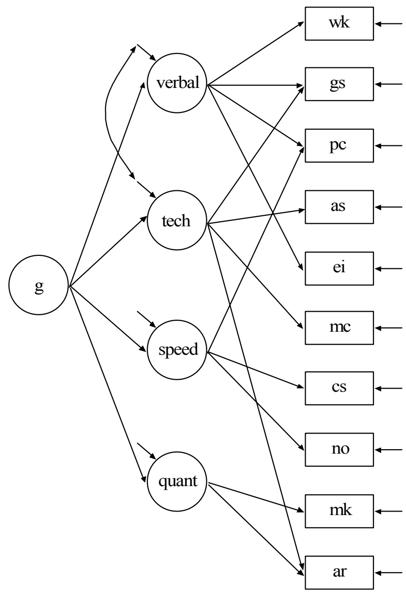
Bi-Factor Model
(Hierarchical Factor Model)



Input Excerpts Holzinger-Swineford General-Specific (Bi-Factor) Factor Model

```
MODEL:     g BY visual-arithmet;
           spatial BY visual-lozenges;
           verbal BY general-wordm;
           speed BY addition-straight;
           recogn BY wordr-object;
           memory BY numberf object figurew;

!          uncorrelated factors because of the general factor:

           g WITH spatial-memory @0;
           spatial WITH verbal-memory @0;
           verbal WITH speed-memory @0;
           speed WITH recogn-memory @0;
           recogn WITH memory @0;

!          correlated residual ("doublet factor"):

           addition WITH arithmet;

OUTPUT:    STANDARDIZED MODINDICES(3.84) SAMPSTAT FSDTERMINACY;
```


Second-Order Factor Model



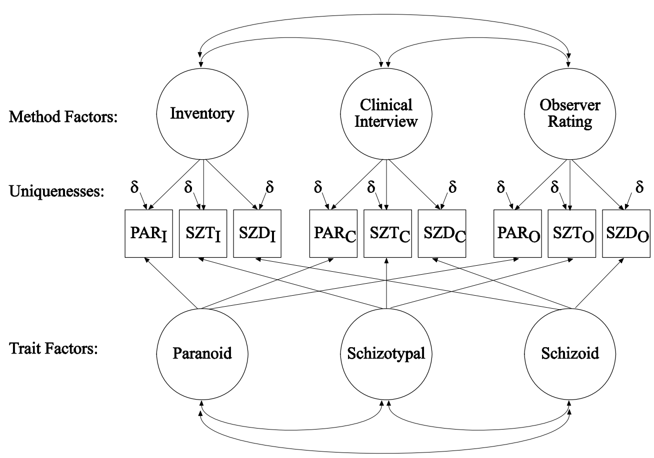
Input For Second-Order Factor Analysis Model

```
TITLE:      Second-order factor analysis model
DATA:       FILE IS asvab.dat;
            !  Armed services vocational aptitude battery
            NOBSERVATIONS = 20422;
            TYPE=COVARIANCE;
VARIABLE:   NAMES ARE ar wk pc mk gs no cs as mc ei;
            USEV = wk gs pc as ei mc cs no mk ar;

            !WK   Word Knowledge
            !GS   General Science
            !PC   Paragraph Comprehension
            !AS   Auto and Shop Information
            !EI   Electronics information
            !MC   Mechanical Comprehension
            !CS   Coding Speed
            !NO   Numerical Operations
            !MK   Mathematical Knowledge
            !AR   Arithmetic Reasoning

ANALYSIS:   ESTIMATOR = ML;
```

Input For Second-Order Factor Analysis Model (Continued)

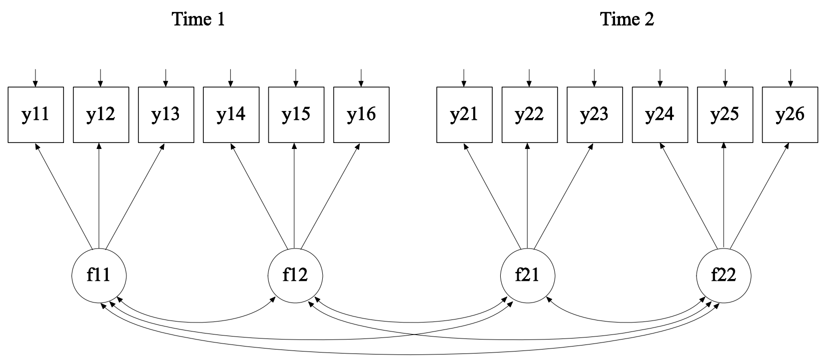
```
MODEL:      verbal BY wk gs pc ei;
            tech BY gs mc ar;
            speed BY pc cs no;
            quant BY mk ar;
            g BY verbal tech speed quant;
            tech WITH verbal;

OUTPUT:     SAMPSTAT MOD(0) STAND TECH1 RESIDUAL;
```


Multi-Trait, Multi-Method (MTMM) Model



Source: Brown (2006)

Longitudinal Factor Analysis Model



ESEM:
Exploratory Structural Equation Modeling

Factor Analysis And Structural Equation Modeling

- Exploratory factor analysis (EFA) is one of the most frequently used multivariate analysis technique in statistics
- 1966 Jennrich solved a significant EFA rotation problem by deriving the direct quartimin rotation
- Jennrich was the first to develop standard errors for rotated solutions although these have still not made their way into most statistical software programs
- 1969 development of confirmatory factor analysis (CFA) by Joreskog
- Joreskog developed CFA further into structural equation modeling (SEM) in LISREL where CFA was used for the measurement part of the model

CFA Simple Structure Λ

$$\Lambda = \begin{pmatrix} X & 0 \\ X & 0 \\ X & 0 \\ 0 & X \\ 0 & X \\ 0 & X \end{pmatrix}$$

where X is a factor loading parameter to be estimated

- CFA simple structure is often too restrictive in practice

Quote From Browne (2001)

"Confirmatory factor analysis procedures are often used for exploratory purposes. Frequently a confirmatory factor analysis, with pre-specified loadings, is rejected and a sequence of modifications of the model is carried out in an attempt to improve fit. The procedure then becomes exploratory rather than confirmatory --- In this situation the use of exploratory factor analysis, with rotation of the factor matrix, appears preferable. --- The discovery of misspecified loadings ... is more direct through rotation of the factor matrix than through the examination of model modification indices."

