

Mplus Short Courses  
Topics 7 and 8, part A

**Multilevel Modeling With Latent  
Variables Using Mplus**

Linda K. Muthén  
Bengt Muthén

Copyright © 2008 Muthén & Muthén

[www.statmodel.com](http://www.statmodel.com)

8/14/2007

1

**Table Of Contents**

General Latent Variable Modeling Framework	4
Complex Survey Data Analysis	11
Intraclass Correlation	12
Design Effects	14
Two-Level Regression Analysis	23
Two-Level Logistic Regression	44
Two-Level Path Analysis	50
Two-Level Factor Analysis	65
SIMS Variance Decomposition	77
Aggression Items	82
Two-Level Factor Analysis With Covariates	86
Multiple Group, Two-Level Factor Analysis	106
Two-Level SEM	122
Practical Issues Related To The Analysis Of Multilevel Data	133
Technical Aspects Of Multilevel Modeling	136
Multivariate Approach To Multilevel Modeling	145
Twin Modeling	150
Multilevel Growth Models	152
Three-Level Modeling	156
Multilevel Discrete-Time Survival Analysis	175
References	180

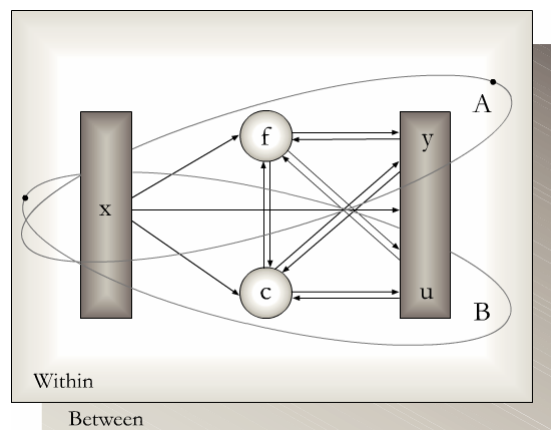
2

## 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

## General Latent Variable Modeling Framework



4

## Mplus

Several programs in one

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

Fully integrated in the general latent variable framework

5

## Overview

### Single-Level Analysis

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

6

## Overview (Continued)

### Multilevel Analysis

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

7

## Analysis With Multilevel Data

Used when data have been obtained by cluster sampling and/or unequal probability sampling to avoid biases in parameter estimates, standard errors, and tests of model fit and to learn about both within- and between-cluster relationships.

### Analysis Considerations

- Sampling perspective
  - Aggregated modeling – SUDAAN
    - TYPE = COMPLEX
      - Clustering, sampling weights, stratification (Asparouhov, 2005)

8

## **Analysis With Multilevel Data (Continued)**

- Multilevel perspective
  - Disaggregated modeling – multilevel modeling
    - TYPE = TWOLEVEL
      - Clustering, sampling weights, stratification
  - Multivariate modeling
    - TYPE = GENERAL
      - Clustering, sampling weights, stratification
- Combined sampling and multilevel perspective
  - TYPE = COMPLEX TWOLEVEL
    - Clustering, sampling weights, stratification

9

## **Analysis With Multilevel Data (Continued)**

### **Analysis Areas**

- Multilevel regression analysis
- Multilevel path analysis
- Multilevel factor analysis
- Multilevel SEM
- Multilevel growth modeling
- Multilevel latent class analysis
- Multilevel latent transition analysis
- Multilevel growth mixture modeling

10

## Complex Survey Data Analysis

11

## Intraclass Correlation

Consider nested, random-effects ANOVA for unit  $i$  in cluster  $j$ ,

$$y_{ij} = \nu + \eta_j + \varepsilon_{ij}; i = 1, 2, \dots, n_j; j = 1, 2, \dots, J. \quad (44)$$

Random sample of  $J$  clusters (e.g. schools).

With timepoint as  $i$  and individual as  $j$ , this is a repeated measures model with random intercepts.

Consider the covariance and variances for cluster members  $i = k$  and  $i = l$ ,

$$\text{Cov}(y_{kj}, y_{lj}) = V(\eta), \quad (45)$$

$$V(y_{kj}) = V(y_{lj}) = V(\eta) + V(\varepsilon), \quad (46)$$

resulting in the intraclass correlation

$$\rho(y_{kj}, y_{lj}) = V(\eta) / [V(\eta) + V(\varepsilon)]. \quad (47)$$

Interpretation: Between-cluster variability relative to total variation, intra-cluster homogeneity.

12

## NLSY Household Clusters

Household Type (# of respondents)	# of Households*	Intraclass Correlations for Siblings	
		Year	Heavy Drinking
Single	5,944	1982	0.19
Two	1,985	1983	0.18
Three	634	1984	0.12
Four	170	1985	0.09
Five	32	1988	0.04
Six	5	1989	0.06

Total number of households: 8,770  
 Total number of respondents: 12,686  
 Average number of respondents per household: 1.4

\*Source: NLS User's Guide, 1994, p.247

13

## Design Effects

Consider cluster sampling with equal cluster sizes and the sampling variance of the mean.

$V_C$ : correct variance under cluster sampling

$V_{SRS}$ : variance assuming simple random sampling

$V_C \geq V_{SRS}$  but cluster sampling more convenient, less expensive.

$$DEFF = V_C / V_{SRS} = 1 + (s - 1) \rho, \quad (47)$$

where  $s$  is the common cluster size and  $\rho$  is the intraclass correlation (common range: 0.00 – 0.50).

14

## Random Effects ANOVA Example

200 clusters of size 10 with intraclass correlation 0.2 analyzed as:

- TYPE = TWOLEVEL
- TYPE = COMPLEX
- Regular analysis, ignoring clustering

$$DEFF = 1 + 9 * 0.2 = 2.8$$

15

## Input For Two-Level Random Effects ANOVA Analysis

```
TITLE:      Random effects ANOVA data
            Two-level analysis with balanced data

DATA:      FILE = anova.dat;

VARIABLE:  NAMES = y cluster;
            USEV = y;
            CLUSTER = cluster;

ANALYSIS:  TYPE = TWOLEVEL;

MODEL:
            %WITHIN%
            y;
            %BETWEEN%
            y;
```

16

## Output Excerpts Two-Level Random Effects ANOVA Analysis

### Model Results

	Estimates	S.E.	Est./S.E.
Within Level			
Variances			
Y	0.779	0.025	31.293
Between Level			
Means			
Y	0.003	0.038	0.076
Variances			
Y	0.212	0.028	7.496

17

## Input For Complex Random Effects ANOVA Analysis

```
TITLE:      Random effects ANOVA data
           Complex analysis with balanced data

DATA:      FILE = anova.dat;

VARIABLE:  NAMES = y cluster;
           USEV = y;
           CLUSTER = cluster;

ANALYSIS:  TYPE = COMPLEX;
```

18

## Output Excerpts Complex Random Effects ANOVA Analysis

### Model Results

	Estimates	S.E.	Est./S.E.
Means			
Y	0.003	0.038	0.076
Variances			
Y	0.990	0.036	27.538

19

## Input For Random Effects ANOVA Analysis Ignoring Clustering

```
TITLE:      Random effects ANOVA data
            Ignoring clustering

DATA:      FILE = anova.dat;

VARIABLE:  NAMES = y cluster;
            USEV = y;
            CLUSTER = cluster;
!

ANALYSIS:  TYPE = MEANSTRUCTURE;
```

20

## Output Excerpts Random Effects ANOVA Analysis Ignoring Clustering

### Model Results

	Estimates	S.E.	Est./S.E.
Means			
Y	0.003	0.022	0.131
Variiances			
Y	0.990	0.031	31.623

Note: The estimated mean has SE = 0.022 instead of the correct 0.038

21

## Further Readings On Complex Survey Data

- Asparouhov, T. (2005). Sampling weights in latent variable modeling. Structural Equation Modeling, 12, 411-434.
- Chambers, R.L. & Skinner, C.J. (2003). Analysis of survey data. Chichester: John Wiley & Sons.
- Kaplan, D. & Ferguson, A.J (1999). On the utilization of sample weights in latent variable models. Structural Equation Modeling, 6, 305-321.
- Korn, E.L. & Graubard, B.I (1999). Analysis of health surveys. New York: John Wiley & Sons.
- Patterson, B.H., Dayton, C.M. & Graubard, B.I. (2002). Latent class analysis of complex sample survey data: application to dietary data. Journal of the American Statistical Association, 97, 721-741.
- Skinner, C.J., Holt, D. & Smith, T.M.F. (1989). Analysis of complex surveys. West Sussex, England: Wiley.
- Stapleton, L. (2002). The incorporation of sample weights into multilevel structural equation models. Structural Equation Modeling, 9, 475-502.

22

## Two-Level Regression Analysis

23

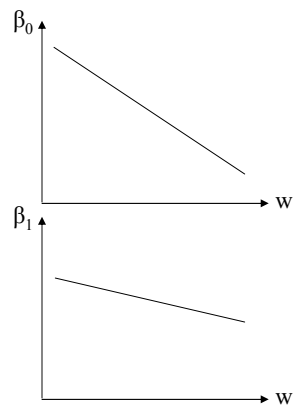
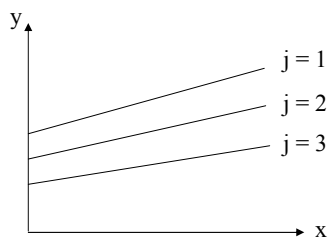
## Cluster-Specific Regressions

Individual  $i$  in cluster  $j$

$$(1) y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$$

$$(2a) \beta_{0j} = \gamma_{00} + \gamma_{01} w_j + u_{0j}$$

$$(2b) \beta_{1j} = \gamma_{10} + \gamma_{11} w_j + u_{1j}$$



24

## Two-Level Regression Analysis With Random Intercepts And Random Slopes In Multilevel Terms

Two-level analysis (individual  $i$  in cluster  $j$ ):

$y_{ij}$  : individual-level outcome variable  
 $x_{ij}$  : individual-level covariate  
 $w_j$  : cluster-level covariate

Random intercepts, random slopes:

$$\text{Level 1 (Within)} : y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}, \quad (1)$$

$$\text{Level 2 (Between)} : \beta_{0j} = \gamma_{00} + \gamma_{01} w_j + u_{0j}, \quad (2a)$$

$$\text{Level 2 (Between)} : \beta_{1j} = \gamma_{10} + \gamma_{11} w_j + u_{1j}. \quad (2b)$$

- Mplus gives the same estimates as HLM/MLwiN ML (not REML):
  - $V(r)$  (residual variance for level 1)
  - $\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, V(u_0), V(u_1), Cov(u_0, u_1)$
- Centering of  $x$ : subtracting grand mean or group (cluster) mean

