Beyond Conventional HLM

Bengt Muthen
UCLA
bmuthen@ucla.edu

AERA HLM SIG, April 8, 2006

1

HLM ≠ HLM

2

Mplus = Multilevel modeling and much more

Visit www.statmodel.com to check out

- Free web videos
- Recent papers
- Mplus Discussion
- Short Courses

3

Multilevel Analysis Areas

- **Multilevel regression analysis**
- Multilevel path analysis
- Multilevel factor analysis
- Multilevel SEM
- **Multilevel growth modeling**
- Multilevel 2-part growth modeling
- Multilevel survival analysis
- Multilevel latent class analysis
- Multilevel growth mixture modeling

Multilevel modeling needed for all analysis areas, not just the bolded ones

4

Multilevel Regression

- But, is there really anything new to say about multilevel regression beyond conventional HLM?

  Yes!

5

Two-Level Regression

- Raudenbush & Bryk (2002; p. 140, Table 5.11, group-mean centering):

$$Y_{ij} = \beta_{0j} + \beta_w(X_{ij} - \bar{X}_{.j}) + \varepsilon_{ij} \quad (1)$$

$$\beta_{0j} = \mu + \beta_b \bar{X}_{.j} + \varepsilon_j \quad (2)$$

Contextual effect: $\beta_c = \beta_b - \beta_w$

- Mplus considers the latent variable decomposition:

$$X_{ij} = X_{jb} + X_{ijw} \quad (\bar{X}_{.j} \approx X_{jb},\ X_{ij} - \bar{X}_{.j} \approx X_{ijw})$$

$$Y_{ij} = Y_{jb} + Y_{ijw}$$

$$= \underbrace{\mu + \beta_b X_{jb} + \varepsilon_j}_{\text{Between} = \beta_{0j}} + \underbrace{\beta_w X_{ijw} + \varepsilon_{ij}}_{\text{Within}}$$

Compare to (1) and (2)

6

Between-Level Bias In Two-Level Regression

Reference: Asparouhov & Muthen (2006). Constructing covariates in multilevel regression.

$$\text{Between}: Z_{jb} = \bar{X}_{.j} = \frac{1}{l}\sum_{i=1}^{l} X_{ij} = X_{jb} + \frac{1}{l}\sum_{i=1}^{l} X_{ijw}$$

$$\text{Bias}: E(\hat{\beta}_b) - \beta_b = \frac{Cov(Y_{jb}, Z_{jb})}{Var(Z_{jb})} - \beta_b$$

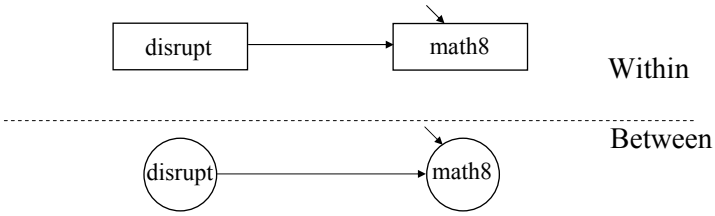
$$= \frac{Cov(Y_{jb}, X_{jb})}{Var(X_{jb}) + Var(X_{ijw})/l} - \beta_b$$

$$= \beta_b \frac{\psi_b}{\psi_b + \psi_w/l} - \beta_b = \mathbf{-\beta_b \frac{\psi_w}{l\psi_b}}$$

- German TIMSS: n = 1,980 students, 98 schools, average school size = 20 (=l)
- y(math8) icc =0.55, x(disrupt) icc = 0.21. The disrupt icc results in $\psi_w/\psi_b = 4$
- Bias in the naive β_b estimate = -1×(-1) × 4/20 = 0.2

7

Two-Level Regression In The German TIMSS (Continued)



Mplus input:

```
Model:

%Within%
math8 ON disrupt;
%Between%
math8 ON disrupt;
```

Estimates:
Mplus: β_b = -1.35 (.36), contextual effect β_c = -1.25
Naive: β_b = -1.18 (.29), contextual effect β_c = -1.08

8

Two-Level Path Analysis

9

LSAY Data

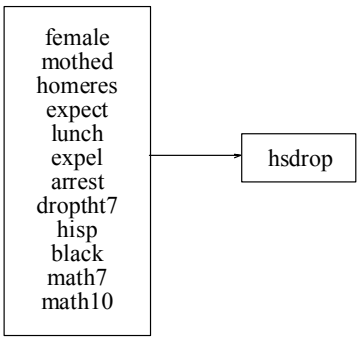
Longitudinal Study of American Youth (LSAY)