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research, 36 , 111-150

A New Approach: Exploratory SEM

- Allow EFA measurement model parts (EFA sets)
- Integrated with CFA measurement parts
- Allowing EFA sets access to other SEM parameters, such as
  - Correlated residuals
  - Regressions on covariates
  - Regressions between factors of different EFA sets
  - Regressions between factors of EFA and CFA sets
  - Multiple groups
  - EFA loading matrix equalities across time or group
  - Mean structures
- Available for continuous, categorical, and censored outcomes

Examples

- MIMIC with cross-loadings (see Web Talks)
- Longitudinal EFA (test-retest) (see Web Talks)
- Multiple-group EFA

Example: Aggressive Behavior Male-Female EFA in Baltimore Cohort 3

- 261 males and 248 females in third grade
- Teacher-rated aggressive-disruptive behavior
- Outcomes treated as non-normal continuous variables
- Two types of analyses:
  - EFA in each group separately using Geomin rotation
  - Multiple-group EFA analysis of males and females jointly

Summary Of Separate Male/Female EFAs

Variables	StdYX Loadings for Males			StdYX Loadings for Females		
	Verbal	Person	Property	Verbal	Person	Property
Stubborn	**0.82**	-0.05	0.01	**0.88**	0.03	-0.22
Breaks Rules	0.47	0.34	0.01	**0.76**	0.06	-0.17
Harms Others & Property	-0.01	**0.63**	0.31	0.45	0.03	0.36
Breaks Things	-0.02	0.02	**0.66**	-0.02	0.19	**0.43**
Yells At Others	**0.66**	0.23	-0.03	**0.97**	-0.23	0.05
Takes Others' Property	0.27	0.08	**0.52**	0.02	**0.79**	0.10
Fights	0.22	**0.75**	-0.00	**0.81**	-0.01	0.18
Harms Property	0.03	-0.02	**0.93**	0.27	0.20	**0.57**
Lies	**0.58**	0.01	0.27	0.42	0.50	-0.00
Talks Back to Adults	**0.61**	-0.02	0.30	**0.69**	0.09	-0.02
Teases Classmates	0.46	0.44	-0.04	**0.71**	-0.01	0.10
Fights With Classmates	0.30	**0.64**	0.08	**0.83**	0.03	0.21
Loses Temper	**0.64**	0.16	0.04	**1.05**	-0.29	-0.01

Male And Female Factor Loading Estimates From Multiple-Group ESEM Using Invariant Loadings

Variables	StdYX Loadings for Males			StdYX Loadings for Females		
	Verbal	Person	Property	Verbal	Person	Property
Stubborn	**0.80**	-0.00	-0.02	**0.86**	-0.00	-0.01
Breaks Rules	0.53	0.27	0.01	**0.59**	0.20	0.01
Harms Others & Property	0.00	**0.57**	0.35	0.00	**0.56**	0.24
Breaks Things	-0.01	-0.02	**0.67**	-0.03	-0.03	**0.63**
Yells At Others	**0.66**	0.25	-0.03	**0.69**	0.18	-0.01
Takes Others' Property	0.32	0.04	0.53	0.39	0.03	0.31
Fights	0.28	**0.74**	-0.03	0.35	0.61	-0.02
Harms Property	0.11	0.03	**0.83**	0.19	0.04	**0.68**
Lies	**0.58**	0.01	0.30	**0.67**	0.00	0.16
Talks Back to Adults	**0.64**	-0.03	0.29	**0.71**	-0.02	0.15
Teases Classmates	0.44	0.40	0.02	0.49	0.30	0.01
Fights With Classmates	0.33	**0.65**	0.05	0.41	**0.53**	0.03
Loses Temper	**0.64**	0.19	0.02	**0.74**	0.14	-0.29

Bayesian Analysis

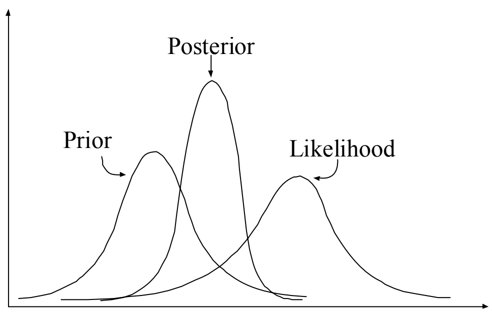
Bayesian Analysis

Why do we have to learn about Bayes?

- More can be learned about parameter estimates and model fit
- Better small-sample performance, large-sample theory not needed
- Analyses can be made less computationally demanding
- New types of models can be analyzed

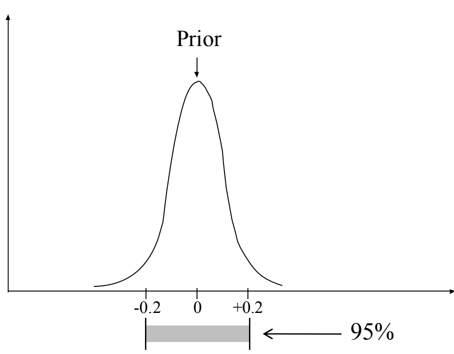
Prior, Likelihood, And Posterior

- Frequentist view: Parameters are fixed. ML estimates have an asymptotically-normal distribution
- Bayesian view: Parameters are variables that have a prior distribution. Estimates have a possibly non-normal posterior distribution. Does not depend on large-sample theory
  – Diffuse (non-informative) priors vs informative priors



Where Do Parameter Priors Come From?

- Previous studies
- Hypotheses based on substantive theory
  – Example: Zero cross-loadings in CFA



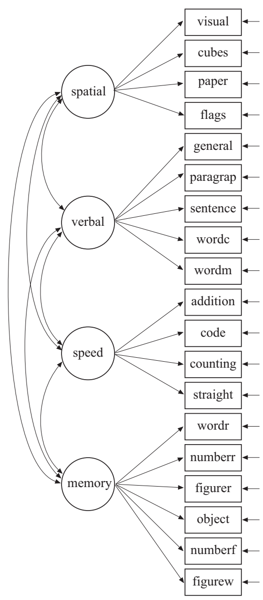
$\lambda \sim N(0,0.01)$

Bayesian Factor Analysis

Holzinger-Swineford Data

- 19 tests hypothesized to measure four mental abilities: Spatial, verbal, speed, and memory
- n=145 7^{th} and 8^{th} grade students from Grant-White elementary school

Source: Muthén, B. & Asparouhov, T. (2010). Bayesian SEM: A more flexible representation of substantive theory. Submitted for publication.