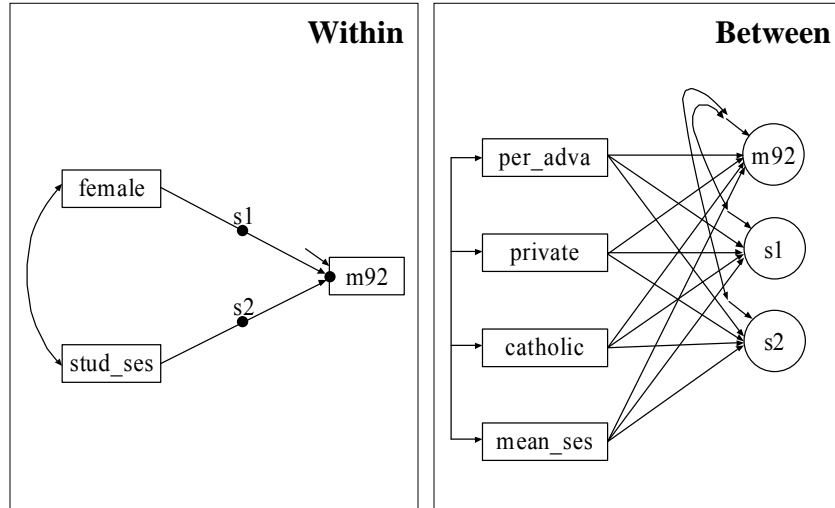
25

## NELS Data

- The data—National Education Longitudinal Study (NELS:88)
  - Base year Grade 8—followed up in Grades 10 and 12
  - Students sampled within 1,035 schools—approximately 26 students per school,  $n = 14,217$
  - Variables—reading, math, science, history-citizenship-geography, and background variables

26

## NELS Math Achievement Regression



27

## Input For NELS Math Achievement Regression

```

TITLE:      NELS math achievement regression

DATA:      FILE IS completev2.dat;
           ! National Education Longitudinal Study (NELS)
           FORMAT IS f8.0 12f5.2 f6.3 f11.4 23f8.2
           f18.2 f8.0 4f8.2;

VARIABLE:  NAMES ARE school r88 m88 s88 h88 r90 m90 s90 h90 r92
           m92 s92 h92 stud_ses f2pnlwt transfer minor coll_asp
           algebra retain aca_back female per_mino hw_time
           salary dis_fair clas_dis mean_col per_high unsafe
           num_frie teaqual par_invo ac_track urban size rural
           private mean_ses catholic stu_teach per_adva tea_exce
           tea_res;

           USEV = m92 female stud_ses per_adva private catholic
           mean_ses;

           !per_adva = percent teachers with an MA or higher

           WITHIN = female stud_ses;
           BETWEEN = per_adva private catholic mean_ses;
           MISSING = blank;
           CLUSTER = school;
           CENTERING = GRANDMEAN (stud_ses per_adva mean_ses);
    
```

28

## Input For NELS Math Achievement Regression (Continued)

```

ANALYSIS: TYPE = TWOLEVEL RANDOM MISSING;

MODEL:
    %WITHIN%
    s1 | m92 ON female;
    s2 | m92 ON stud_ses;

    %BETWEEN%
    m92 s1 s2 ON per_adva private catholic mean_ses;
    m92 WITH s1 s2;

OUTPUT: TECH8 SAMPSTAT;

```

29

## Output Excerpts NELS Math Achievement Regression

N = 10,933

### Summary of Data

Number of clusters      902

Size (s) Cluster ID with Size s

1	89863	75862	52654	1995	32661	89239	56214	
2	41743	81263	45025	26790	60281	82860	56241	21474
	4570	27159	11662	87842	38454			
3	65407	61407	83048	42640	41412	67708	83085	39685
	40402	93469	98582	68595	11517	17543	75498	81069
	66512							
4	31646	68153	85508	26234	83390	60835	74400	20770
	5095	10904	93569	38063	86733	66125	51670	10910
	98461	44395	95317	64112	50880	77381	12835	47555
	9208	93859	35719	67574	20048	34139	25784	80675
5	14464	74791	18219	10468	72193	97616	15773	877
	9471	83234	68254	68028	70718	3496	6842	45854

30

## Output Excerpts NELS Math Achievement Regression (Continued)

22	79570	15426	97947	93599	85125	10926	4603
23	6411	60328	70024	67835			
24	36988	22874	50626	19091			
25	56619	59710	34292	18826	62209		
26	44586	67832	16515				
27	82887						
28	847	76909					
30	36177						
31	12786	53660	47120	94802			
32	80553						
34	53272						
36	89842	31572					
42	99516						
43	75115						

Average cluster size 12.187  
 Estimated Intraclass Correlations for the Y Variables

	Intraclass	
Variable	Correlation	
M92	0.107	

31

## Output Excerpts NELS Math Achievement Regression (Continued)

### Tests of Model Fit

Loglikelihood		
H0 Value		-39390.404
Information Criteria		
Number of Free parameters		21
Akaike (AIC)		78822.808
Bayesian (BIC)		78976.213
Sample-Size Adjusted BIC		78909.478
	(n* = (n + 2) / 24)	

### Model Results

	Estimates	S.E.	Est./S.E.
<b>Within Level</b>			
Residual			
Variances			
M92	70.577	1.149	61.442
<b>Between Level</b>			
S1			
ON			
PER_ADVA	0.084	0.841	0.100
PRIVATE	-0.134	0.844	-0.159
CATHOLIC	-0.736	0.780	-0.944
MEAN_SES	-0.232	0.428	-0.542

32

## Output Excerpts NELS Math Achievement Regression (Continued)

		Estimates	S.E.	Est./S.E.
S2	ON			
	PER_ADVA	1.348	0.521	2.587
	PRIVATE	-1.890	0.706	-2.677
	CATHOLIC	-1.467	0.562	-2.612
	MEAN_SES	1.031	0.283	3.640
M92	ON			
	PER_ADVA	0.195	0.727	0.268
	PRIVATE	1.505	1.108	1.358
	CATHOLIC	0.765	0.650	1.178
	MEAN_SES	3.912	0.399	9.814
S1	WITH			
	M92	-4.456	1.007	-4.427
S2	WITH			
	M92	0.128	0.399	0.322
Intercepts				
	M92	55.136	0.185	297.248
	S1	-0.819	0.211	-3.876
	S2	4.841	0.152	31.900
Residual Variances				
	M92	8.679	1.003	8.649
	S1	5.740	1.411	4.066
	S2	0.307	0.527	0.583

33

## Cross-Level Influence

Between-level (level 2) variable  $w$  influencing within-level (level 1)  $y$  variable:

**Random intercept**

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + r_{ij}$$

$$\beta_{0j} = \underbrace{\gamma_{00} + \gamma_{01} w_j + u_{0j}}$$

Mplus:

```
MODEL:
  %WITHIN%;
  y ON x; ! estimates beta1
  %BETWEEN%;
  y ON w; ! y is the same as beta0
           ! estimates gamma01
```

34

## Cross-Level Influence (Continued)

Cross-level interaction, or between-level (level 2) variable moderating a within level (level 1) relationship:

### Random slope

$$y_{ij} = \beta_0 + \beta_{1j} x_{ij} + r_{ij}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} w_j + u_{1j}$$

Mplus:

MODEL:

```
%WITHIN%;
beta1 | y ON x;
%BETWEEN%;
beta1 ON w;           ! estimates gammall
```

35

## Random Slopes

- In single-level modeling random slopes  $\beta_i$  describe variation across individuals  $i$ ,

$$y_i = \alpha_i + \beta_i x_i + \varepsilon_i, \quad (100)$$

$$\alpha_i = \alpha + \zeta_{0i}, \quad (101)$$

$$\beta_i = \beta + \zeta_{1i}, \quad (102)$$

resulting in heteroscedastic residual variances

$$V(y_i | x_i) = V(\beta_i) x_i^2 + \theta. \quad (103)$$

- In two-level modeling random slopes  $\beta_j$  describe variation across clusters  $j$

$$y_{ij} = a_j + \beta_j x_{ij} + \varepsilon_{ij}, \quad (104)$$

$$a_j = a + \zeta_{0j}, \quad (105)$$

$$\beta_j = \beta + \zeta_{1j}, \quad (106)$$

A small variance for a random slope typically leads to slow convergence of the ML-EM iterations. This suggests respecifying the slope as fixed.

Mplus allows random slopes for predictors that are

- Observed covariates
- Observed dependent variables
- Continuous latent variables

36

## **Further Readings On Multilevel Regression Analysis**

- Ludtke Marsh, Robitzsch, Trautwein, Asparouhov, Muthen (2007). Analysis of group level effects using multilevel modeling: Probing a latent covariate approach. Submitted for publication.
- Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. Second edition. Newbury Park, CA: Sage Publications.
- Snijders, T. & Bosker, R. (1999). Multilevel analysis. An introduction to basic and advanced multilevel modeling. Thousand Oakes, CA: Sage Publications.

37

## **Logistic And Probit Regression**

38

## Categorical Outcomes: Logit And Probit Regression

Probability varies as a function of x variables (here  $x_1, x_2$ )

$$P(u = 1 | x_1, x_2) = F[\beta_0 + \beta_1 x_1 + \beta_2 x_2], \quad (22)$$

$P(u = 0 | x_1, x_2) = 1 - P[u = 1 | x_1, x_2]$ , where  $F[z]$  is either the standard normal ( $\Phi[z]$ ) or logistic ( $1/[1 + e^{-z}]$ ) distribution function.

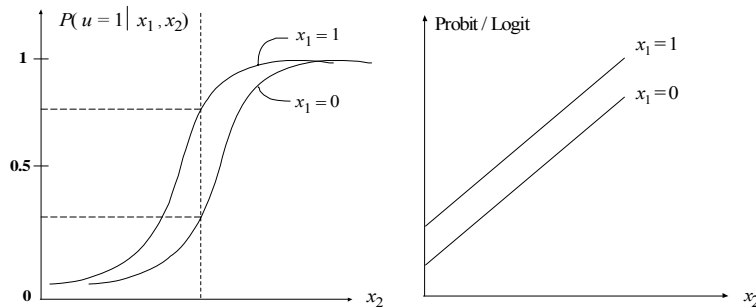
**Example:** Lung cancer and smoking among coal miners

- $u$  lung cancer ( $u = 1$ ) or not ( $u = 0$ )
- $x_1$  smoker ( $x_1 = 1$ ), non-smoker ( $x_1 = 0$ )
- $x_2$  years spent in coal mine

39

## Categorical Outcomes: Logit And Probit Regression

$$P(u = 1 | x_1, x_2) = F[\beta_0 + \beta_1 x_1 + \beta_2 x_2], \quad (22)$$



40

## Interpreting Logit And Probit Coefficients

- Sign and significance
- Odds and odds ratios
- Probabilities

41

## Logistic Regression And Log Odds

$$\begin{aligned} \text{Odds}(u = 1 | x) &= P(u = 1 | x) / P(u = 0 | x) \\ &= P(u = 1 | x) / (1 - P(u = 1 | x)). \end{aligned}$$

The logistic function

$$P(u = 1 | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

gives a log odds linear in  $x$ ,

$$\text{logit} = \log [\text{odds}(u = 1 | x)] = \log [P(u = 1 | x) / (1 - P(u = 1 | x))]$$

$$= \log \left[ \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} / \left( 1 - \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \right) \right]$$

$$= \log \left[ \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} * \frac{1 + e^{-(\beta_0 + \beta_1 x)}}{e^{-(\beta_0 + \beta_1 x)}} \right]$$

$$= \log \left[ e^{(\beta_0 + \beta_1 x)} \right] = \beta_0 + \beta_1 x$$

42

## Logistic Regression And Log Odds (Continued)

- $\text{logit} = \log \text{odds} = \beta_0 + \beta_1 x$
- When  $x$  changes one unit, the *logit* (*log odds*) changes  $\beta_1$  units
- When  $x$  changes one unit, the *odds* changes  $e^{\beta_1}$  units

