Mplus Short Courses
Day 2

Growth Modeling With Latent Variables Using Mplus

Linda K. Muthén
Bengt Muthén

1

Table Of Contents

2

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 versions
  - V1: November 1998
  - V2: February 2001
  - V3: March 2004
  - V4: February 2006
- Mplus team: Linda & Bengt Muthén, Thuy Nguyen, Tihomir Asparouhov, Michelle Conn

3

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

4

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

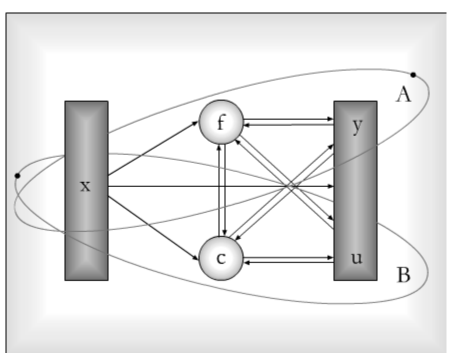
Categorical Latent Variables

- 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.

5

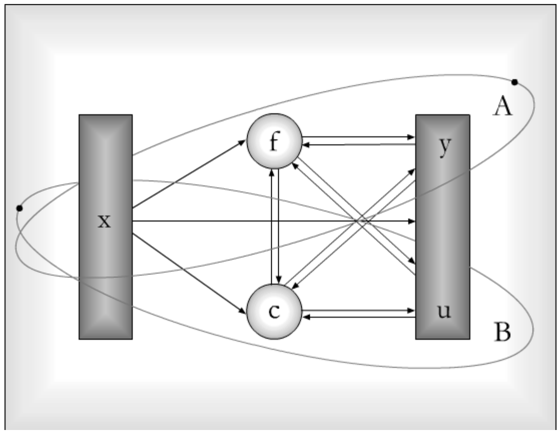
General Latent Variable Modeling Framework



- 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

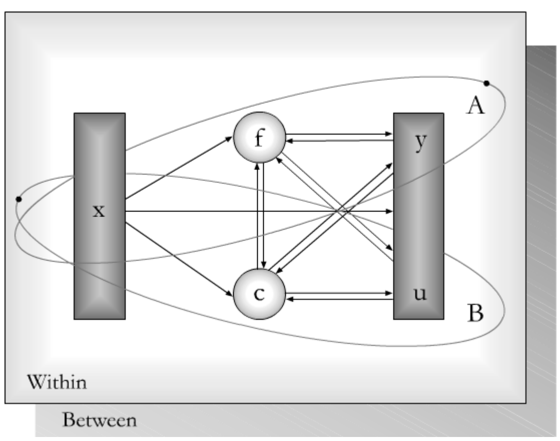
6

General Latent Variable Modeling Framework



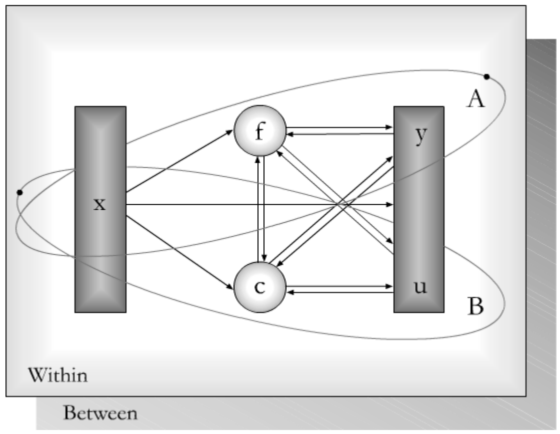
7

General Latent Variable Modeling Framework



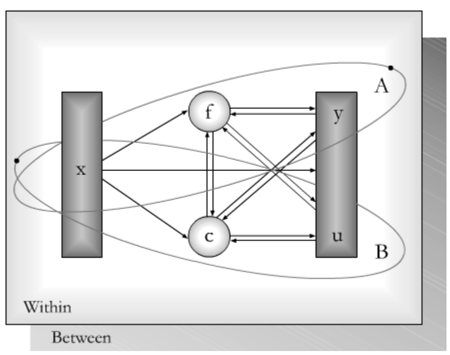
8

General Latent Variable Modeling Framework



9

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

10

Overview

Single-Level Analysis

	Cross-Sectional	Longitudinal
Continuous Observed And Latent Variables	***Day 1*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling	***Day 2*** Growth Analysis
Adding Categorical Observed And Latent Variables	***Day 3*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling Latent Class Analysis Factor Mixture Analysis Structural Equation Mixture Modeling	***Day 4*** Latent Transition Analysis Latent Class Growth Analysis Growth Analysis Growth Mixture Modeling Discrete-Time Survival Mixture Analysis Missing Data Analysis