- 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

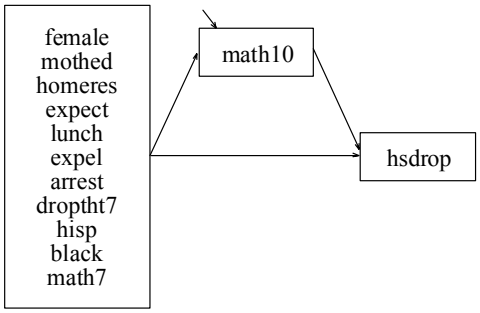
10

Path Analysis With A Binary Outcome And A Continuous Mediator With Missing Data

Logistic Regression

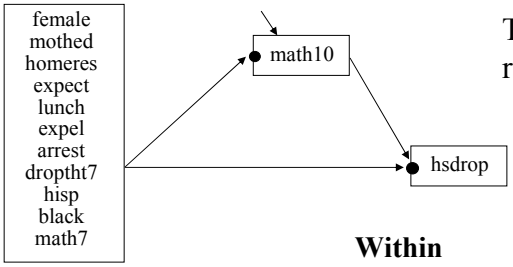


Path Model



11

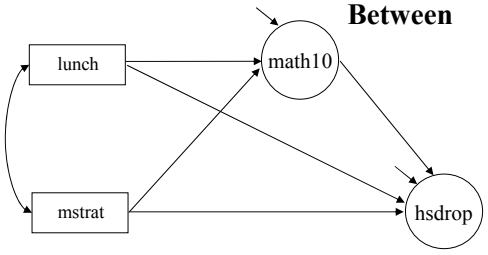
Two-Level Path Analysis



Two-level linear and logistic regression combined

Within

Between

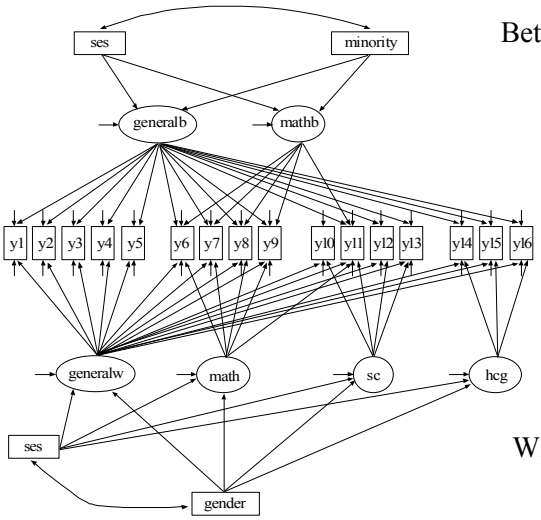


Lunch: % of students eligible for full lunch assistance

Mstrat: ratio of students to full time math teachers

12

Two-Level Factor Analysis

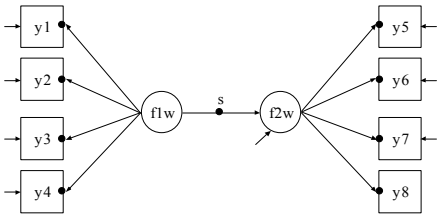


Between

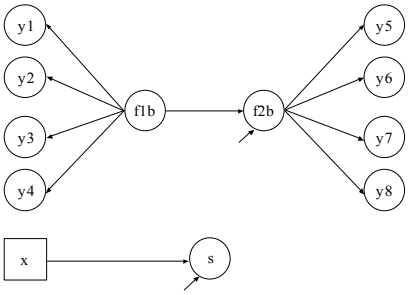
Within

13

Two-Level SEM



Within



Between

14

Growth Modeling

- Is there really anything new to be said about growth modeling beyond conventional HLM?

  Yes!

15

Longitudinal Study Of Frequency Of Heavy Drinking Ages 18 - 25

The alcohol data were taken from the National Longitudinal Survey of Youth (NLSY) that is sponsored by the Bureau of Labor Statistics. These data are made available to the public by Ohio State University. The data were obtained as a multistage probability sample with oversampling of blacks, Hispanics, and economically disadvantaged non-blacks and non-Hispanics.

Data for the analysis include 1172 individuals of cohort 1964 who in 1989 gave a positive answer to the question “Have you ever had a drink of an alcoholic beverage?" Alcohol outcomes were recorded at ages 18, 19, 20, 24, and 25. Covariates include gender, ethnicity, family history or problem drinking, early onset of drinking (age 14 or younger), high school dropout, and college education.