43

## Two-Level Logistic Regression

With  $j$  denoting cluster,

$$\text{logit}_{ij} = \log (P(u_{ij} = 1)/P(u_{ij} = 0)) = \alpha_j + \beta_j * x_{ij}$$

where

$$\begin{aligned}\alpha_j &= \alpha + u_{0j} \\ \beta_j &= \beta + u_{1j}\end{aligned}$$

High/low  $\alpha_j$  value means high/low logit (high log odds)

44

## Predicting Juvenile Delinquency From First Grade Aggressive Behavior

- Cohort 1 data from the Johns Hopkins University Preventive Intervention Research Center
- n= 1,084 students in 40 classrooms, Fall first grade
- Covariates: gender and teacher-rated aggressive behavior

45

## Input For Two-Level Logistic Regression

```
TITLE: Hopkins Cohort 1 2-level logistic regression
DATA: FILE = Cohort1_classroom_ALL.DAT;
VARIABLE:
      NAMES = prcid juv99 gender stub1F bkRule1F harm01F
              bkThin1F yell1F takeP1F fight1F lies1F
              tease1F;
      CLUSTER = classrm;
      USEVAR = juv99 male aggress;
      CATEGORICAL = juv99;
      MISSING = ALL (999);
      WITHIN = male aggress;
DEFINE:
      male = 2 - gender;
      aggress = stub1F + bkRule1F + harm01F + bkThin1F +
                yell1F + takeP1F + fight1F + lies1F + tease1F;
```

46

## Input For Two-Level Logistic Regression (Continued)

```
ANALYSIS:
      TYPE = TWOLEVEL MISSING;
      PROCESS = 2;
MODEL:
      %WITHIN%
      juv99 ON male aggress;
      %BETWEEN%
OUTPUT:
      TECH1 TECH8;
```

47

## Output Excerpts Two-Level Logistic Regression

```
MODEL RESULTS
```

	Estimates	S.E	Est./S.E.
Within Level			
JUV99			
ON			
MALE	1.071	0.149	7.193
AGGRESS	0.060	0.010	6.191
Between Level			
Thresholds			
JUV99\$1	2.981	0.205	14.562
Variances			
JUV99	0.807	0.250	3.228

48

## Understanding The Between-Level Intercept Variance

- Intra-class correlation
  - $ICC = 0.807 / (\pi^2/3 + 0.807)$
- Odds ratios
  - Larsen & Merlo (2005). Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. *American Journal of Epidemiology*, 161, 81-88.
  - Larsen proposes MOR:  
"Consider two persons with the same covariates, chosen randomly from two different clusters. The MOR is the median odds ratio between the person of higher propensity and the person of lower propensity."  
$$MOR = \exp(\sqrt{2 * \sigma^2} * \Phi^{-1}(0.75))$$
  
In the current example,  $ICC = 0.20$ ,  $MOR = 2.36$
- Probabilities
  - Compare  $\alpha_j = 1$  SD and  $\alpha_k = -1$  SD from the mean

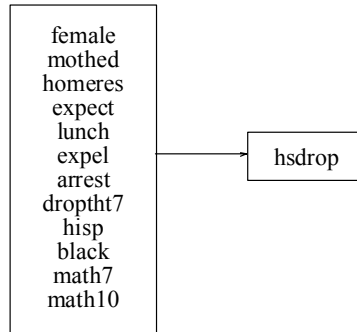
49

## Two-Level Path Analysis

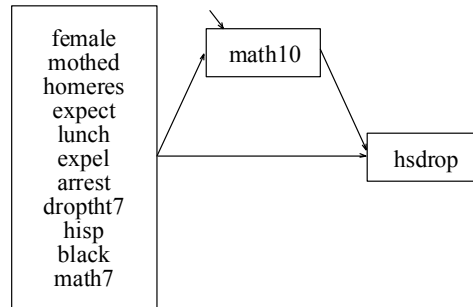
50

## A Path Model With A Binary Outcome And A Mediator With Missing Data

### Logistic Regression



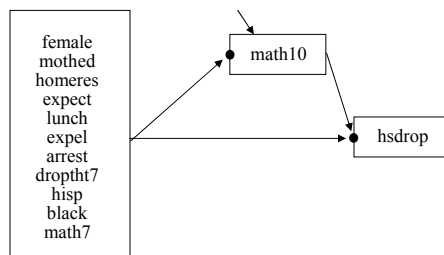
### Path Model



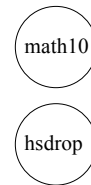
51

## Two-Level Path Analysis

### Within



### Between



52

## Input For A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable

```
TITLE:      a twolevel path analysis with a categorical outcome
            and missing data on the mediating variable
DATA:      FILE = lsayfull_dropout.dat;
VARIABLE:  NAMES = female mothed homeres math7 math10 expel
            arrest hisp black hsdrop expect lunch droptht7
            schcode;
            MISSING = ALL (9999);
            CATEGORICAL = hsdrop;
            CLUSTER = schcode;
            WITHIN = female mothed homeres expect math7 lunch
            expel arrest droptht7 hisp black;
ANALYSIS:  TYPE = TWOLEVEL MISSING;
            ESTIMATOR = ML;
            ALGORITHM = INTEGRATION;
            INTEGRATION = MONTECARLO (500);
```

53

## Input For A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable (Continued)

```
MODEL:
            %WITHIN%
            hsdrop ON female mothed homeres expect math7 math10
            lunch expel arrest droptht7 hisp black;
            math10 ON female mothed homeres expect math7 lunch
            expel arrest droptht7 hisp black;
            %BETWEEN%
            hsdrop*1; math10*1;
OUTPUT:    PATTERNS SAMPSTAT STANDARDIZED TECH1 TECH8;
```

54

**Output Excerpts A Two-Level Path Analysis Model  
With A Categorical Outcome And Missing Data  
On The Mediating Variable**

**Summary Of Data**

	Number of patterns	2		
	Number of clusters	44		
Size (s)	Cluster ID with Size s			
12	304			
13	305			
36	307	122		
38	106	112		
39	138	109		
40	103			
41	308			
42	146	120		
43	102	101		
44	303	143		
45	141			

55

**Output Excerpts A Two-Level Path Analysis Model  
With A Categorical Outcome And Missing Data  
On The Mediating Variable (Continued)**

Size (s)	Cluster ID with Size s					
46	144					
47	140					
49	108					
50	126	111	110			
51	127	124				
52	137	117	147	118	301	136
53	142	131				
55	145	123				
57	135	105				
58	121					
59	119					
73	104					
89	302					
93	309					
118	115					

56

**Output Excerpts A Two-Level Path Analysis Model  
With A Categorical Outcome And Missing Data  
On The Mediating Variable (Continued)**

**Model Results**

	Estimates	S.E.	Est./S.E.	Std	StdYX
Within Level					
HSDROP ON					
FEMALE	0.323	0.171	1.887	0.323	0.077
MOTHEd	-0.253	0.103	-2.457	-0.253	-0.121
HOMERES	-0.077	0.055	-1.401	-0.077	-0.061
EXPECT	-0.244	0.065	-3.756	-0.244	-0.159
MATH7	-0.011	0.015	-0.754	-0.011	-0.055
MATH10	-0.031	0.011	-2.706	-0.031	-0.197
LUNCH	0.008	0.006	1.324	0.008	0.074
EXPEL	0.947	0.225	4.201	0.947	0.121
ARREST	0.068	0.321	0.212	0.068	0.007
DROPTHT7	0.757	0.284	2.665	0.757	0.074
HISP	-0.118	0.274	-0.431	-0.118	-0.016
BLACK	-0.086	0.253	-0.340	-0.086	-0.013

57

**Output Excerpts A Two-Level Path Analysis Model  
With A Categorical Outcome And Missing Data  
On The Mediating Variable (Continued)**

	Estimates	S.E.	Est./S.E.	Std	StdYX
MATH10 ON					
FEMALE	-0.841	0.398	-2.110	-0.841	-0.031
MOTHEd	0.263	0.215	1.222	0.263	0.020
HOMERES	0.568	0.136	4.169	0.568	0.070
EXPECT	0.985	0.162	6.091	0.985	0.100
MATH7	0.940	0.023	40.123	0.940	0.697
LUNCH	-0.039	0.017	-2.308	-0.039	-0.059
EXPEL	-1.293	0.825	-1.567	-1.293	-0.026
ARREST	-3.426	1.022	-3.353	-3.426	-0.054
DROPTHT7	-1.424	1.049	-1.358	-1.424	-0.022
HISP	-0.501	0.728	-0.689	-0.501	-0.010
BLACK	-0.369	0.733	-0.503	-0.369	-0.009

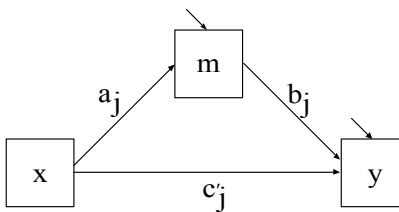
58

## Output Excerpts A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
Residual Variances					
MATH10	62.010	2.162	28.683	62.010	0.341
Between Level					
Means					
MATH10	10.226	1.340	7.632	10.226	5.276
Thresholds					
HSDROP\$1	-1.076	0.560	-1.920		
Variances					
HSDROP	0.286	0.133	2.150	0.286	1.000
MATH10	3.757	1.248	3.011	3.757	1.000

59

## Two-Level Mediation



Indirect effect:

$$\alpha + \beta + Cov(a_j, b_j)$$

Bauer, Preacher & Gil (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 11, 142-163.