11

Overview (Continued)

Multilevel Analysis

	Cross-Sectional	Longitudinal
Continuous Observed And Latent Variables	***Day 5*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling	***Day 5*** Growth Analysis
Adding Categorical Observed And Latent Variables	***Day 5*** Latent Class Analysis Factor Mixture Analysis	***Day 5*** Growth Mixture Modeling

12

Typical Examples Of Growth Modeling

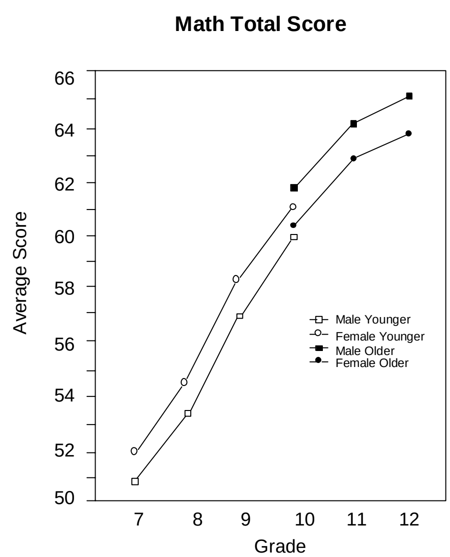
13

LSAY Data

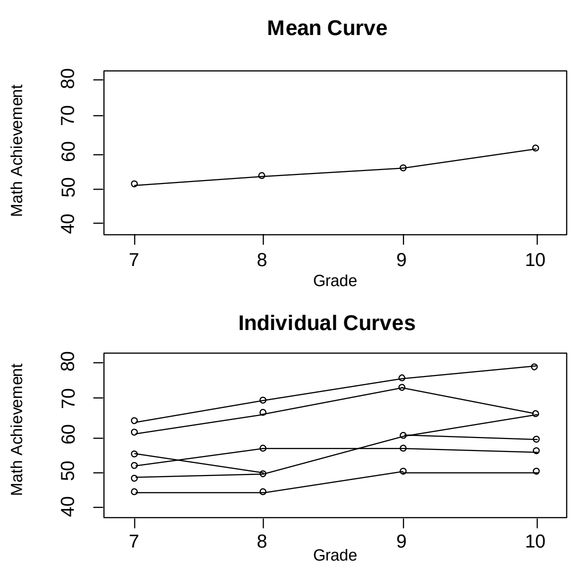
Longitudinal Study of American Youth (LSAY)

- Two cohorts measured each year beginning in 1987
  - Cohort 1 - Grades 10, 11, and 12
  - Cohort 2 - Grades 7, 8, 9, and 10
- 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

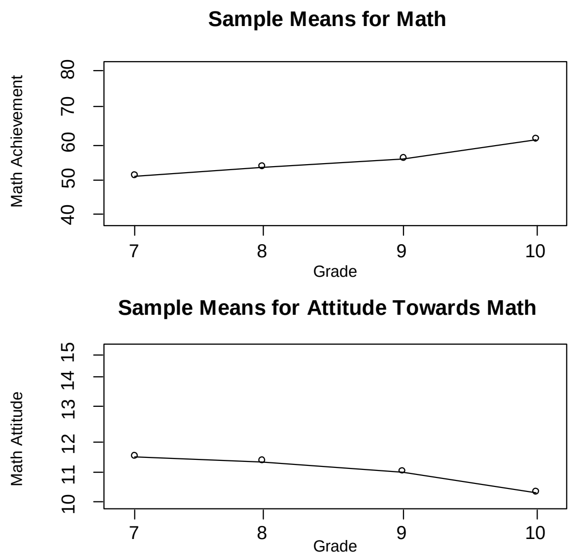
14


Math Total Score
Average Score
66
64
62
60
58
56
54
52
50
7
8
9
10
11
12
Grade
Male Younger
Female Younger
Male Older
Female Older