Factor loading pattern

	Spatial	Verbal	Speed	Memory
visual	x	0	0	0
cubes	x	0	0	0
paper	x	0	0	0
flags	x	0	0	0
general	0	x	0	0
paragrap	0	x	0	0
sentence	0	x	0	0
wordc	0	x	0	0
wordm	0	x	0	0
addition	0	0	x	0
code	0	0	x	0
counting	0	0	x	0
straight	0	0	x	0
wordr	0	0	0	x
numberr	0	0	0	x
figurer	0	0	0	x
object	0	0	0	x
numberf	0	0	0	x
figurew	0	0	0	x

ML Tests Of Model Fit For Holzinger-Swineford (N = 145)

- Confirmatory factor analysis (CFA) model rejected:
  - ML likelihood-ratio chi-square $p = 0.0002$
    (CFI = 0.93, RMSEA = 0.057, SRMR = 0.063)
    - Modification indices point to three cross-loadings (Modindices by Sörbom, 1989)
    - Model still rejected when those are freed
- Exploratory factor analysis (EFA) with 4 factors using oblique rotation (Geomin):
  - ML likelihood-ratio chi-square $p = 0.248$
    (CFI = 0.99, RMSEA = 0.025, SRMR = 0.030)
    - The pattern of major loadings is as hypothesized
    - Several significant cross-loadings

Bayesian CFA Using MCMC For Holzinger-Swineford

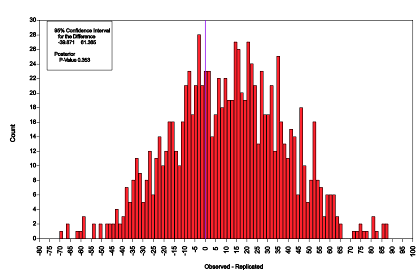
- CFA: Cross-loadings fixed at zero - the model is rejected
- A more realistic hypothesis: Small cross-loadings allowed
- Cross-loadings are not all identified in terms of ML
- Different alternative: Bayesian CFA with informative cross-loading priors: $\lambda \sim N(0, 0.01)$.

  This means that 95% of the prior is in the range -0.2 to 0.2

Bayesian Posterior Predictive Checking For The CFA Model

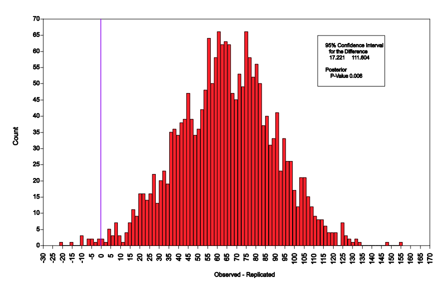
CFA with small cross-loadings not rejected by Bayes PPC:

p = 0.353



Conventional CFA model rejected by Bayes PPC:

p = 0.006:



Summary Of Analyses Of Holzinger-Swineford Data

- Conventional, frequentist, CFA model rejected
- Bayesian CFA not rejected with cross-loadings
- The Bayesian approach uses an intermediate hypothesis:
  - Less strict than conventional CFA
  - Stricter than EFA, where the hypothesis only concerns the number of factors
- Bayes modification indices obtained by estimated cross-loadings

Measurement Invariance And Population Heterogeneity

Models To Study Measurement Invariance And Population Heterogeneity

To further study a set of factors or latent variables established by an EFA/CFA, questions can be asked about the invariance of the measures and the heterogeneity of populations.

Measurement Invariance – Does the factor model hold in other populations or at other time points?

- Same number of factors
- Zero loadings in the same positions
- Equality of factor loadings
- Equality of intercepts
  - Test difficulty

Population Heterogeneity – Are the factor means, variances, and covariances the same for different populations?

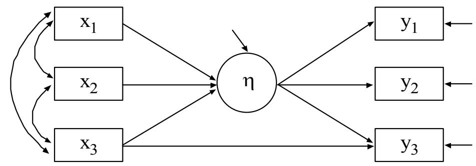
Models To Study Measurement Invariance And Population Heterogeneity (Continued)

Models To Study Measurement Invariance and Population Heterogeneity

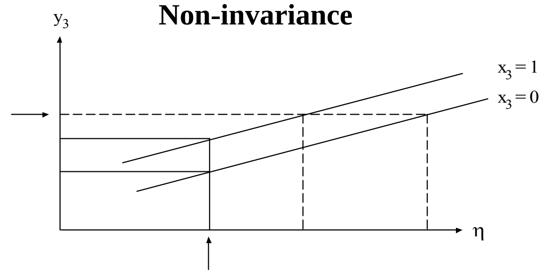
- CFA with covariates
  - Parsimonious
  - Small sample advantage
  - Advantageous with many groups
- Multiple group analysis
  - More parameters to represent non-invariance
    - Factor loadings and observed residual variances/covariances in addition to intercepts
    - Factor variances and covariances in addition to means
  - Interactions

CFA With Covariates

Non-invariance



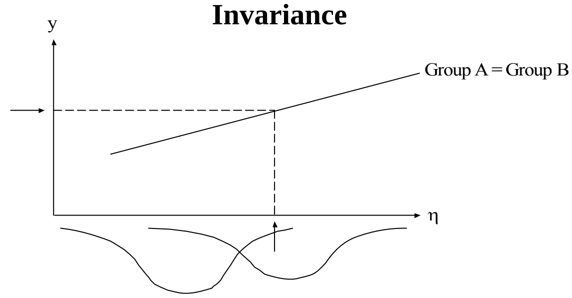
Non-invariance



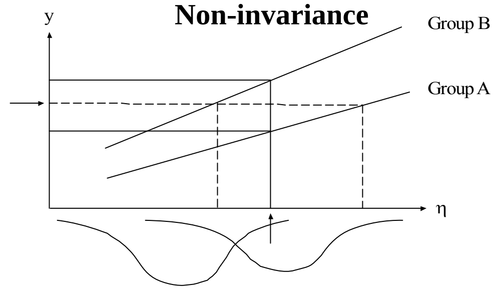
Conditional on η, y is different for the two groups

Multiple Group Analysis

Invariance



Non-invariance



CFA With Covariates (MIMIC)

NELS Data

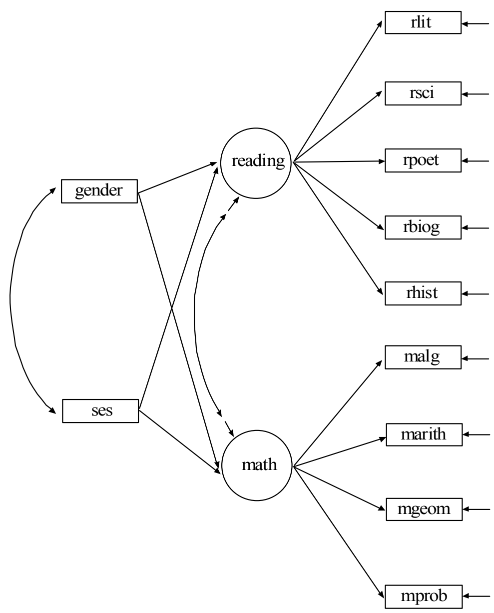
The NELS data consist of 16 testlets developed to measure the achievement areas of reading, math, science, and other school subjects. The sample consists of 4,154 eighth graders from urban, public schools.