16

Longitudinal Study Of Frequency Of Heavy Drinking Ages 18 – 25 (Continued)

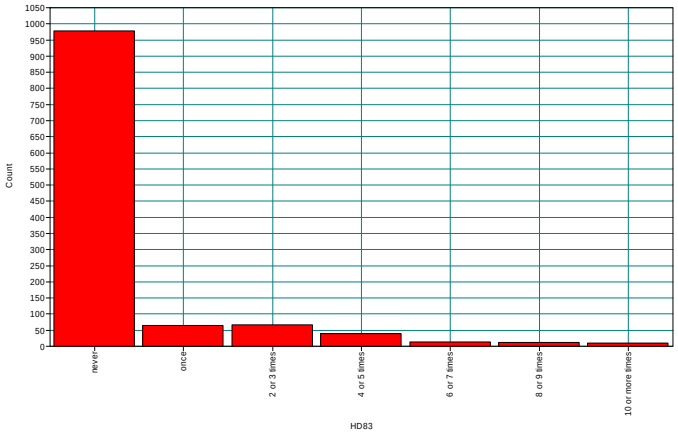
The outcome measured at the five occasions was the response to the question: "How often have you had 6 or more drinks on one occasion during the last 30 days?" The responses were recorded as:

0: never
1: once
2: 2 or 3 times
3: 4 or 5 times
4: 6 or 7 times
5: 8 or 9 times
6: 10 or more times

17

Longitudinal Study Of Frequency Of Heavy Drinking Ages 18 – 25 (Continued)

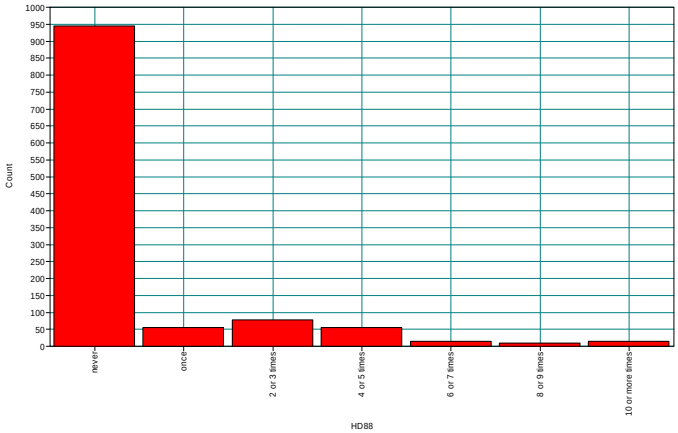
HD83 Histogram – Heavy Drinking At Age 19



18

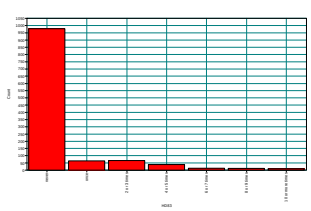
Longitudinal Study Of Frequency Of Heavy Drinking Ages 18 – 25 (Continued)

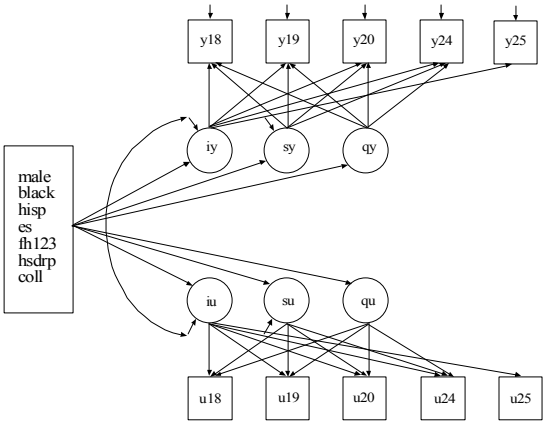
HD88 Histogram – Heavy Drinking At Age 24



19

Two-Part Growth Modeling Of Frequency Of Heavy Drinking Ages 18 – 25





Olsen and Schafer (2001)

20

Two-Part Growth Modeling Of Frequency Of Heavy Drinking Ages 18 – 25

Regular growth modeling, treating outcome as continuous. Non-normality robust ML (MLR)

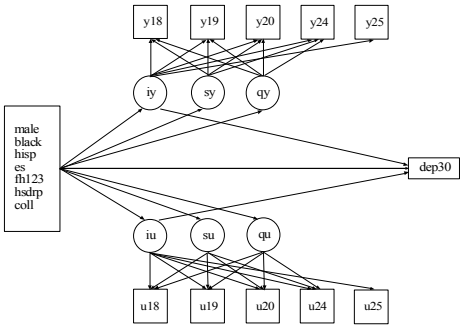
	Estimates	S.E.	Est./S.E.	Std	StdYX
I ON					
MALE	0.769	0.076	10.066	0.653	0.326
BLACK	-0.336	0.083	-4.034	-0.286	-0.127
HISP	-0.227	0.103	-2.208	-0.193	-0.071
ES	0.291	0.128	2.283	0.247	0.088
FH123	0.286	0.137	2.089	0.243	0.075
HSDRP	-0.024	0.104	-0.232	-0.020	-0.008
COLL	-0.131	0.086	-1.527	-0.111	-0.052