15


Mean Curve
Math Achievement
80
70
60
50
40
7
8
9
10
Grade
Individual Curves
Math Achievement
80
70
60
50
40
7
8
9
10
Grade

16



17

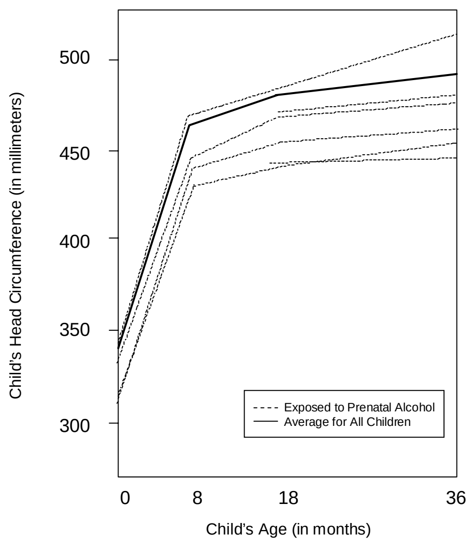
Maternal Health Project Data

Maternal Health Project (MHP)

- Mothers who drank at least three drinks a week during their first trimester plus a random sample of mothers who used alcohol less often
- Mothers measured at fourth month and seventh month of pregnancy, at delivery, and at 8, 18, and 36 months postpartum
- Offspring measured at 0, 8, 18 and 36 months
- Variables for mothers - demographic, lifestyle, current environment, medical history, maternal psychological status, alcohol use, tobacco use, marijuana use, other illicit drug use
- Variables for offspring - head circumference, height, weight, gestational age, gender, and ethnicity

18

MHP: Offspring Head Circumference



19

Basic Modeling Ideas

20

Longitudinal Data: Three Approaches

Three modeling approaches for the regression of outcome on time (n is sample size, T is number of timepoints):

- **Use all *n* x *T* data points to do a single regression analysis**: Gives an intercept and a slope estimate for all individuals - does not account for individual differences or lack of independence of observations
- **Use each individual's *T* data points to do *n* regression analyses:** Gives an intercept and a slope estimate for each individual. Accounts for individual differences, but does not account for similarities among individuals
- **Use all *n* x *T* data points to do a single random effect regression analysis:** Gives an intercept and a slope estimate for each individual. Accounts for similarities among individuals by stipulating that all individuals' random effects come from a single, common population and models the non-independence of observations as show on the next page

21

Individual Development Over Time

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

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

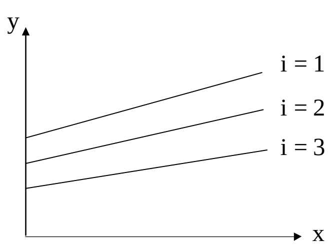
(2b) $\eta_{1i} = \alpha_1 + \gamma_1\, w_i + \zeta_{1i}$