60

## Input For Two-Level Mediation

```
MONTECARLO:
  NAMES ARE y m x;
  WITHIN = x;
  NOOBSERVATIONS = 1000;
  NCSIZES = 1;
  CSIZES = 100 (10);
  NREP = 100;

MODEL POPULATION:
  %WITHIN%
  c | y ON x;
  b | y ON m;
  a | m ON x;
  x*1; m*1; y*1;
  %BETWEEN%
  y WITH m*0.1 b*0.1 a*0.1 c*0.1;
  m WITH b*0.1 a*0.1 c*0.1;
  a WITH b*0.1 c*0.1;
  b WITH c*0.1;
  y*1 m*1 a*1 b*1 c*1;
  [a*0.4 b*0.5 c*0.6];
```

61

## Input For Two-Level Mediation (Continued)

```
ANALYSIS:
  TYPE = TWOLEVEL RANDOM;

MODEL:
  %WITHIN%
  c | y ON x;
  b | y ON m;
  a | m ON x;
  m*1; y*1;
  %BETWEEN%
  y WITH M*0.1 b*0.1 a*0.1 c*0.1;
  m WITH b*0.1 a*0.1 c*0.1;
  a WITH b*0.1 (cab);
  a WITH c*0.1;
  b WITH c*0.1;
  y*1 m*1 a*1 b*1 c*1;
  [a*0.4] (ma);
  [b*0.5] (mb);
  [c*0.6];

MODEL CONSTRAINT:
  NEW(m*0.3);
  m=ma*mb+cab;
```

62

## Output Excerpts Two Level Mediation

	Estimates			S.E.	M. S. E.	95%	% Sig
	Population	Average	Std.Dev.	Average		Cover	Coeff
<b>Within Level</b>							
Residual variances							
Y	1.000	1.0020	0.0530	0.0530	0.0028	0.960	1.000
M	1.000	1.0011	0.0538	0.0496	0.0029	0.910	1.000
<b>Between Level</b>							
Y	WITH						
B	0.100	0.1212	0.1246	0.114	0.0158	0.910	0.210
A	0.100	0.1086	0.1318	0.1162	0.0173	0.910	0.190
C	0.100	0.0868	0.1121	0.1237	0.0126	0.940	0.090
M	WITH						
B	0.100	0.1033	0.1029	0.1085	0.0105	0.940	0.120
A	0.100	0.0815	0.1081	0.1116	0.0119	0.950	0.070
C	0.100	0.1138	0.1147	0.1165	0.0132	0.970	0.160
A	WITH						
B	0.100	0.0964	0.1174	0.1101	0.0137	0.920	0.150
C	0.100	0.0756	0.1376	0.1312	0.0193	0.910	0.110

63

## Output Excerpts Two-Level Mediation (Continued)

B	WITH						
C	0.100	0.0892	0.1056	0.1156	0.0112	0.960	0.070
Y	WITH						
M	0.100	0.1034	0.1342	0.1285	0.0178	0.940	0.140
<b>Means</b>							
Y	0.000	0.0070	0.1151	0.1113	0.0132	0.950	0.050
M	0.000	-0.0031	0.1102	0.1056	0.0120	0.950	0.050
C	0.600	0.5979	0.1229	0.1125	0.0150	0.930	1.000
B	0.500	0.5022	0.1279	0.1061	0.0162	0.890	1.000
A	0.400	0.3854	0.0972	0.1072	0.0096	0.970	0.970
<b>Variiances</b>							
Y	1.000	1.0071	0.1681	0.1689	0.0280	0.910	1.000
M	1.000	1.0113	0.1782	0.1571	0.0316	0.930	1.000
C	1.000	0.9802	0.1413	0.1718	0.0201	0.980	1.000
B	1.000	0.9768	0.1443	0.1545	0.0212	0.950	1.000
A	1.000	1.0188	0.1541	0.1587	0.0239	0.950	1.000
<b>New/Additional Parameters</b>							
M	0.300	0.2904	0.1422	0.1316	0.0201	0.950	0.550

64

## Two-Level Factor Analysis

65

## Two-Level Factor Analysis

- Recall random effects ANOVA (individual  $i$  in cluster  $j$ ):

$$y_{ij} = \nu + \eta_j + \varepsilon_{ij} = y_{B_j} + y_{W_{ij}}$$

- Two-level factor analysis ( $r = 1, 2, \dots, p$  items):

$$y_{rij} = \nu_r + \lambda_{B_r} \eta_{B_j} + \varepsilon_{B_{rj}} + \lambda_{W_r} \eta_{W_{ij}} + \varepsilon_{W_{rij}}$$

(between-cluster variation)      (within-cluster variation)

66

## Two-Level Factor Analysis (Continued)

- Covariance structure:

$$V(\mathbf{y}) = V(\mathbf{y}_B) + V(\mathbf{y}_w) = \Sigma_B + \Sigma_w,$$

$$\Sigma_B = \mathbf{A}_B \Psi_B \mathbf{A}_B' + \Theta_B,$$

$$\Sigma_w = \mathbf{A}_w \Psi_w \mathbf{A}_w' + \Theta_w.$$

- Two interpretations:
  - variance decomposition, including decomposing the residual
  - random intercept model

67

## Two-Level Factor Analysis And Design Effects

Muthén & Satorra (1995; Sociological Methodology): Monte Carlo study using two-level data (200 clusters of varying size and varying intraclass correlations), a latent variable model with 10 variables, 2 factors, conventional ML using the regular sample covariance matrix  $S_T$ , and 1,000 replications (d.f. = 34).

$$\mathbf{A}_B = \mathbf{A}_w = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \Psi_B, \Theta_B \text{ reflecting different icc's}$$

$$y_{ij} = \nu + \lambda(\eta_{Bj} + \eta_{w_{ij}}) + \varepsilon_{Bj} + \varepsilon_{w_{ij}}$$

$$V(\mathbf{y}) = \Sigma_B + \Sigma_w = \lambda(\Psi_B + \Psi_w) \lambda' + \Theta_B + \Theta_w$$

68

## Two-Level Factor Analysis And Design Effects (Continued)

### Inflation of $\chi^2$ due to clustering

Intraclass Correlation		Cluster Size			
		7	15	30	60
0.05	Chi-square mean	35	36	38	41
	Chi-square var	68	72	80	96
	5%	5.6	7.6	10.6	20.4
	1%	1.4	1.6	2.8	7.7
0.10	Chi-square mean	36	40	46	58
	Chi-square var	75	89	117	189
	5%	8.5	16.0	37.6	73.6
	1%	1.0	5.2	17.6	52.1
0.20	Chi-square mean	42	52	73	114
	Chi-square var	100	152	302	734
	5%	23.5	57.7	93.1	99.9
	1%	8.6	35.0	83.1	99.4

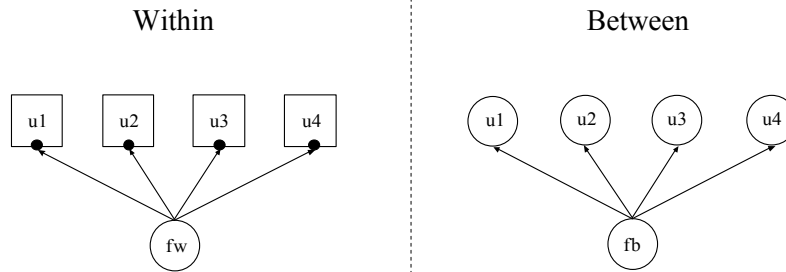
69

## Two-Level Factor Analysis And Design Effects (Continued)

- Regular analysis, ignoring clustering
  - Inflated chi-square, underestimated SE's
- TYPE = COMPLEX
  - Correct chi-square and SE's but only if model aggregates,  
e.g.  $A_B = A_W$
- TYPE = TWOLEVEL
  - Correct chi-square and SE's

70

## Two-Level Factor Analysis (IRT)



$$u^*_{ij} = \lambda (f_{B_j} + f_{w_{ij}}) + \varepsilon_{ij}$$

71

## Input For A Two-Level Factor Analysis (IRT) Model With Categorical Outcomes

```
TITLE:      this is an example of a two-level factor analysis
            model with categorical outcomes
DATA:      FILE = catrepl.dat;
VARIABLE:  NAMES ARE u1-u6 clus;
            CATEGORICAL = u1-u6;
            CLUSTER = clus;
ANALYSIS:  TYPE = TWOLEVEL;
            ESTIMATION = ML;
            ALGORITHM = INTEGRATION;
MODEL:
            %WITHIN%
            fw BY u1@1
            u2 (1)
            u3 (2)
            u4 (3)
            u5 (4)
            u6 (5);
```

72

## Input For A Two-Level Factor Analysis (IRT) Model With Categorical Outcomes (Continued)

```
%BETWEEN%  
fb BY u1@1  
u2 (1)  
u3 (2)  
u4 (3)  
u5 (4)  
u6 (5);  
OUTPUT: TECH1 TECH8;
```

73

## Output Excerpts A Two-Level Factor Analysis (IRT) Model With Categorical Outcomes

### Tests Of Model Fit

Loglikelihood

H0 Value -3696.117

Information Criteria

Number of Free Parameters 13

Akaike (AIC) 7418.235

Bayesian (BIC) 7481.505

Sample-Size Adjusted BIC 7440.217

(n\* = (n + 2) / 24)

74

**Output Excerpts A Two-Level Factor Analysis  
(IRT) Model With Categorical Outcomes  
(Continued)**

**Model Results**

Within Level		Estimates	S.E.	Est./S.E.
FW	BY			
	U1	1.000	0.000	0.000
	U2	0.915	0.146	6.264
	U3	1.087	0.169	6.437
	U4	1.058	0.164	6.441
	U5	1.191	0.185	6.449
	U6	1.143	0.178	6.439
Variances				
	FW	0.834	0.191	4.360

75

**Output Excerpts Two-Level Factor Analysis  
(IRT) Model With Categorical Outcomes (Continued)**

Between Level		Estimates	S.E.	Est./S.E.
FB	BY			
	U1	1.000	0.000	0.000
	U2	0.915	0.146	6.264
	U3	1.087	0.169	6.437
	U4	1.058	0.164	6.441
	U5	1.191	0.185	6.449
	U6	1.143	0.178	6.439
Thresholds				
	U1\$1	-0.206	0.096	-2.150
	U2\$1	0.001	0.091	0.007
	U3\$1	-0.016	0.100	-0.156
	U4\$1	-0.064	0.098	-0.652
	U5\$1	-0.033	0.105	-0.315
	U6\$1	-0.021	0.102	-0.209
Variances				
	FB	0.496	0.139	3.562

76

## **SIMS Variance Decomposition**

The Second International Mathematics Study (SIMS; Muthén, 1991, JEM).

- National probability sample of school districts selected proportional to size; a probability sample of schools selected proportional to size within school district, and two classes randomly drawn within each school
- 3,724 students observed in 197 classes from 113 schools with class sizes varying from 2 to 38; typical class size of around 20
- Eight variables corresponding to various areas of eighth-grade mathematics
- Same set of items administered as a pretest in the Fall of eighth grade and as a posttest in the Spring.

77

## **SIMS Variance Decomposition (Continued)**

Muthén (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28, 338-354.

- Research questions: “The substantive questions of interest in this article are the variance decomposition of the subscores with respect to within-class student variation and between-class variation and the change of this decomposition from pretest to posttest. In the SIMS ... such variance decomposition relates to the effects of tracking and differential curricula in eighth-grade math. On the one hand, one may hypothesize that effects of selection and instruction tend to increase between-class variation relative to within-class variation, assuming that the classes are homogeneous, have different performance levels to begin with, and show faster growth for higher initial performance level. On the other hand, one may hypothesize that eighth-grade exposure to new topics will increase individual differences among students within each class so that posttest within-class variation will be sizable relative to posttest between-class variation.”