Data for the analysis include five reading testlets and four math testlets. The entire sample is used.

Variables

rlit – reading literature
rsci – reading science
rpoet – reading poetry
rbiog – reading biography
rhist – reading history

malg – math algebra
marith – math arithmetic
mgeom – math geometry
mprob – math probability



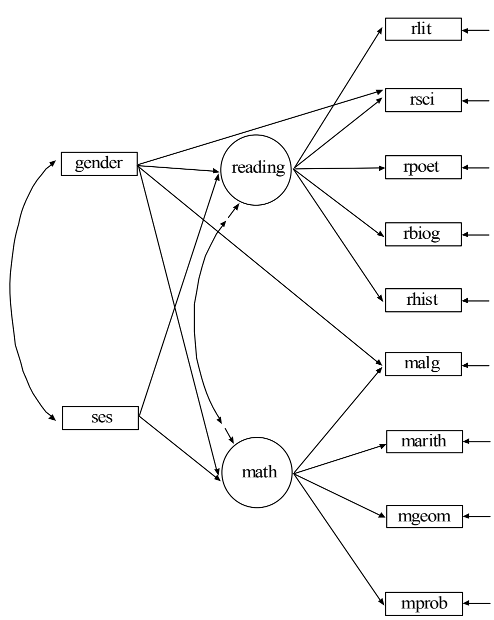
Input For NELS CFA With Covariates

```
TITLE:      CFA with covariates using NELS data

DATA:       FILE IS ft21.dat;

VARIABLE:   NAMES ARE ses rlit rsci rpoet rbiog rhist malg
            marith mgeom mprob searth schem slife smeth hgeog
            hcit hhist gender schoolid minorc;

            USEVARIABLES ARE rlit-mprob ses gender;

MODEL:      reading BY rlit-rhist;
            math BY malg-mprob;

            reading math ON ses gender;   ! female = 0, male = 1

OUTPUT:     STANDARDIZED MODINDICES (ALL 3.84);
```

Output Excerpts Modification Indices For Direct Effects NELS CFA With Covariates

Modification Indices

		M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
RSCI	**ON GENDER**	**31.730**	**0.253**	**0.253**	**0.073**
RPOET	ON GENDER	12.715	-0.124	-0.124	-0.045
RHIST	ON SES	6.579	0.062	0.062	0.038
MALG	**ON GENDER**	**26.616**	**-0.120**	**-0.120**	**-0.051**
MARITH	ON GENDER	10.083	0.075	0.075	0.032
MGEON	ON SES	4.201	0.040	0.040	0.032
MPROB	ON GENDER	7.922	0.143	0.143	0.037



Output Excerpts NELS CFA With Covariates And Two Direct Effects (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
READING ON					
SES	.343	.014	24.854	.406	.437
GENDER	-.222	.028	-7.983	-.262	-.131
MATH ON					
SES	.419	.015	28.807	.411	.444
GENDER	**.092**	**.032**	**2.873**	**.090**	**.045**
RSCI ON					
GENDER	**.254**	**.045**	**5.649**	**.254**	**.073**
MALG ON					
GENDER	**-.121**	**.023**	**-5.171**	**-.121**	**-.051**

Structural Equation Modeling (SEM)

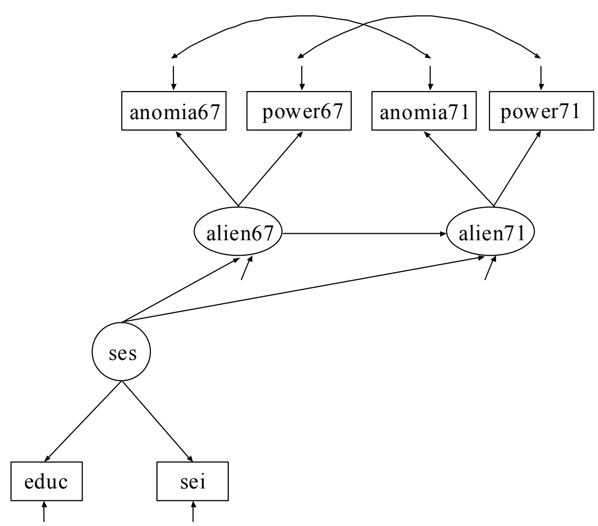
Structural Equation Modeling (SEM)

Used to study relationships among multiple outcomes often involving latent variables

- Estimate and test direct and indirect effects in a system of regression equations for latent variables without the influence of measurement error
- Estimate and test theories about the absence of relationships among latent variables

Model identification, estimation, testing, and modification are the same as for CFA.

Classic Wheaton Et Al. SEM



Input For Classic Wheaton Et Al. SEM

```
TITLE:      Classic structural equation model with multiple
            indicators used in a study of the stability of
            alienation.

DATA:       FILE IS wheacov.dat
            TYPE IS COVARIANCE;
            NOBS ARE 932;

VARIABLE:   NAMES ARE anomia67 power67 anomia71 power71 educ
            sei;

MODEL:      ses       BY educ sei;
            alien67   BY anomia67 power67;
            alien71   BY anomia71 power71;

            alien71   ON alien67 ses;
            alien67   ON ses;

            anomia67  WITH anomia71;
            power67   WITH power71;

OUTPUT:     SAMPSTAT STANDARDIZED MODINDICES (0);
```


Output Excerpts
Classic Wheaton Et Al. SEM

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                              4.771
             Degrees of Freedom                     4
             P-Value                            .3111