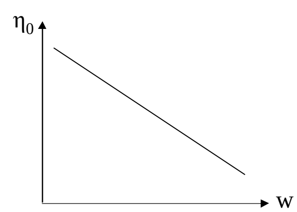
t = timepoint i = individual

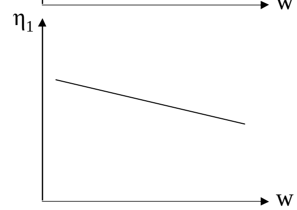
w = time-invariant covariate

y = outcome x = time score

η_0 = intercept η_1 = slope

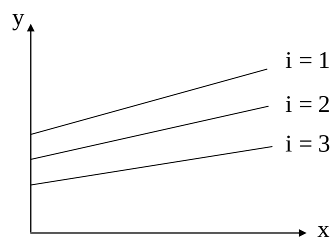


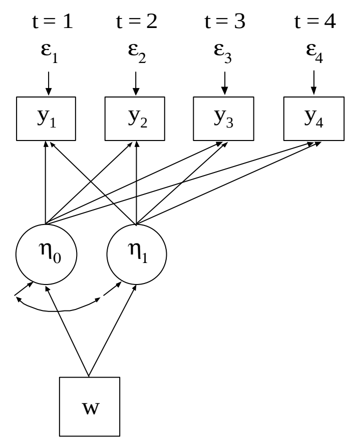




22

Individual Development Over Time





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

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

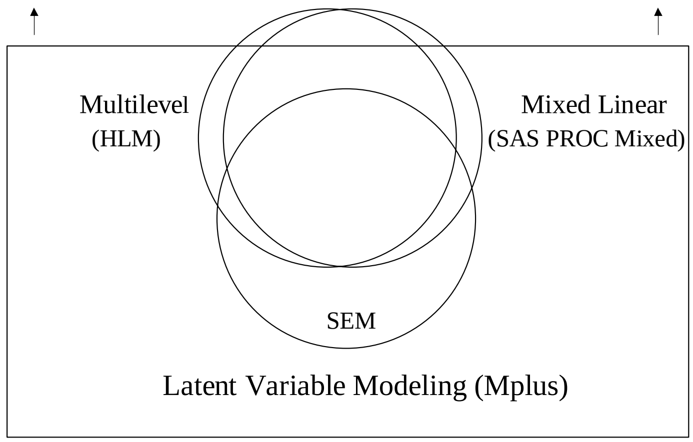
(2b) $\eta_{1i} = \alpha_1 + \gamma_1\, w_i + \zeta_{1i}$

23

Growth Modeling Frameworks

24

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

Random Effects: Multilevel And Mixed Linear Modeling

Individual i (i = 1, 2, …, n) observed at time point t (t = 1, 2, … T).

Multilevel model with two levels (e.g. Raudenbush & Bryk, 2002, HLM).

- Level 1: $y_{ti} = \eta_{0i} + \eta_{1i} x_{ti} + \kappa_i w_{ti} + \varepsilon_{ti}$ (39)
- Level 2: $\eta_{0i} = \alpha_0 + \gamma_0 w_i + \zeta_{0i}$ (40)
  $\eta_{1i} = \alpha_1 + \gamma_1 w_i + \zeta_{1i}$ (41)
  $\kappa_i = \alpha + \gamma w_i + \zeta_i$ (42)

27

Random Effects: Multilevel And Mixed Linear Modeling (Continued)

Mixed linear model:

$$y_{ti} = \textit{fixed part} + \textit{random part} \quad (43)$$
$$= \alpha_0 + \gamma_0 w_i + (\alpha_1 + \gamma_1 w_i)\, x_{ti} + (\alpha + \gamma w_i)\, w_{ti} \quad (44)$$
$$+ \zeta_{0i} + \zeta_{1i} x_{ti} + \zeta_i w_{ti} + \varepsilon_{ti}\,. \quad (45)$$

E.g. "*time* x w_i" refers to γ_1 (e.g. Rao, 1958; Laird & Ware, 1982; Jennrich & Sluchter, 1986; Lindstrom & Bates, 1988; BMDP5V; Goldstein, 1995, MLn; SAS PROC MIXED - Littell et al. 1996 and Singer, 1999).

28

Random Effects: SEM And Multilevel Modeling

SEM (Tucker, 1958; Meredith & Tisak, 1990; McArdle & Epstein 1987; SEM software):

Measurement part:

$$y_{ti} = \eta_{0i} + \eta_{1i}\, x_t + \kappa_t\, w_{ti} + \varepsilon_{ti}\,. \tag{46}$$

Compare with level 1 of multilevel:

$$y_{ti} = \eta_{0i} + \eta_{1i}\, x_{ti} + \kappa_i\, w_{ti} + \varepsilon_{ti}\,. \tag{47}$$

Multilevel approach:

- x_{ti} as data: Flexible individually-varying times of observation
- Slopes for time-varying covariates vary over individuals

29

Random Effects: SEM And Multilevel Modeling (Continued)

SEM approach:

- x_t as parameters: Flexible growth function form
- Slopes for time-varying covariates vary over time points