78

## SIMS Variance Decomposition (Continued)

$$y_{rij} = \nu_r + \lambda_{Br} \eta_{Bj} + \varepsilon_{Brj} + \lambda_{wr} \eta_{wij} + \varepsilon_{wrij}$$

$$V(y_{rij}) = \text{BF} + \text{BE} + \text{WF} + \text{WE}$$

Between reliability:  $\text{BF} / (\text{BF} + \text{BE})$

– BE often small (can be fixed at 0)

Within reliability:  $\text{WF} / (\text{WF} + \text{WE})$

– sum of a small number of items gives a large WE

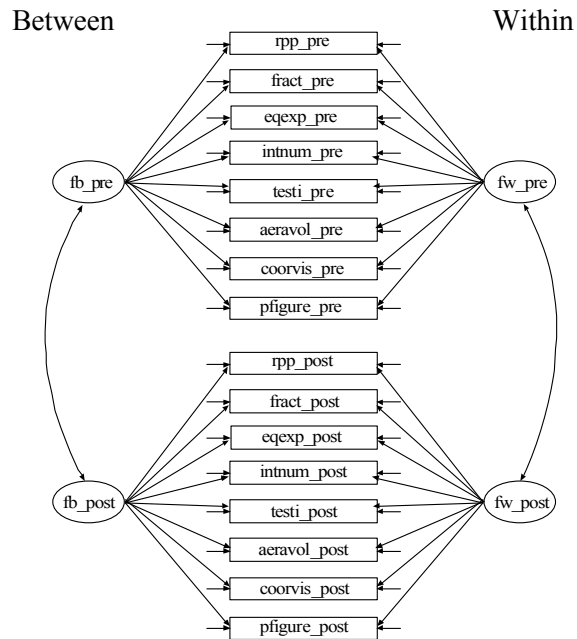
Intraclass correlation:

$$\text{ICC} = (\text{BF} + \text{BE}) / (\text{BF} + \text{BE} + \text{WF} + \text{WE})$$

Large measurement error  $\rightarrow$  large WE  $\rightarrow$  small ICC

$$\text{True ICC} = \text{BF} / (\text{BF} + \text{WF})$$

79



80

**Table 4: Variance Decomposition of SIMS Achievement Scores  
(percentages of total variance in parenthesis)**

		ANOVA							FACTOR ANALYSIS				
		Pretest			Posttest			% Increase In Variance		Error-free Prop. Between		Error-free % Increase In Variance	
		Between	Within	Prop-Between	Between	Within	Prop-Between	Between	Within	Pre	Post	Between	Within
RPP	8	1.542 (34.0)	2.990 (66.0)	.34	2.084 (38.5)	3.326 (61.5)	.38	35	11	.54	.52	29	41
FRACT	8	1.460 (38.2)	2.366 (61.8)	.38	1.906 (40.8)	2.767 (59.2)	.41	31	17	.60	.58	29	41
EQEXP	6	.543 (26.9)	1.473 (73.1)	.27	1.041 (38.7)	1.646 (61.3)	.39	92	18	.65	.64	113	117
INTNUM	2	.127 (25.2)	.358 (70.9)	.29	.195 (30.6)	.442 (69.4)	.31	54	24	.63	.61	29	41
TESTI	5	.580 (33.3)	1.163 (66.7)	.33	.664 (34.5)	1.258 (65.5)	.34	15	8	.58	.56	29	41
AREAVOL	2	.094 (17.2)	.451 (82.8)	.17	.156 (24.1)	.490 (75.9)	.24	66	9	.54	.52	29	41
COORVIS	3	.173 (20.9)	.656 (79.1)	.21	.275 (28.7)	.680 (68.3)	.32	59	4	.57	.55	29	41
PFigure	5	.363 (22.9)	1.224 (77.1)	.23	.711 (42.9)	1.451 (67.1)	.33	96	19	.60	.54	87	136

81

**Second-Generation JHU PIRC Trial Aggression Items**

Item Distributions for Cohort 3: Fall 1st Grade (n=362 males in 27 classrooms)

	<i>Almost Never (scored as 1)</i>	<i>Rarely (scored as 2)</i>	<i>Sometimes (scored as 3)</i>	<i>Often (scored as 4)</i>	<i>Very Often (scored as 5)</i>	<i>Almost Always (scored as 6)</i>
<b>Stubborn</b>	42.5	21.3	18.5	7.2	6.4	4.1
<b>Breaks Rules</b>	37.6	16.0	22.7	7.5	8.3	8.0
<b>Harms Others</b>	69.3	12.4	9.40	3.9	2.5	2.5
<b>Breaks Things</b>	79.8	6.60	5.20	3.9	3.6	0.8
<b>Yells at Others</b>	61.9	14.1	11.9	5.8	4.1	2.2
<b>Takes Others' Property</b>	72.9	9.70	10.8	2.5	2.2	1.9
<b>Fights</b>	60.5	13.8	13.5	5.5	3.0	3.6
<b>Harms Property</b>	74.9	9.90	9.10	2.8	2.8	0.6
<b>Lies</b>	72.4	12.4	8.00	2.8	3.3	1.1
<b>Talks Back to Adults</b>	79.6	9.70	7.80	1.4	0.8	1.4
<b>Teases Classmates</b>	55.0	14.4	17.7	7.2	4.4	1.4
<b>Fights With Classmates</b>	67.4	12.4	10.2	5.0	3.3	1.7
<b>Loses Temper</b>	61.6	15.5	13.8	4.7	3.0	1.4

82

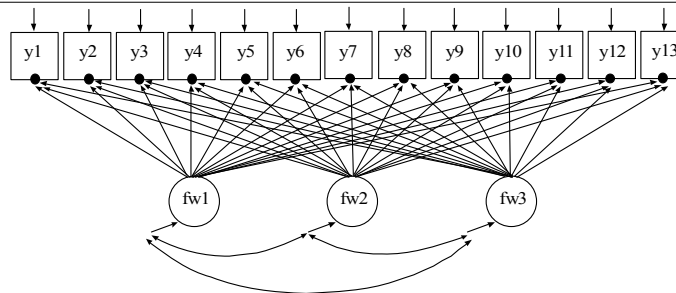
## Hypothesized Aggressiveness Factors

- Verbal aggression
  - Yells at others
  - Talks back to adults
  - Loses temper
  - Stubborn
- Property aggression
  - Breaks things
  - Harms property
  - Takes others' property
  - Harms others
- Person aggression
  - Fights
  - Fights with classmates
  - Teases classmates

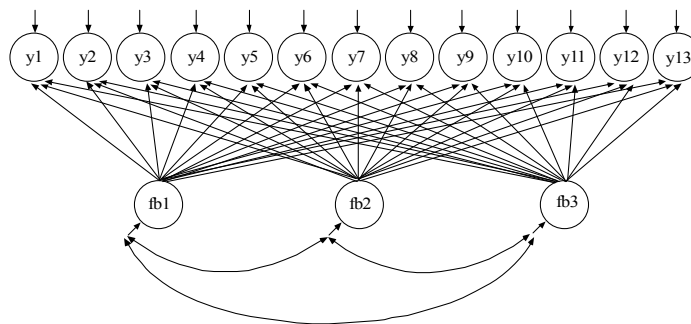
83

## Two-Level Factor Analysis

Within



Between



84

## Promax Rotated Loadings

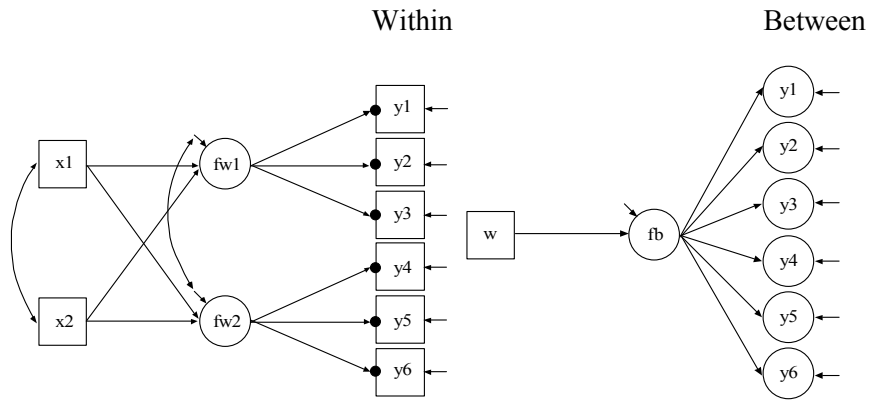
	Within-Level Loadings			Between-Level Loadings		
	1	2	3	1	2	3
<b>Stubborn</b>	0.07	<b>0.70</b>	0.05	-0.19	<b>1.03</b>	0.07
<b>Breaks Rules</b>	0.25	0.31	0.37	0.15	0.28	0.31
<b>Harms Others</b>	<b>0.52</b>	0.16	0.27	0.35	-0.20	<b>0.72</b>
<b>Breaks Things</b>	<b>0.84</b>	0.16	-0.01	<b>0.71</b>	0.01	0.41
<b>Yells at Others</b>	0.15	<b>0.64</b>	0.13	0.38	<b>0.74</b>	-0.01
<b>Takes Others' Property</b>	<b>0.57</b>	0.00	0.37	<b>0.86</b>	-0.04	0.12
<b>Fights</b>	0.20	0.21	<b>0.63</b>	0.09	0.03	<b>0.89</b>
<b>Harms Property</b>	<b>0.73</b>	0.21	0.10	<b>0.90</b>	-0.05	0.16
<b>Lies</b>	0.48	0.28	0.24	<b>0.86</b>	0.33	-0.21
<b>Talks Back to Adults</b>	0.29	<b>0.71</b>	0.23	0.41	0.58	-0.04
<b>Teases Classmates</b>	0.11	0.19	<b>0.62</b>	0.37	0.31	0.30
<b>Fights With Classmates</b>	0.10	0.31	<b>0.63</b>	-0.19	0.38	<b>0.88</b>
<b>Loses Temper</b>	0.12	<b>0.75</b>	0.04	0.17	<b>0.78</b>	0.12

85

## Two-Level Factor Analysis With Covariates

86

## Two-Level Factor Analysis With Covariates



87

## Input For Two-Level Factor Analysis With Covariates

```

TITLE:      this is an example of a two-level CFA with
             continuous factor indicators with two factors on the
             within level and one factor on the between level

DATA:       FILE IS ex9.8.dat;

VARIABLE:   NAMES ARE y1-y6 x1 x2 w clus;
             WITHIN = x1 x2;
             BETWEEN = w;
             CLUSTER IS clus;

ANALYSIS:   TYPE IS TWOLEVEL;

MODEL:      %WITHIN%
             fw1 BY y1-y3;
             fw2 BY y4-y6;
             fw1 ON x1 x2;
             fw2 ON x1 x2;
             %BETWEEN%
             fb BY y1-y6;
             fb ON w;
    
```

88

## Input For Monte Carlo Simulations For Two-Level Factor Analysis With Covariates

```
TITLE:          This is an example of a two-level CFA with
                  continuous factor indicators with two
                  factors on the within level and one factor
                  on the between level

MONTECARLO:
  NAMES ARE y1-y6 x1 x2 w;
  NOOBSERVATIONS = 1000;
  NCSIZES = 3;
  CSIZES = 40 (5) 50 (10) 20 (15);
  SEED = 58459;
  NREPS = 1;
  SAVE = ex9.8.dat;
  WITHIN = x1 x2;
  BETWEEN = w;

ANALYSIS:      TYPE = TWOLEVEL;
```