Two-Part growth modeling

IY ON					
MALE	0.262	0.052	5.065	0.610	0.305
BLACK	-0.096	0.059	-1.619	-0.223	-0.099
HISP	-0.130	0.066	-1.963	-0.301	-0.111
ES	0.082	0.062	1.333	0.191	0.068
FH123	0.213	0.076	2.815	0.495	0.152
HSDRP	0.084	0.065	1.289	0.195	0.078
COLL	-0.015	0.053	-0.280	-0.035	-0.016
IU ON					
MALE	2.041	0.176	11.594	0.949	0.474
BLACK	-1.072	0.203	-5.286	-0.499	-0.222
HISP	-0.545	0.234	-2.331	-0.254	-0.093
ES	0.364	0.234	1.560	0.169	0.060
FH123	0.562	0.275	2.045	0.262	0.080
HSDRP	-0.238	0.216	-1.103	-0.111	-0.044
COLL	-0.259	0.196	-1.317	-0.120	-0.056

21

Two-Part Growth Modeling With A Distal Outcome



	Estimates	S.E.	Est./S.E.	Std	StdYX
DEP30 ON					
IU	0.440	0.141	3.120	0.949	0.427
IY	0.874	0.736	1.187	0.373	0.168
DEP30 ON					
MALE	-0.098	0.291	-0.337	-0.098	-0.022
BLACK	0.415	0.294	1.414	0.415	0.083
HISP	0.025	0.326	0.075	0.025	0.004
ES	0.237	0.286	0.830	0.237	0.038
FH123	0.498	0.325	1.532	0.498	0.069
HSDRP	0.565	0.312	1.812	0.565	0.101
COLL	-0.384	0.276	-1.390	-0.384	-0.081

22

Two-Part Growth Modeling Extensions In Mplus

Distal outcome

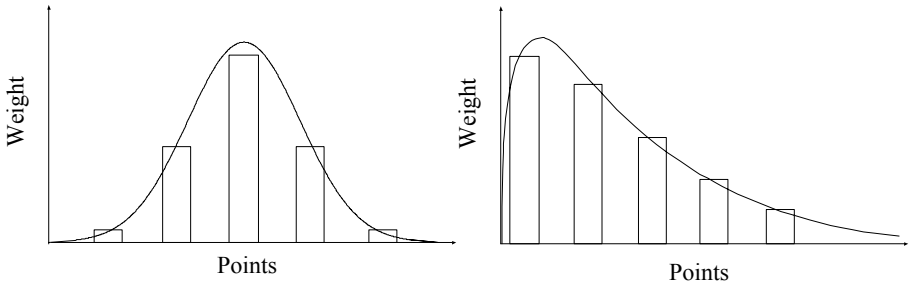
Parallel processes

Trajectory classes (mixtures)

Multilevel

23

Non-Parametric Estimation Of The Random Effect Distribution Using Mixtures



Estimated weights and points
(class probabilities and class means)

24

Multilevel Growth Modeling
Three-Level Analysis

- Is there really anything new to be said about multilevel growth modeling beyond conventional HLM?

  Yes!

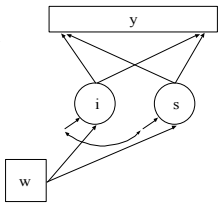
25

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

- Wide: Multivariate, Single-Level Approach

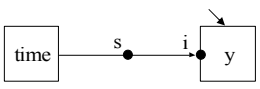
  $y_{ti} = i_i + s_i \times \text{time}_{ti} + \varepsilon_{ti}$

  i_i regressed on w_i
  s_i regressed on w_i

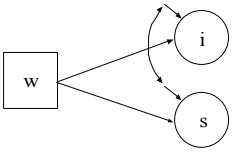


- Long: Univariate, 2-Level Approach (CLUSTER = id)

  Within



  Between



  The intercept i is called y in Mplus

26

Growth Modeling Approached in Two Ways: Data Arranged As Wide Versus Long (Continued)

- Wide (one person):

			t1	t2	t3	t1	t2	t3	
Person i:		id	y1	y2	y3	x1	x2	x3	w

- Long (one cluster):

Person i:	t1	id	y1	x1	w
	t2	id	y2	x2	w
	t3	id	y3	x3	w

27

Three-Level Modeling In Multilevel Terms

Time point t, individual i, cluster j.

y_{tij} : individual-level, outcome variable
a_{1tij} : individual-level, time-related variable (age, grade)
a_{2tij} : individual-level, time-varying covariate
x_{ij} : individual-level, time-invariant covariate
w_j : cluster-level covariate

Three-level analysis (Mplus considers Within and Between)

$$\textit{Level 1 (Within)} : y_{tij} = \pi_{0ij} + \pi_{1ij}\, a_{1tij} + \pi_{2tij}\, a_{2tij} + e_{tij}, \quad (1)$$

$$\textit{Level 2 (Within)} : \begin{cases} \pi_{0ij} = \beta_{00j} + \beta_{01j}\, x_{ij} + r_{0ij}, \\ \pi_{1ij} = \beta_{10j} + \beta_{11j}\, x_{ij} + r_{1ij}, \\ \pi_{2tij} = \beta_{20tj} + \beta_{21tj}\, x_{ij} + r_{2tij}. \end{cases} \quad (2)$$