RMSEA (Root Mean Square Error Of Approximation)
             Estimate                            .014
             90 Percent C.I.                     .000  .053
             Probability RMSEA <= .05            .928
```

Output Excerpts
Classic Wheaton Et Al. SEM (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
SES BY					
EDUC	1.000	.000	.000	2.607	.841
SEI	5.221	.422	12.367	13.612	.642
ALIEN67 BY					
ANOMIA67	1.000	.000	.000	2.663	.775
POWER67	.979	.062	15.896	2.606	.852
ALIEN71 BY					
ANOMIA71	1.000	.000	.000	2.850	.805
POWER71	.922	.059	15.500	2.627	.832

Output Excerpts
Classic Wheaton Et Al. SEM (Continued)

ALIEN71 ON					
ALIEN67	.607	.051	11.895	.567	.567
SES	-.227	.052	-4.337	-.208	-.208
ALIEN67 ON					
SES	-.575	.056	-10.197	-.563	-.563
ANOMIA67 WITH					
ANOMIA71	1.622	.314	5.173	1.622	.133
POWER67 WITH					
POWER71	.340	.261	1.302	.340	.035

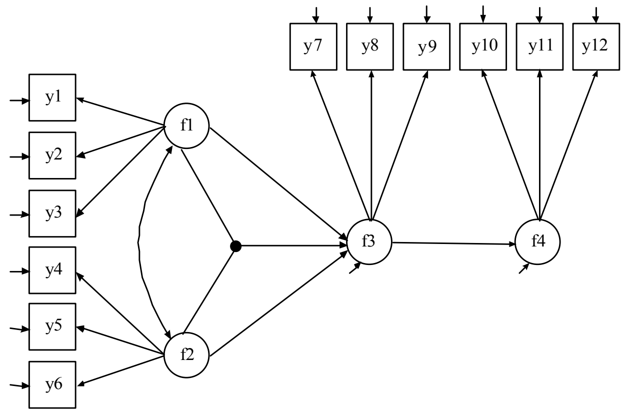
Output Excerpts
Classic Wheaton Et Al. SEM (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
Residual Variances					
ANOMIA67	4.730	.453	10.438	4.730	.400
POWER67	2.564	.403	6.362	2.564	.274
ANOMIA71	4.397	.515	8.537	4.397	.351
POWER71	3.072	.434	7.077	3.072	.308
EDUC	2.804	.507	5.532	2.804	.292
SEI	264.532	18.125	14.595	264.532	.588
ALIEN67	4.842	.467	10.359	.683	.683
ALIEN71	4.084	.404	10.104	.503	.503
Variances					
SES	6.796	.649	10.476	1.000	1.000

Modeling Issues In SEM

- Model building strategies
  - Bottom up
  - Measurement versus structural parts
- Number of indicators
  - Identifiability
  - Robustness to misspecification
- Believability
  - Measures
  - Direction of arrows
  - Other models
- Quality of estimates
  - Parameters, s.e.'s, power
  - Monte Carlo study within the substantive study

Structural Equation Model With Interaction Between Latent Variables



Klein & Moosbrugger (2000)
Marsh et al. (2004)

Growth Modeling

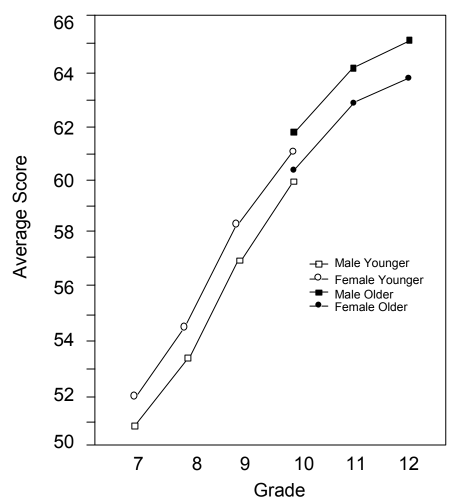
Typical Examples Of Growth Modeling

LSAY Data

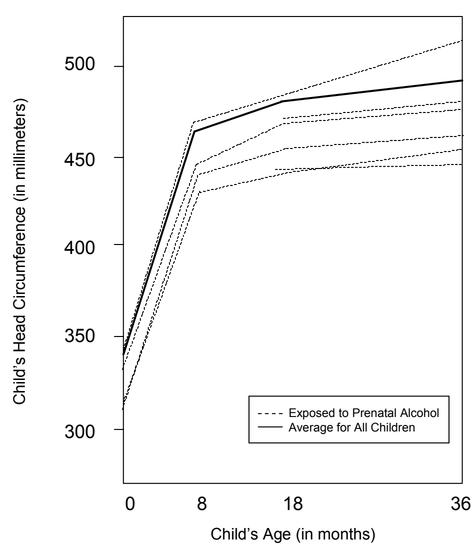
Longitudinal Study of American Youth (LSAY)

- Two cohorts measured each year beginning in 1987
  - Cohort 1 - Grades 10, 11, and 12
  - Cohort 2 - Grades 7, 8, 9, 10, 11, and 12
- Each cohort contains approximately 60 schools with approximately 60 students per school