89

## Input For Monte Carlo Simulations For Two-Level Factor Analysis With Covariates (Continued)

```
MODEL POPULATION:

  %WITHIN%
  x1-x2@1;
  fw1 BY y1@1 y2-y3*1;
  fw2 BY y4@1 y5-y6*1;
  fw1-fw2*1;
  y1-y6*1;
  fw1 ON x1*.5 x2*.7;
  fw2 ON x1*.7 x2*.5;

  %BETWEEN%
  [w@0]; w*1;
  fb BY y1@1 y2-y6*1;
  y1-y6*.3;
  fb*.5;
  fb ON w*1;
```

90

## **Input For Monte Carlo Simulations For Two-Level Factor Analysis With Covariates (Continued)**

```
MODEL:
    %WITHIN%
    fw1 BY y1@1 y2-y3*1;
    fw2 BY y4@1 y5-y6*1;
    fw1-fw2*1;
    y1-y6*1;
    fw1 ON x1*.5 x2*.7;
    fw2 ON x1*.7 x2*.5;

    %BETWEEN%
    fb BY y1@1 y2-y6*1;
    y1-y6*.3;
    fb*.5;
    fb ON w*1;

OUTPUT:
    TECH8 TECH9;
```

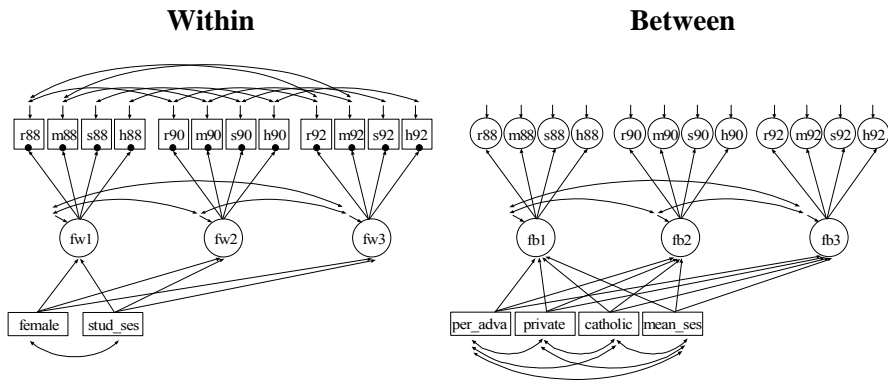
91

## **NELS Data**

- The data—National Education Longitudinal Study (NELS:88)
  - Base year Grade 8—followed up in Grades 10 and 12
  - Students sampled within 1,035 schools—approximately 26 students per school,  $n = 14,217$
  - Variables—reading, math, science, history-citizenship-geography, and background variables
- Data for the analysis—reading, math, science, history-citizenship-geography

92

## NELS Two-Level Longitudinal Factor Analysis With Covariates



93

## Input For NELS Two-Level Longitudinal Factor Analysis With Covariates

```

TITLE:      two-level factor analysis with covariates using the NELS
            data

DATA:      FILE = NELS.dat;
            FORMAT = 2f7.0 f11.4 12f5.2 11f8.2;

VARIABLE:  NAMES = id school f2pnlwt r88 m88 s88 h88 r90 m90 s90 h90
            r92 m92 s92 h92 stud_ses female per_mino urban size rural
            private mean_ses catholic stu_tec per_adva;
            !Variable Description
            !m88 = math IRT score in 1988
            !m90 = math IRT score in 1990
            !m92 = math IRT score in 1992
            !r88 = reading IRT score in 1988
            !r90 = reading IRT score in 1990
            !r92 = reading IRT score in 1992
    
```

94

## Input For NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

```
!s88 = science IRT score in 1988
!s90 = science IRT score in 1990
!s92 = science IRT score in 1992
!h88 = history IRT score in 1988
!h90 = history IRT score in 1990
!h92 = history IRT score in 1992
!female = scored 1 vs 0
!stud_ses = student family ses in 1990 (flses)
!per_adva = percent teachers with an MA or higher
!private = private school (scored 1 vs 0)
!catholic = catholic school (scored 1 vs 0)
!private = 0, catholic = 0 implies public school

MISSING = BLANK;
CLUSTER = school;

USEV = r88 m88 s88 h88 r90 m90 s90 h90 r92 m92 s92 h92
female stud_ses per_adva private catholic mean_ses;
WITHIN = female stud_ses;
BETWEEN = per_adva private catholic mean_ses;
```

95

## Input For NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

```
ANALYSIS: TYPE = TWOLEVEL MISSING;
MODEL: %WITHIN%
fw1 BY r88-h88;
fw2 BY r90-h90;
fw3 BY r92-h92;
r88 WITH r90; r90 WITH r92; r88 WITH r92;
m88 WITH m90; m90 WITH m92; m88 WITH m92;
s88 WITH s90; s90 WITH s92;
h88 WITH h90; h90 WITH h92;
fw1-fw3 ON female stud_ses;

%BETWEEN%
fb1 BY r88-h88;
fb2 BY r90-h90;
fb3 BY r92-h92;
fb1-fb3 ON per_adva private catholic mean_ses;
OUTPUT: SAMPSTAT STANDARDIZED TECH1 TECH8 MODINDICES;
```

96

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates

### Summary Of Data

Number of patterns      15  
 Number of clusters     913

Average cluster size 15.572

Estimated Intraclass Correlations for the Y Variables

Variable	Intraclass Correlation	Variable	Intraclass Correlation	Variable	Intraclass Correlation
R88	0.067	M88	0.129	S88	0.100
H88	0.105	R90	0.076	M90	0.117
S90	0.110	H90	0.106	R92	0.073
M92	0.111	S92	0.099	H92	0.091

97

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

### Tests Of Model Fit

Chi-Square Test of Model Fit

Value	4883.539*
Degrees of Freedom	146
P-Value	0.0000
Scaling Correction Factor for MLR	1.046

Chi-Square Test of Model Fit for the Baseline Model

Value	150256.855
Degrees of Freedom	202
P-Value	0.0000

CFI/TLI

CFI	0.968
TLI	0.956

Loglikelihood

H0 Value	-487323.777	
H1 Value	-484770.257	98

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

### Information Criteria

Number of Free Parameters	94
Akaike (AIC)	974835.554
Bayesian (BIC)	975546.400
Sample-Size Adjusted BIC	975247.676
(n* = (n + 2) / 24)	
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.048
SRMR (Standardized Root Mean Square Residual)	
Value for Between	0.041
Value for Within	0.027

99

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

### Model Results

		Estimates	S.E.	Est./S.E.	Std	StdYX
Within Level						
FW1	BY					
R88		1.000	0.000	0.000	6.528	0.812
M88		0.940	0.010	94.856	6.135	0.804
S88		1.005	0.010	95.778	6.559	0.837
H88		1.041	0.011	97.888	6.796	0.837
FW2	BY					
R90		1.000	0.000	0.000	8.038	0.842
M90		0.911	0.008	109.676	7.321	0.838
S90		1.003	0.010	99.042	8.065	0.859
H90		0.939	0.008	113.603	7.544	0.855

100

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

FW3	BY					
R92		1.000	0.000	0.000	8.460	0.832
M92		0.939	0.009	101.473	7.946	0.845
S92		1.003	0.011	90.276	8.482	0.861
H92		0.934	0.009	102.825	7.905	0.858
FW1	ON					
FEMALE		-0.403	0.128	-3.150	-0.062	-0.031
STUD_SES		3.378	0.096	35.264	0.517	0.418
FW2	ON					
FEMALE		-0.621	0.157	-3.945	-0.077	-0.039
STUD_SES		4.169	0.110	37.746	0.519	0.420
FW3	ON					
FEMALE		-1.027	0.169	-6.087	-0.121	-0.064
STUD_SES		4.418	0.122	36.124	0.522	0.422

101

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

Residual Variances						
R88		22.021	0.383	57.464	22.021	0.341
M88		20.618	0.338	61.009	20.618	0.354
S88		18.383	0.323	56.939	18.383	0.299
H88		19.805	0.370	53.587	19.805	0.300
R90		26.546	0.491	54.033	26.546	0.291
M90		22.756	0.375	60.748	22.756	0.298
S90		23.150	0.383	60.516	23.150	0.262
H90		21.002	0.403	52.124	21.002	0.270
R92		31.821	0.617	51.562	31.821	0.308
M92		25.213	0.485	52.018	25.213	0.285
S92		25.155	0.524	47.974	25.155	0.259
H92		22.479	0.489	46.016	22.479	0.265
FW1		35.081	0.699	50.201	0.823	0.823
FW2		53.079	1.005	52.806	0.822	0.822
FW3		58.438	1.242	47.041	0.817	0.817

102

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

Between Level						
FB1	BY					
R88		1.000	0.000	0.000	1.952	0.933
M88		1.553	0.070	22.138	3.031	0.979
S88		1.061	0.058	18.255	2.071	0.887
H88		1.065	0.053	19.988	2.078	0.814
FB2	BY					
R90		1.000	0.000	0.000	2.413	0.923
M90		1.407	0.058	24.407	3.395	1.003
S90		1.220	0.062	19.697	2.943	0.946
H90		0.973	0.047	20.496	2.348	0.829
FB3	BY					
R92		1.000	0.000	0.000	2.472	0.947
M92		1.435	0.065	22.095	3.546	0.997
S92		1.160	0.065	17.889	2.868	0.938
H92		0.963	0.041	23.244	2.380	0.871

103

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

Between Level						
FB1	ON					
PER_ADVA		0.217	0.292	0.742	0.111	0.024
PRIVATE		0.303	0.344	0.883	0.155	0.042
CATHOLIC		-0.696	0.277	-2.512	-0.357	-0.088
MEAN_SES		2.513	0.206	12.185	1.288	0.672
FB2	ON					
PER_ADVA		0.280	0.338	0.828	0.116	0.025
PRIVATE		0.453	0.392	1.155	0.188	0.051
CATHOLIC		-0.538	0.334	-1.609	-0.223	-0.055
MEAN_SES		3.054	0.239	12.805	1.266	0.660
FB3	ON					
PER_ADVA		0.473	0.375	1.261	0.192	0.041
PRIVATE		0.673	0.435	1.547	0.272	0.074
CATHOLIC		-0.206	0.372	-0.554	-0.084	-0.021
MEAN_SES		3.142	0.258	12.169	1.271	0.663

104

## Output Excerpts NELS Two-Level Longitudinal Factor Analysis With Covariates (Continued)

### Residual Variances

R88	0.564	0.104	5.437	0.564	0.129
M88	0.399	0.093	4.292	0.399	0.042
S88	1.160	0.126	9.170	1.160	0.213
H88	2.203	0.203	10.839	2.203	0.338
R90	1.017	0.160	6.352	1.017	0.149
M90	-0.068	0.055	-1.225	-0.068	-0.006
S90	1.025	0.172	5.945	1.025	0.106
H90	2.518	0.216	11.636	2.518	0.313
R92	0.706	0.182	3.886	0.706	0.104
M92	0.076	0.076	1.000	0.076	0.006
S92	1.120	0.190	5.901	1.120	0.120
H92	1.810	0.211	8.599	1.810	0.242
FB1	1.979	0.245	8.066	0.520	0.520
FB2	3.061	0.345	8.875	0.526	0.526
FB3	3.010	0.409	7.363	0.493	0.493