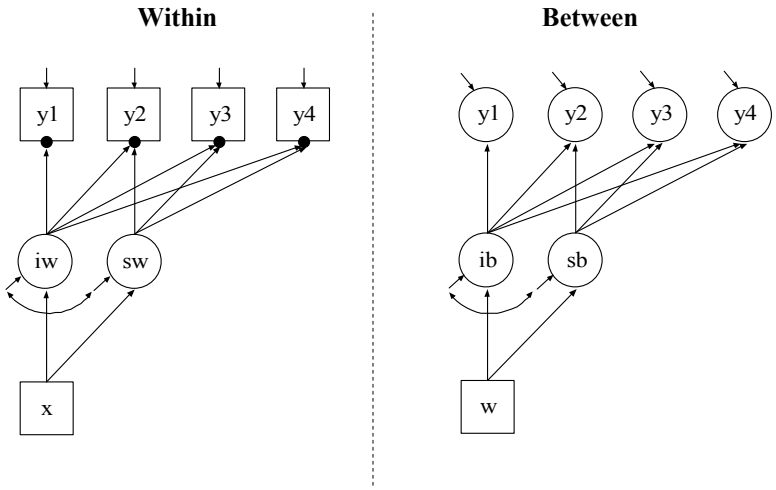
iw

$$\textit{Level 3 (Between)} : \begin{cases} \beta_{00j} = \gamma_{000} + \gamma_{001}\, w_j + u_{00j}, \\ \beta_{10j} = \gamma_{100} + \gamma_{101}\, w_j + u_{10j}, \\ \beta_{20tj} = \gamma_{200t} + \gamma_{201t}\, w_j + u_{20tj}, \\ \beta_{01j} = \gamma_{010} + \gamma_{011}\, w_j + u_{01j}, \\ \beta_{11j} = \gamma_{110} + \gamma_{111}\, w_j + u_{11j}, \\ \beta_{21tj} = \gamma_{2t0} + \gamma_{2t1}\, w_j + u_{2tj}. \end{cases} \quad (3)$$

ib

28

Two-Level Growth Modeling (Three-Level Modeling)



29

Conventional Growth Modeling of LSAY Math Achievement, G7-G10:

Chi-square (32) = 179.58. Level 3 Estimates and SEs:

	Estimates	S.E.	Est./S.E.	Std	StdYX
SB ON					
LUNCH	-1.271	0.402	-3.160	-1.919	-0.397
MSTRAT	1.724	1.022	1.688	2.605	0.185
Residual Variances					
MATH7	0.000	0.000	0.000	0.000	0.000
MATH8	0.000	0.000	0.000	0.000	0.000
MATH9	0.000	0.000	0.000	0.000	0.000
MATH10	0.000	0.000	0.000	0.000	0.000
IB	5.866	1.401	4.186	0.736	0.736
SB	0.354	0.138	2.564	0.809	0.809

Allowing Time-Specific Level 3 Residual Variances:

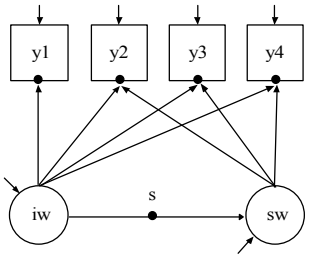
Chi-square (28) = 83.69. Level 3 Estimates and SEs:

SB ON					
LUNCH	-1.312	0.367	-3.576	-2.495	-0.516
MSTRAT	2.281	0.771	2.957	4.338	0.308
Residual Variances					
MATH7	1.396	0.749	1.863	1.396	0.159
MATH8	1.414	0.480	2.946	1.414	0.154
MATH9	0.382	0.381	1.002	0.382	0.042
MATH10	-0.121	0.518	-0.234	-0.121	-0.012
IB	5.211	1.410	3.694	0.704	0.704
SB	0.177	0.155	1.143	0.640	0.640

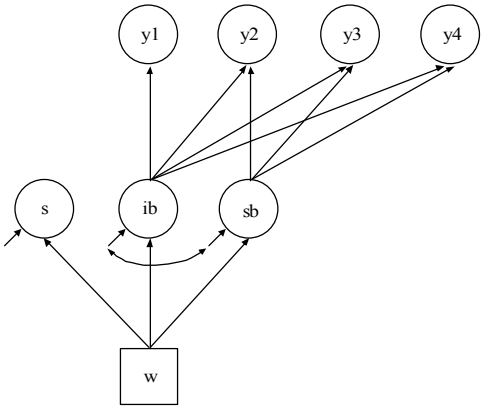
30

Multilevel Modeling With A Random Slope For Latent Variables

Student (Within)