- Variables - math and science achievement items, math and science attitude measures, and background variables from parents, teachers, and school principals
- Approximately 60 items per test with partial item overlap across grades - adaptive tests

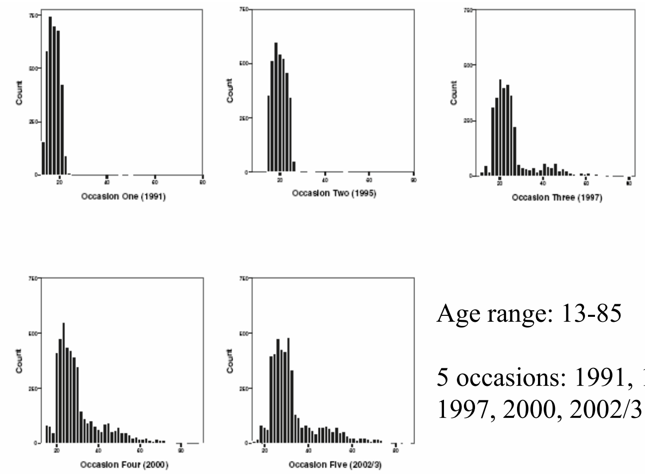
Math Total Score



MHP: Offspring Head Circumference



Loneliness In Twins



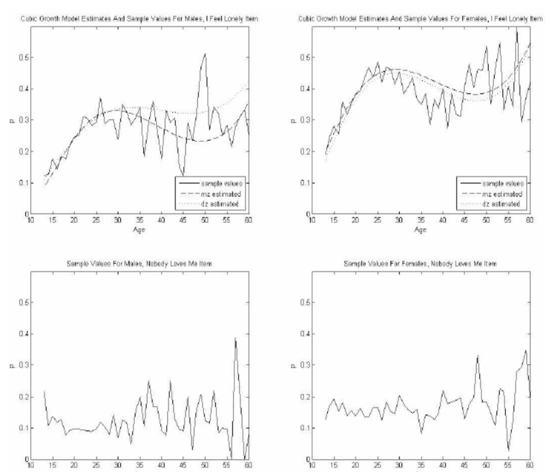
Age range: 13-85

5 occasions: 1991, 1995, 1997, 2000, 2002/3

Boomsma, D.I., Cacioppo, J.T., Muthen, B., Asparouhov, T., & Clark, S. (2007). Longitudinal Genetic Analysis for Loneliness in Dutch Twins. Twin Research and Human Genetics, 10, 267-273.

Loneliness In Twins

Males | Females

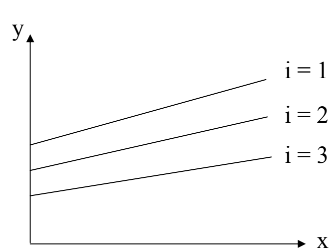


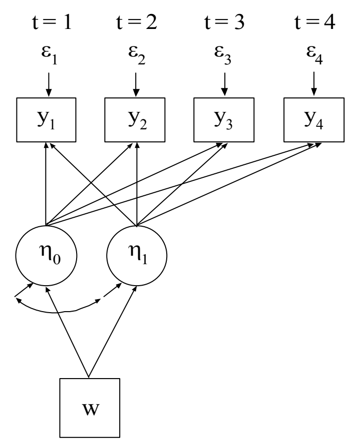
I feel lonely

Nobody loves me

Basic Modeling Ideas

Individual Development Over Time



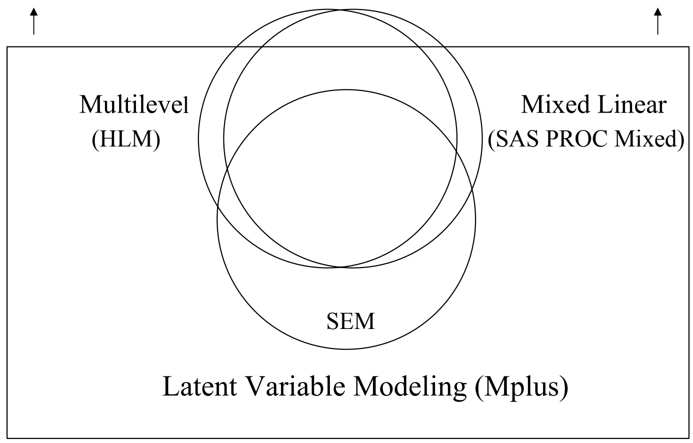


(1) $y_{ti} = \eta_{0i} + \eta_{1i}\, x_t + \varepsilon_{ti}$

(2a) $\eta_{0i} = \alpha_0 + \gamma_0\, w_i + \zeta_{0i}$

(2b) $\eta_{1i} = \alpha_1 + \gamma_1\, w_i + \zeta_{1i}$

Growth Modeling Frameworks/Software



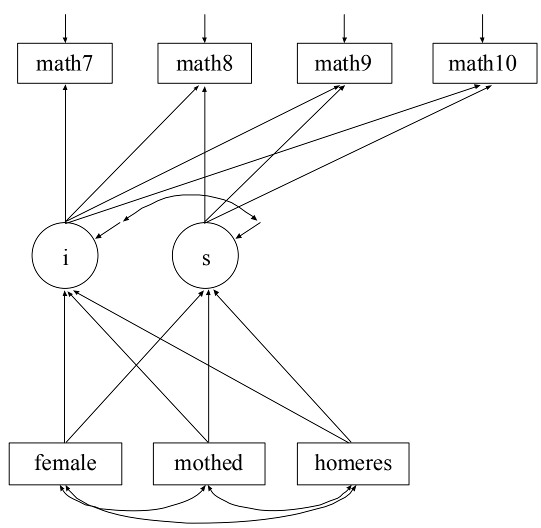
Comparison Summary Of Multilevel, Mixed Linear, And SEM Growth Models

- Multilevel and mixed linear models are the same
- SEM differs from the multilevel and mixed linear models in two ways
  - Treatment of time scores
    - Time scores are data for multilevel and mixed linear models -- individuals can have different times of measurement
    - Time scores are parameters for SEM growth models -- time scores can be estimated
  - Treatment of time-varying covariates
    - Time-varying covariates have random effect coefficients for multilevel and mixed linear models -- coefficients vary over individuals
    - Time-varying covariates have fixed effect coefficients for SEM growth models -- coefficients vary over time

Advantages Of Growth Modeling In A Latent Variable Framework

- Flexible curve shape
- Individually-varying times of observation
- Regressions among random effects
- Multiple processes
- Modeling of zeroes
- Multiple populations
- Multiple indicators
- Embedded growth models
- Categorical latent variables: growth mixtures

LSAY Growth Model With Time-Invariant Covariates



Input Excerpts For LSAY Linear Growth Model With Time-Invariant Covariates