105

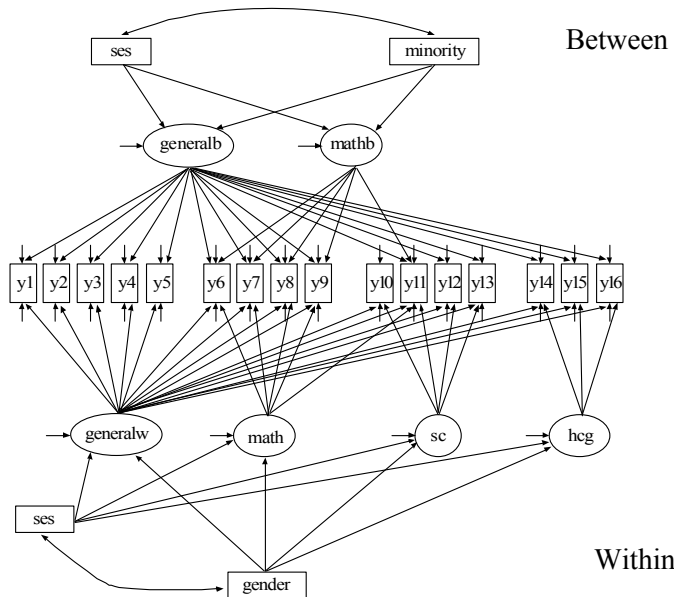
## Multiple-Group, Two-Level Factor Analysis With Covariates

106

## NELS Data

- The data—National Education Longitudinal Study (NELS:88)
  - Base year Grade 8—followed up in Grades 10 and 12
  - Students sampled within 1,035 schools—approximately 26 students per school
  - Variables—reading, math, science, history-citizenship-geography, and background variables
- Data for the analysis—reading, math, science, history-citizenship-geography, gender, individual SES, school SES, and minority status, n = 14,217 with 913 schools (clusters)

107



108

## Input For NELS:88 Two-Group, Two-Level Model For Public And Catholic Schools

```
TITLE:      NELS:88 with listwise deletion
           disaggregated model for two groups, public and
           catholic schools

DATA:      FILE IS EX831.DAT;;

VARIABLE:  NAMES = ses y1-y16 gender cluster minority group;
           CLUSTER = cluster;
           WITHIN = gender;
           BETWEEN = minority;
           GROUPING = group(1=public 2=catholic);

DEFINE:    minority = minority/5;

ANALYSIS:  TYPE = TWOLEVEL;
           H1ITER = 2500;
           MITER = 1000;
```

109

## Input For NELS:88 Two-Group, Two-Level Model For Public And Catholic Schools (Continued)

```
MODEL:     %WITHIN%
           generalw BY y1* y2-y6 y8-y16 y7@1;
           mathw BY y6* y8* y9* y11 y7@1;
           scw BY y10 y11*.5 y12*.3 y13*.2;
           hcgw BY y14*.7 y16*2 y15@1;
           generalw WITH mathw-hcgw@0;
           mathw WITH scw-hcgw@0;
           scw WITH hcgw@0;
           generalw mathw scw hcgw ON gender ses;

           %BETWEEN%
           generalb BY y1* y2-y6 y8-y16 y7@1;
           mathb BY y6* y8 y9 y11 y7@1;
           y1-y16@0;
           generalb WITH mathb@0;
           generalb mathb ON ses minority;
```

110

## Output Excerpts NELS:88 Two-Group, Two-Level Model For Public And Catholic Schools

### Summary Of Data

```

Group PUBLIC
Number of clusters      195
Size (s) Cluster ID with Size s
 1      68114  68519
 2      72872
 7      72765
 8      45991  72012
 9      68071
10      7298  72187
11      72463  7105  72405
12      24083  68971  7737  68390
13      45861  72219  72049
14      68511  72148  72175  72176  25464
15      68023  25071  68748  45928  7915  78324
16      45362  7403  72415  77204  77219  72456
17      45502  68487  45824  7203  24948  7829  72612  7892
      25835  7591  68155  68295
  
```

111

## Output Excerpts NELS:88 Two-Group, Two-Level Model For Public And Catholic Schools (Continued)

```

18      72133  25580  24910  68614  25074  72990  68328  25404
      7348
19      7671  68662  68671  45385  7438  7332  25615  72799
      68340  72956  25642  25658  24856  78283  68030
20      72617  72715  7211  25422  7330  72292  72060  72993
      7451  68461  78162  78232  72170  25130
21      45394  7193  68180  24589  7205  25894  25958  68391
      77254  77634  68448  45271  7584  25227  78598
22      68254  68397  68648  72768  7192  7117  7119  68753
      24813
23      68456  25361  7157  25702  25804  45620  24858  7658
      25163  45041  77351  45183  77684  78101  68788  68817
      7792  78311  68048  68453
24      77222  24053  7000  77403  24138  68297  78011  25536
      7778  72042  25360  25977  45747  7616  78886
25      68906  68720  25354  68427  72833  77268  7269  68520
      77537  72075
26      72973  45555  24828  68315  45087  25328  77710  25848
27      45831  25618  68652  72080  45900  25208  45452  7103
  
```

112

**Output Excerpts NELS:88 Two-Group, Two-Level  
Model For Public And Catholic Schools (Continued)**

28	25666	68809	25076	25224	68551
30	7343	45978	25722	45924	
31	77109	7230	68855		
32	25178				
33	45330	25745	25825		
35	25667				
36	72129				
37	25834				
38	45287				
39	45197	7090			
43	45366				

113

**Output Excerpts NELS:88 Two-Group, Two-Level  
Model For Public And Catholic Schools (Continued)**

Group PUBLIC

Number of clusters 195  
Average cluster size 21.292

Estimated Intraclass Correlations for the Y Variables

Variable	Intraclass Correlation	Variable	Intraclass Correlation	Variable	Intraclass Correlation
Y1	.111	Y7	.100	Y12	.115
Y2	.105	Y8	.124	Y13	.185
Y3	.213	Y9	.069	Y14	.094
Y4	.160	Y10	.147	Y15	.132
Y5	.081	Y11	.105	Y16	.159
Y6	.159				

114

## Output Excerpts NELS:88 Two-Group, Two-Level Model For Public And Catholic Schools (Continued)

Group CATHOLIC

Number of clusters 40  
Average cluster size 26.016

Estimated Intraclass Correlations for the Y Variables

Variable	Intraclass Correlation	Variable	Intraclass Correlation	Variable	Intraclass Correlation
Y1	.010	Y7	.029	Y12	.056
Y2	.039	Y8	.061	Y13	.176
Y3	.180	Y9	.056	Y14	.078
Y4	.091	Y10	.079	Y15	.071
Y5	.055	Y11	.056	Y16	.154
Y6	.118				

115

## Output Excerpts NELS:88 Two-Group, Two-Level Model For Public And Catholic Schools (Continued)

### Tests Of Model Fit

Loglikelihood

Value	1716.922*
Degrees of Freedom	575
P-Value	0.0000
Scaling Correction Factor for MLR	0.872

Chi-Square Test of Model

Value	35476.471
Degrees of Freedom	608
P-Value	0.0000

CFI/TLI

CFI	0.967
TLI	0.965

Loglikelihood

H0 Value	-130332.921
H1 Value	-129584.053

116

**Output Excerpts NELS:88 Two-Group, Two-Level  
Model For Public And Catholic Schools (Continued)**

		Estimates	S.E.	Est./S.E.	Std	StdYX
<b>Group Public Within Level</b>						
GENERALW	ON					
	GENDER	-0.193	0.029	-6.559	-0.256	-0.128
	SES	0.233	0.016	14.269	0.309	0.279
MATHW	ON					
	GENDER	0.266	0.025	10.534	0.510	0.255
	SES	0.054	0.014	3.879	0.103	0.093
SCW	ON					
	GENDER	0.452	0.032	14.005	0.961	0.480
	SES	0.018	0.015	1.244	0.039	0.035
HCGW	ON					
	GENDER	0.152	0.023	6.588	0.681	0.341
	SES	0.002	0.007	0.239	0.007	0.007

117

**Output Excerpts NELS:88 Two-Group, Two-Level  
Model For Public And Catholic Schools (Continued)**

		Estimates	S.E.	Est./S.E.	Std	StdYX
<b>Group Catholic Within Level</b>						
GENERALW	ON					
	GENDER	-0.294	0.059	-5.000	-0.403	-0.201
	SES	0.169	0.021	7.892	0.232	0.193
MATHW	ON					
	GENDER	0.332	0.051	6.478	0.627	0.313
	SES	-0.030	0.017	-1.707	-0.056	-0.047
SCW	ON					
	GENDER	0.555	0.063	8.860	1.226	0.613
	SES	-0.022	0.014	-1.592	-0.049	-0.041
HCGW	ON					
	GENDER	0.160	0.029	5.610	0.785	0.392
	SES	0.001	0.007	0.089	0.003	0.002

118

**Output Excerpts NELS:88 Two-Group, Two-Level  
Model For Public And Catholic Schools (Continued)**

	Estimates	S.E.	Est./S.E.	Std	StdYX
<b>Group Public Between Level</b>					
GENERALB ON					
SES	0.505	0.079	6.390	1.244	0.726
MINORITY	-0.217	0.088	-2.452	-0.534	-0.188
MATHB ON					
SES	0.198	0.070	2.825	0.984	0.574
MINORITY	-0.031	0.087	-0.354	-0.153	-0.054
GENERALB WITH MATHB	0.000	0.000	0.000	0.000	0.000
Intercepts					
GENERALB	0.000	0.000	0.000	0.000	0.000
MATHB	0.000	0.000	0.000	0.000	0.000

119

**Output Excerpts NELS:88 Two-Group, Two-Level  
Model For Public And Catholic Schools (Continued)**

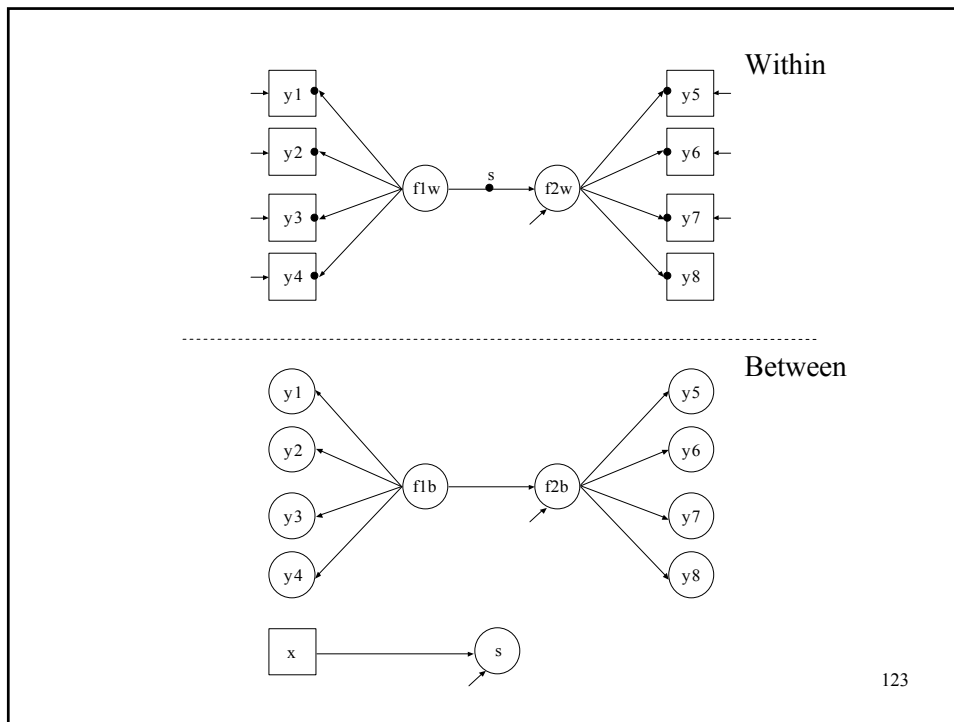
	Estimates	S.E.	Est./S.E.	Std	StdYX
<b>Group Catholic Between Level</b>					
GENERALB ON					
SES	0.262	0.067	3.929	0.975	0.538
MINORITY	-0.327	0.069	-4.707	-0.216	-0.573
MATHB ON					
SES	0.205	0.071	2.901	0.746	0.412
MINORITY	-0.213	0.095	-2.241	-0.778	-0.367
GENERALB WITH MATHB	0.000	0.000	0.000	0.000	0.000
Intercepts					
GENERALB	0.466	0.163	2.854	1.734	1.734
MATHB	0.573	0.177	3.239	2.087	2.087