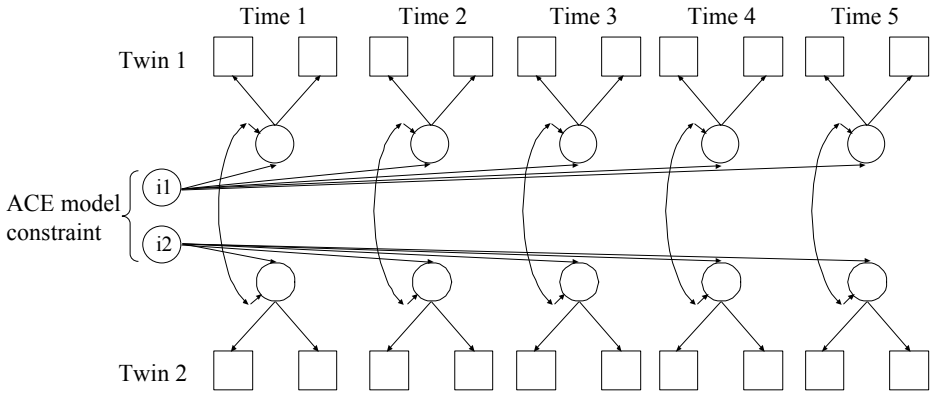


School (Between)



31

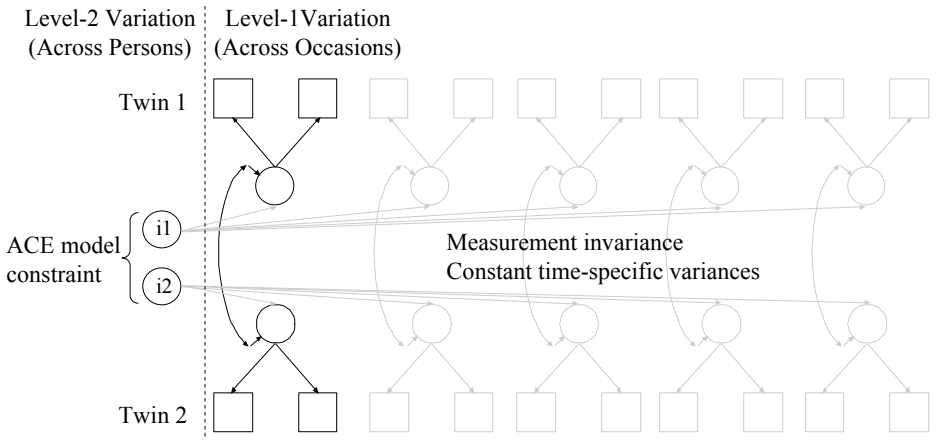
Multiple Indicator Growth: Wide Data Format, Single-Level Approach



20 variables, 12 factors, 10 dimensions of integration for ML
ML very hard, WLS easy

32

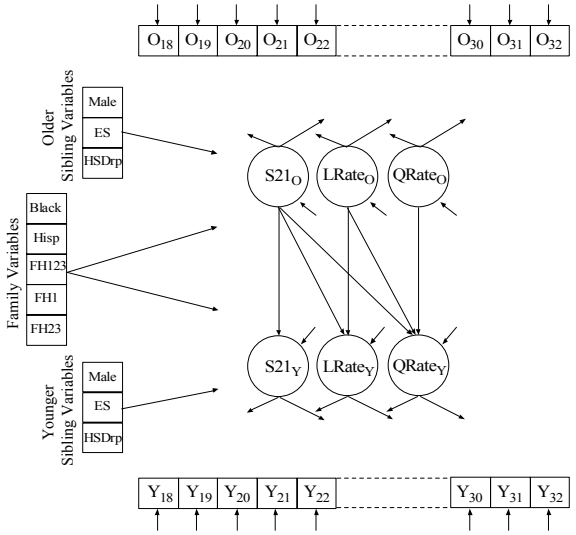
Long Format, Two-Level Approach



4 variables, 2 Level-2 and 2 Level-1 factors, 4 dimensions of integration for ML
ML feasible, WLS in development