Structural part (same as level 2, except for κ_t):

$$\eta_{0i} = \alpha_0 + \gamma_0\, w_i + \zeta_{0i}\,, \tag{48}$$

$$\eta_{1i} = \alpha_1 + \gamma_1\, w_i + \zeta_{1i}\,, \tag{49}$$

κ_t *not involved (parameter).*

30

Random Effects: Mixed Linear Modeling And SEM

Mixed linear model in matrix form:

$$\boldsymbol{y}_i = (y_{1i}, y_{2i}, \ldots, y_{Ti})' \qquad (51)$$

$$= \boldsymbol{X}_i \boldsymbol{\alpha} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i . \qquad (52)$$

Here, ***X, Z*** are design matrices with known values, $\boldsymbol{\alpha}$ contains fixed effects, and ***b*** contains random effects. Compare with (43) - (45).

31

Random Effects: Mixed Linear Modeling And SEM (Continued)

SEM in matrix form:

$$y_i = \nu + \Lambda\, \eta_i + K\, x_i + \varepsilon_i\,, \qquad (53)$$

$$\eta_i = \alpha + B\, \eta_i + \Gamma\, x_i + \zeta_i\,. \qquad (54)$$

$$y_i = \textit{fixed part} + \textit{random part}$$

$$= \mathrm{v} + \Lambda\, (I - B)^{-1}\, \alpha + \Lambda\, (I - B)^{-1}\, \Gamma\, x_i + K\, x_i$$

$$+ \Lambda\, (I - B)^{-1}\, \zeta_i + \varepsilon_i\,.$$

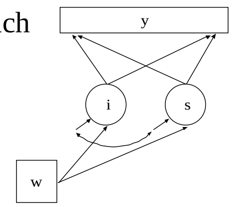
Assume $x_{ti} = x_t$, $\kappa_i = \kappa_t$ in (39). Then (39) is handled by (53) and (40) – (41) are handled by (54), putting x_t in Λ and w_{ti}, w_i in x_i.

Need for Λ_i, K_i, B_i, Γ_i.

32

Growth Modeling Approached In Two Ways: Data Arranged As Wide Versus Long

- Wide: Multivariate, Single-Level Approach

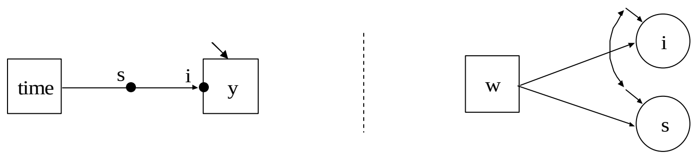


$y_{ti} = i_i + s_i \times \text{time}_{ti} + \varepsilon_{ti}$

i_i regressed on w_i
s_i regressed on w_i

- Long: Univariate, 2-Level Approach (CLUSTER = id)

Within | Between



The intercept *i* is called *y* in Mplus

33

Multilevel Modeling In A Latent Variable Framework

Integrating multilevel and SEM analyses (Asparouhov & Muthén, 2002).

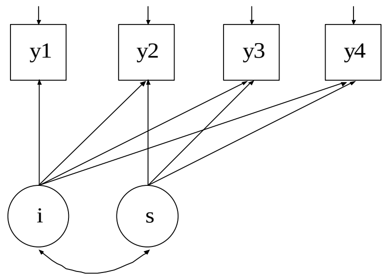
Flexible combination of random effects and other latent variables:

- Multilevel models with random effects (intercepts, slopes)
  - Individually-varying times of observation read as data
  - Random slopes for time-varying covariates
- SEM with factors on individual and cluster levels
- Models combining random effects and factors, e.g.
  - Cluster-level latent variable predictors with multiple indicators
  - Individual-level latent variable predictors with multiple indicators
- Special applications
  - Random coefficient regression (no clustering; heteroscedasticity)
  - Interactions between continuous latent variables and observed variables

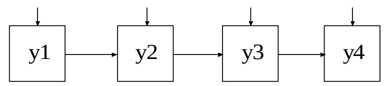
34

Alternative Models For Longitudinal Data

Growth Curve Model



Auto-Regressive Model



Hybrid Models

Curran & Bollen (2001)
McArdle & Hamagami (2001)

35

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

36