```
TITLE:      Growth 7 – 10, no covariates
DATA:       FILE = lsayfull_dropout.dat;
VARIABLE:   NAMES = lsayid schcode female mothed homeres
            math7 math8 math9 math10 math11 math12
            mthcrs7 mthcrs8 mthcrs9 mthcrs10 mthcrs11 mthcrs12;
            MISSING = ALL (999);
            USEVAR = math7-math10 female mothed homeres;
ANALYSIS:   !ESTIMATOR = MLR;
MODEL:      i s | math7@0 math8@1 math9@2 math10@3;
            i s ON female mothed homeres;
```

Alternative language:

```
MODEL:      i BY math7-math10@1;
            s BY math7@0 math8@1 math9@2 math10@3;
            [math7-math10@0];
            [i s];
            i s ON female mothed homeres;
```


Output Excerpts LSAY Growth Model With Time-Invariant Covariates

n = 3116

Tests Of Model Fit for ML

```
Chi-Square Test of Model Fit
             Value                              33.611
             Degrees of Freedom                      8
             P-Value                             0.000
CFI/TLI
             CFI                                 0.998
             TLI                                 0.994
RMSEA (Root Mean Square Error Of Approximation)
             Estimate                            0.032
             90 Percent C.I.                     0.021  0.044
             Probability RMSEA <= .05            0.996
SRMR (Standardized Root Mean Square Residual)
             Value                               0.010
```

Output Excerpts LSAY Growth Model With Time-Invariant Covariates (Continued)

Selected Estimates For ML

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
I ON				
FEMALE	2.123	0.327	6.499	0.000
MOTHED	2.262	0.164	13.763	0.000
HOMERES	1.751	0.104	16.918	0.000
S ON				
FEMALE	-0.134	0.116	-1.153	0.249
MOTHED	0.223	0.059	3.771	0.000
HOMERES	0.273	0.037	7.308	0.000

Output Excerpts LSAY Growth Model With Time-Invariant Covariates (Continued)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
S WITH				
I	4.131	1.244	3.320	0.001
Residual Variances				
I	71.888	3.630	19.804	0.000
S	3.313	0.724	4.579	0.000
Intercepts				
I	38.434	0.497	77.391	0.000
S	2.636	0.181	14.561	0.000

Non-Linear Growth

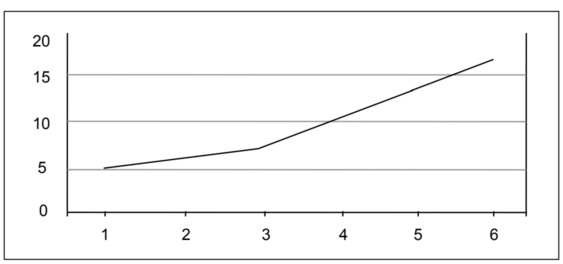
Six Ways To Model Non-Linear Growth

- Estimated time scores
- Quadratic (cubic) growth model
- Fixed non-linear time scores
- Piecewise growth modeling
- Time-varying covariates
- Non-linearity of random effects

Piecewise Growth Modeling

Piecewise Growth Modeling

- Can be used to represent different phases of development
- Can be used to capture non-linear growth
- Each piece has its own growth factor(s)
- Each piece can have its own coefficients for covariates



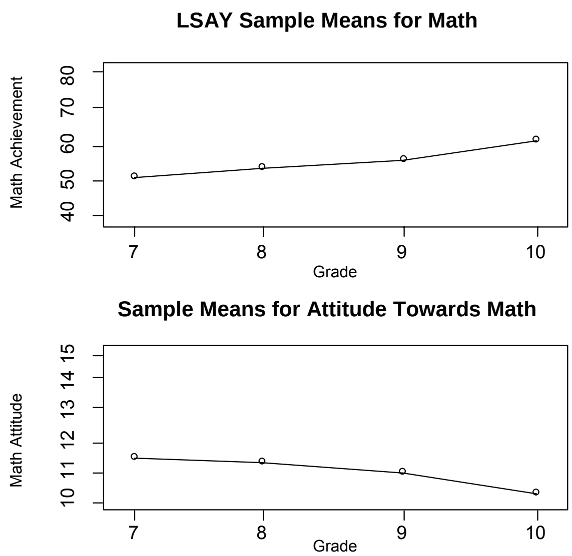
One intercept growth factor, two slope growth factors

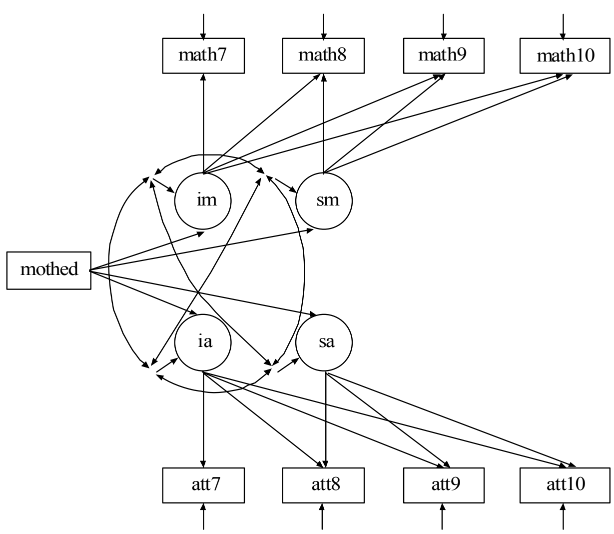
s1:	0	1	2	2	2	2	Time scores piece 1
s2:	0	0	0	1	2	3	Time scores piece 2

Growth Modeling With Parallel Processes

Multiple Processes

- Parallel processes
- Sequential processes


LSAY Sample Means for Math
Math Achievement
40
50
60
70
80
7
8
9
10
Grade
Sample Means for Attitude Towards Math
Math Attitude
10
11
12
13
14
15
7
8
9
10
Grade


math7
math8
math9
math10
im
sm
mothed
ia
sa
att7
att8
att9
att10

Input For LSAY Parallel Process Growth Model (Continued)

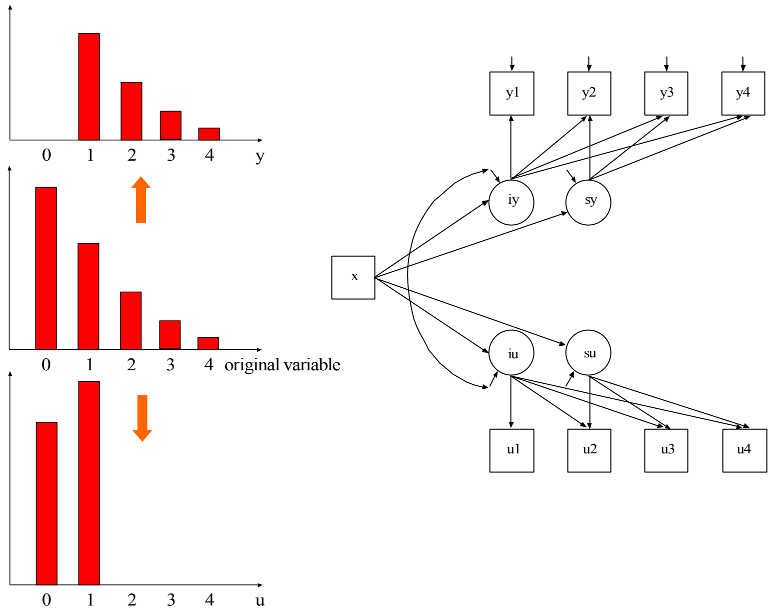
```
MODEL:      im sm | math7@0 math8@1 math9 math10;
            ia sa | att7@0 att8@1 att9@2 att10@3;
            im-sa ON mothed;

OUTPUT:     MODINDICES STANDARDIZED;
```

Alternative language:

```
            im BY math7-math10@1;
            sm BY math7@0 math8@1 math9 math10;

            ia BY att7-att10@1;
            sa BY att7@0 att8@1 att9@2 att10@3;

            [math7-math10@0 att7-att10@0];
            [im sm ia sa];