33

Figure 1. A Longitudinal Growth Model of Heavy Drinking for Two-Sibling Families



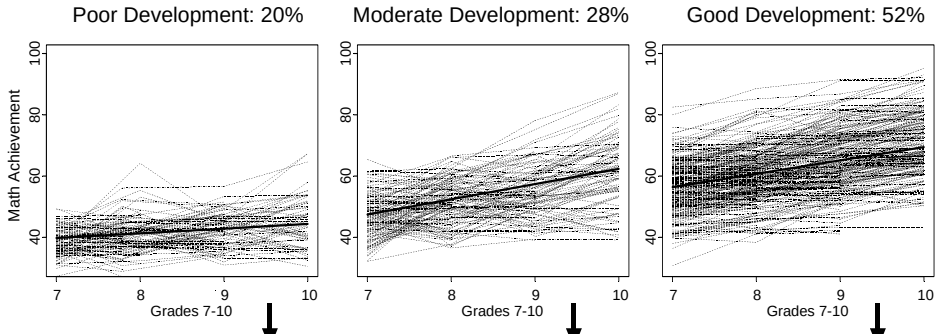
Source: Khoo, S.T. & Muthen, B. (2000). Longitudinal data on families: Growth modeling alternatives. Multivariate Applications in Substance Use Research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds.), Hillsdale, N.J.: Erlbaum, pp. 43-78.

34

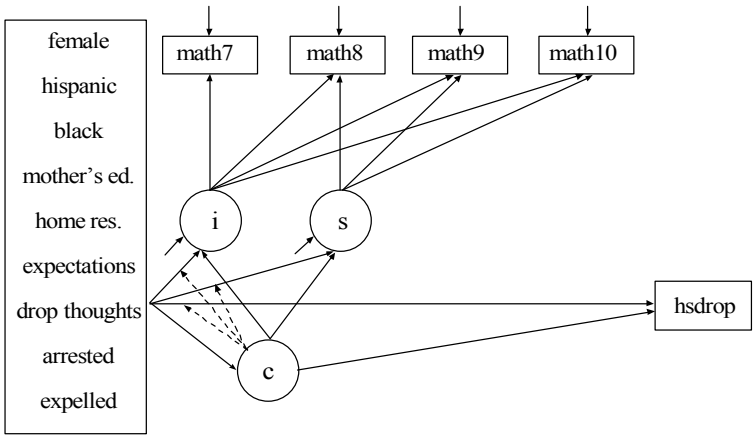
Multilevel Growth Mixture Modeling

35

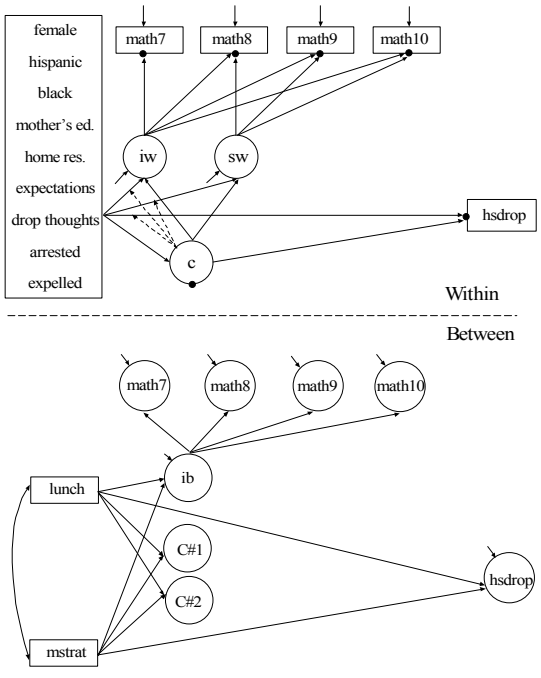
Growth Mixture Modeling: LSAY Math Achievement Trajectory Classes And The Prediction Of High School Dropout



36


female
hispanic
black
mother's ed.
home res.
expectations
drop thoughts
arrested
expelled
math7
math8
math9
math10
i
s
c
hsdrop

37


female
hispanic
black
mother's ed.
home res.
expectations
drop thoughts
arrested
expelled
math7
math8
math9
math10
iw
sw
c
hsdrop
Within
Between
math7
math8
math9
math10
ib
lunch
C#1
C#2
mstrat
hsdrop

38

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Model Results

	Estimates	S.E.	Est./S.E.
Within Level			
C#1 ON			
FEMALE	-0.751	0.188	-3.998
HISP	0.094	0.705	0.133
BLACK	0.900	0.385	2.339
MOTHED	-0.003	0.106	-0.028
HOMERES	-0.060	0.069	0.864
EXPECT	-0.251	0.074	-3.406
DROPTHT7	1.616	0.451	3.583
EXPEL	0.698	0.337	2.068
ARREST	1.093	0.384	2.842

39

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Model Results

	Estimates	S.E.	Est./S.E.
Between Level			
C#1 ON			
LUNCH	2.265	0.706	3.208
MSTRAT	-2.876	2.909	-0.988
C#2 ON			
LUNCH	-0.088	1.343	-0.065
MSTRAT	-0.608	2.324	-0.262

40

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Model Results

	Estimates	S.E	Est./S.E.
Within Level			
HSDROP ON			
FEMALE	0.521	0.232	2.251
HISP	0.208	0.322	0.647
BLACK	-0.242	0.256	-0.944
MOTHED	-0.434	0.121	-3.583
HOMERES	-0.089	0.052	-1.716
EXPECT	-0.333	0.052	-6.417
DROPTHT7	0.629	0.320	1.968
EXPEL	1.212	0.195	6.225
ARREST	0.157	0.263	0.597