            im-sa ON mothed;
```


Two-Part Growth Modeling

Two-Part (Semicontinuous) Growth Modeling

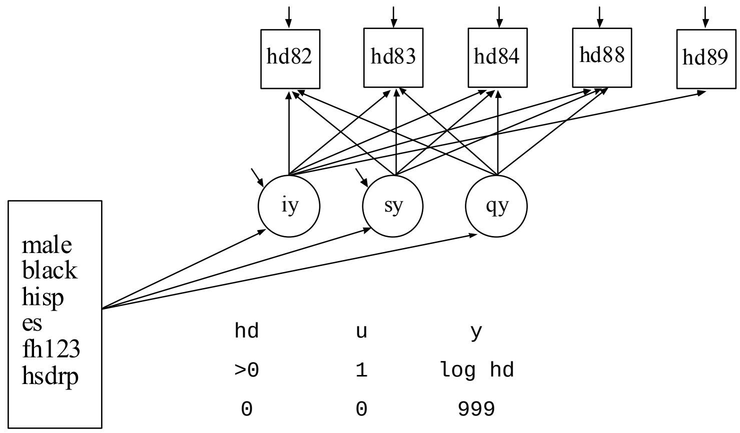


NLSY Heavy Drinking Data

- The data are from the National Longitudinal Study of Youth (NLSY), a nationally representative household study of 12,686 men and women born between 1957 and 1964.
- There are eight birth cohorts, but the current analysis considers only cohort 64 measured in 1982, 1983, 1984, 1988, 1989, and 1994 at ages 18, 19, 20, 24, and 25.
- The outcome is heavy drinking, measured by the question: How often have you had 6 or more drinks on one occasion during the last 30 days?
- The responses are coded as: never (0); once (1); 2 or 3 times (2); 4 or 5 times (3); 6 or 7 times (4); 8 or 9 times (5); and 10 or more times (6).
- Background variables include gender, ethnicity, early onset of regular drinking (es), family history of problem drinking, high school dropout and college education

100

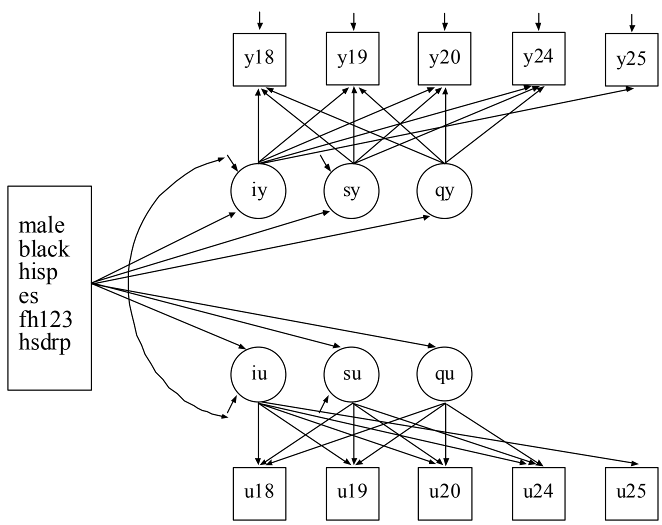
NLSY Heavy Drinking Data



hd	u	y
>0	1	log hd
0	0	999
999	999	999

101

NLSY Heavy Drinking Data



Input For NLSY Heavy Drinking

```
TITLE:     nlsy36425x25dep.inp
           cohort 64
           centering at 25
           hd82-hd89 (ages 18 - 25)
           log age scale: x_t = a*(ln(t-b) - ln(c-b)), where t is
           time, a and b are constants to fit the mean curve
           (chosen as a = 2 and b = 16), and c is the centering
           age, here set at 25.
DATA:      FILE = big.dat;
           FORMAT = 2f5, f2, t14, 5f7, t50, f8, t60, 6f1.0, t67,
           2f2.0, t71, 8f1.0, t79, f2.0, t82, 4f2.0;
DATA TWOPART:
           NAMES = hd82-hd89;
           BINARY = u18 u19 u20 u24 u25;
           CONTINUOUS = y18 y19 y20 y24 y25;
```


Input For NLSY Heavy Drinking (Continued)

```
VARIABLE:  NAMES = id houseid cohort weight82 weight83 weight84
           weight88 weight89 weight94 hd82 hd83 hd84 hd88 hd89
           hd94 dep89 dep94 male black hisp es fh1 fh23 fh123
           hsdrp coll ed89 ed94 cd89 cd94;
           USEOBSERVATIONS = cohort EQ 64 AND (coll GT 0 AND coll
           LT 20);
           USEV = male black hisp es fh123 hsdrp coll u18-u25
           y18-y25;
           CATEGORICAL = u18-u25;
           MISSING = .;
           AUXILIARY = hd82-hd89;
DEFINE:    CUT coll (12.1);
ANALYSIS:  ESTIMATOR = ML;
           ALGORITHM = INTEGRATION;
           COVERAGE = 0.09;
```

Input For NLSY Heavy Drinking (Continued)

```
MODEL:     iu su qu | u18@-3.008 u19@-2.197 u20@-1.621
           u24@-.235 u25@.000;
           iy sy qy | y18@-3.008 y19@-2.197 y20@-1.621 y24@-.235
           y25@.000;
           iu-qy on male black hisp es fh123 hsdrp coll;
OUTPUT:    TECH1 TECH4 TECH8 STANDARDIZED;
PLOT:      TYPE = PLOT3;
           SERIES = y18-y25(sy) | u18-u25(su);
```


Regular Growth Modeling of NLSY Heavy Drinking

Parameter	Estimate	S.E.	Est./S.E.
Regular growth modeling, treating outcome as continuous. Non-normality robust ML (MLR)			
i ON			
male	0.769	0.076	10.066
black	-0.336	0.083	-4.034
hisp	-0.227	0.103	-2.208
es	**0.291**	**0.128**	**2.283**
fh123	0.286	0.137	2.089
hsdrop	-0.024	0.104	-0.232
coll	-0.131	0.086	-1.527

Output Excerpts For Two-Part Growth Modeling of NLSY Heavy Drinking

Parameter	Estimate	S.E.	Est./S.E.
Two-part growth modeling			
iy ON			
male	0.329	0.058	5.651
black	-0.123	0.062	-1.986
hisp	-0.143	0.069	-2.082
es	**0.096**	**0.062**	**1.543**
fh123	0.219	0.076	2.894
hsdrop	0.093	0.063	1.466
coll	-0.030	0.056	-0.526

Output Excerpts For Two-Part Growth Modeling of NLSY Heavy Drinking (Continued)

Parameter	Estimate	S.E.	Est./S.E.
iu ON			
male	1.533	0.164	9.356
black	-0.705	0.172	-4.092
hisp	-0.385	0.199	-1.934
es	**0.471**	**0.194**	**2.430**
fh123	0.287	0.224	1.281
hsdrop	-0.191	0.183	-1.045
coll	-0.325	0.161	-2.017

NLSY Heavy Drinking Conclusions

As an example of differences in results between regular growth modeling and two-part growth modeling, consider the covariate es (early start, that is, early onset of regular drinking scored as 1 if the respondent had 2 or more drinks per week at age 14 or earlier):

Regular growth modeling says that es has a significant, positive influence on heavy drinking at age 25, increasing the frequency of heavy drinking.

Two-part growth modeling says that es has a significant, positive influence on the probability of heavy drinking at age 25, but among those who engage in heavy drinking at age 25 there is no significant difference in heavy drinking frequency with respect to es.

References

For references, see handouts for Topic 1, 2, 3, 4, and 9 at

http://www.statmodel.com/course_materials.shtml

For papers using special Mplus features, see

http://www.statmodel.com/papers.shtml