41

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
Between Level					
CLASS 1					
IB ON					
LUNCH	-1.805	1.310	-1.378	-0.851	-0.176
MSTRAT	-13.365	3.086	-4.331	-6.299	-0.448
HSDROP ON					
LUNCH	1.087	0.543	2.004	1.087	0.290
MSTRAT	-0.178	1.478	-0.120	-0.178	-0.016
IB WITH					
HSDROP	-0.416	0.328	-1.267	-0.196	-0.253

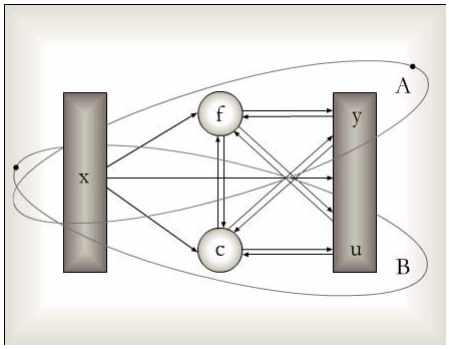
42

News In Mplus Version 4

- Data Twopart, Widetolong, Longtowide, Missing
- Model Constraint
- Type = Complex Twolevel (Stratification, weights, cluster 1, cluster 2)
- Twolevel Model Indirect
- Bootstrapped likelihood ratio test for mixtures
- Continuous-time survival analysis fully integrated

43

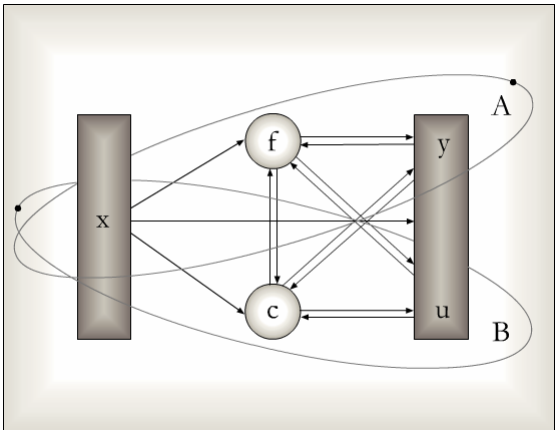
General Latent Variable Modeling Framework



- 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

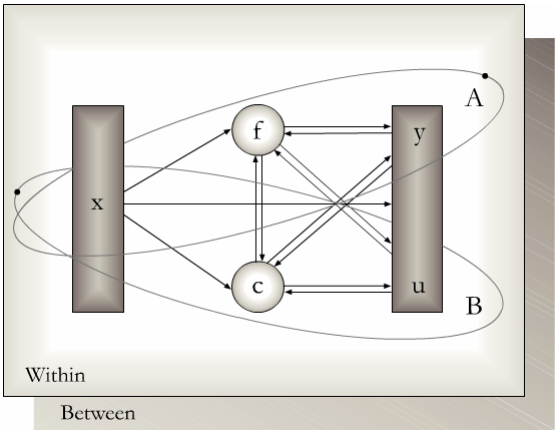
44

General Latent Variable Modeling Framework



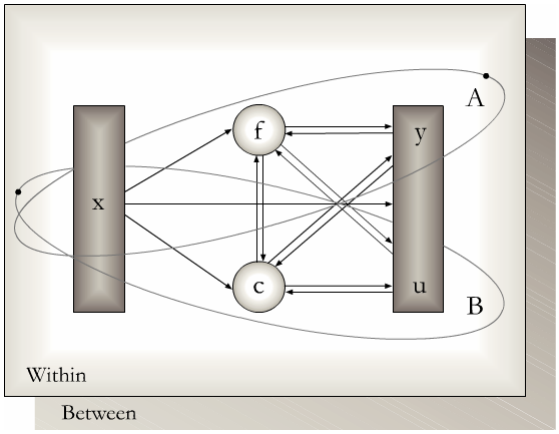
45

General Latent Variable Modeling Framework



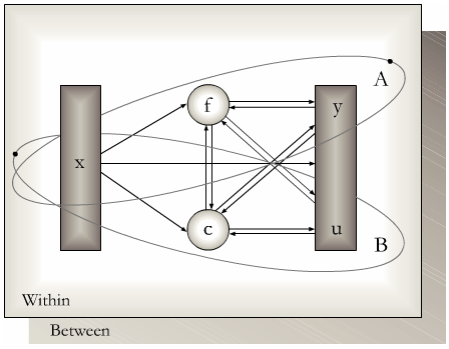
46

General Latent Variable Modeling Framework



47

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

48