120

## **Further Readings On Two-Level Factor Analysis**

- Harnqvist, K., Gustafsson, J.E., Muthén, B., & Nelson, G. (1994). Hierarchical models of ability at class and individual levels. Intelligence, 18, 165-187. (#53)
- Hox, J. (2002). Multilevel analysis. Techniques and applications. Mahwah, NJ: Lawrence Erlbaum
- Longford, N. T., & Muthén, B. (1992). Factor analysis for clustered observations. Psychometrika, 57, 581-597. (#41)
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585. (#24)
- Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society meeting in Princeton, NJ, June 1990. UCLA Statistics Series 62. (#32)
- Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. Journal of Educational Measurement, 28, 338-354. (#37)
- Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox & I. Kreft (eds.), Multilevel Modeling, a special issue of Sociological Methods & Research, 22, 376-398. (#55) 121

## **Two-Level SEM: Random Slopes For Regressions Among Factors**



## Input For A Two-Level SEM With A Random Slope

```

TITLE:      a twolevel SEM with a random slope

DATA:      FILE = etaeta3.dat;

VARIABLE:  NAMES ARE y1-y8 x clus;
           CLUSTER = clus;
           BETWEEN = x;

ANALYSIS:  TYPE = TWOLEVEL RANDOM MISSING;
           ALGORITHM = INTEGRATION;

```

124

## Input For A Two-Level SEM With A Random Slope (Continued)

```
MODEL:      %WITHIN%
            flw BY y1@1
            y2 (1)
            y3 (2)
            y4 (3);
            f2w BY y5@1
            y6 (4)
            y7 (5)
            y8 (6);
            s | f2w ON flw;

            %BETWEEN%
            flb BY y1@1
            y2 (1)
            y3 (2)
            y4 (3);
            f2b BY y5@1
            y6 (4)
            y7 (5)
            y8 (6);
            f2b ON flb;
            s ON x;

OUTPUT:     TECH1 TECH8;
```

125

## Output Excerpts Two-Level SEM With A Random Slope

### Tests Of Model Fit

Loglikelihood

H0 Value	-12689.557
----------	------------

Information Criteria

Number of Free Parameters	30
Akaike (AIC)	25439.114
Bayesian (BIC)	25585.122
Sample-Size Adjusted BIC ( $n^* = (n + 2) / 24$ )	25489.843

126

## Output Excerpts Two-Level SEM With A Random Slope (Continued)

### Model Results

		Estimates	S.E.	Est./S.E.
Within Level				
F1W	BY			
	Y1	1.000	0.000	0.000
	Y2	0.992	0.035	28.597
	Y3	0.978	0.041	23.593
	Y4	1.001	0.037	26.884
F2W	BY			
	Y5	1.000	0.000	0.000
	Y6	0.978	0.028	34.417
	Y7	1.049	0.030	35.174
	Y8	1.008	0.026	38.090
F1W	WITH			
	F2W	0.000	0.000	0.000

127

## Output Excerpts Two-Level SEM With A Random Slope (Continued)

		Estimates	S.E.	Est./S.E.
Variances				
F1W		1.016	0.082	12.325
F2W		0.580	0.063	9.144
Residual Variances				
Y1		0.979	0.063	15.517
Y2		0.949	0.056	16.854
Y3		1.052	0.060	17.406
Y4		0.971	0.053	18.174
Y5		1.039	0.057	18.187
Y6		1.062	0.058	18.292
Y7		0.941	0.058	16.191
Y8		1.076	0.060	17.835

128

## Output Excerpts Two-Level SEM With A Random Slope (Continued)

		Estimates	S.E.	Est./S.E.
Between Level				
F1B	BY			
	Y1	1.000	0.000	0.000
	Y2	0.992	0.035	28.597
	Y3	0.978	0.041	23.593
	Y4	1.001	0.037	26.884
F2B	BY			
	Y5	1.000	0.000	0.000
	Y6	0.978	0.028	34.417
	Y7	1.049	0.030	35.174
	Y8	1.008	0.026	38.090
F2B	ON			
	F1B	0.180	0.080	2.248

129

## Output Excerpts Two-Level SEM With A Random Slope (Continued)

		Estimates	S.E.	Est./S.E.
S	ON			
	X	0.999	0.082	12.150
Intercepts				
	Y1	-0.099	0.063	-1.560
	Y2	-0.011	0.064	-0.175
	Y3	-0.069	0.067	-1.034
	Y4	-0.001	0.065	-0.017
	Y5	0.030	0.062	0.475
	Y6	-0.008	0.064	-0.129
	Y7	0.041	0.064	0.635
	Y8	0.002	0.071	0.035
	S	0.777	0.073	10.604
Variances				
	F1B	0.568	0.096	5.900
Residual Variances				
	F2B	0.237	0.056	4.211
	S	0.420	0.088	4.756

130

## Multilevel Estimation, Testing, Modification, And Identification

### Estimators

- Muthén's limited information estimator (MUML) – random intercepts
  - ESTIMATOR = MUML
  - Muthén's limited information estimator for unbalanced data
  - Maximum likelihood for balanced data
- Full-information maximum likelihood (FIML) – random intercepts and random slopes
  - ESTIMATOR = ML, **MLR**, MLF
  - Full-information maximum likelihood for balanced and unbalanced data
  - Robust maximum likelihood estimator
  - MAR missing data
  - Asparouhov and Muthén

131

## Multilevel Estimation, Testing, Modification, And Identification (Continued)

### Tests of Model Fit

- MUML – chi-square, robust chi-square, CFI, TLI, RMSEA, and SRMR
- FIML – chi-square, robust chi-square, CFI, TLI, RMSEA, and SRMR
- FIML with random slopes – no tests of model fit

### Model Modification

- MUML – modification indices not available
- FIML – modification indices available

**Model identification is the same as for CFA for both the between and within parts of the model.**

132

## **Practical Issues Related To The Analysis Of Multilevel Data**

### **Size Of The Intraclass Correlation**

- Small intraclass correlations can be ignored but important information about between-level variability may be missed by conventional analysis
- The importance of the size of an intraclass correlation depends on the size of the clusters
- Intraclass correlations are attenuated by individual-level measurement error
- Effects of clustering not always seen in intraclass correlations

133

## **Practical Issues Related To The Analysis Of Multilevel Data (Continued)**

### **Within-Level And Between-Level Variables**

- Variables measured on the individual level can be used in both the between and within parts of the model or only in the within part of the model (WITHIN=)
- Variables measured on the between level can be used only in the between part of the model (BETWEEN=)

### **Sample Size**

- There should be at least 30-50 between-level units (clusters)
- Clusters with only one observation are allowed

134

## **Steps In SEM Multilevel Analysis For Continuous Outcomes**

- 1) Explore SEM model using the sample covariance matrix from the total sample
- 2) Estimate the SEM model using the pooled-within sample covariance matrix with sample size  $n - G$
- 3) Investigate the size of the intraclass correlations and DEFF's
- 4) Explore the between structure using the estimated between covariance matrix with sample size  $G$
- 5) Estimate and modify the two-level model suggested by the previous steps

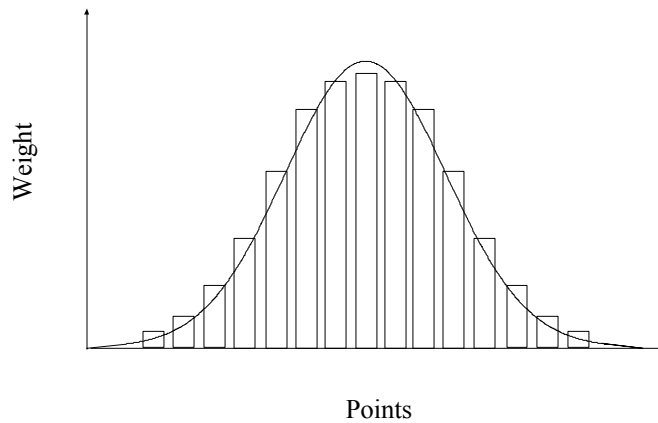
Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox & I. Kreft (eds.), *Multilevel Modeling*, a special issue of *Sociological Methods & Research*, 22, 376-398. (#55)

135

## **Technical Aspects Of Multilevel Modeling**

136

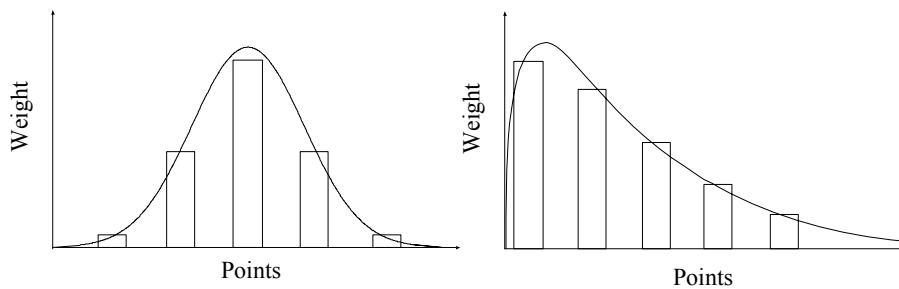
## Numerical Integration With A Normal Latent Variable Distribution



Fixed weights and points

137

## Non-Parametric Estimation Of The Random Effect Distribution



Estimated weights and points  
(class probabilities and class means)

138

## Numerical Integration

Numerical integration is needed with maximum likelihood estimation when the posterior distribution for the latent variables does not have a closed form expression. This occurs for models with categorical outcomes that are influenced by continuous latent variables, for models with interactions involving continuous latent variables, and for certain models with random slopes such as multilevel mixture models.

When the posterior distribution does not have a closed form, it is necessary to integrate over the density of the latent variables multiplied by the conditional distribution of the outcomes given the latent variables. Numerical integration approximates this integration by using a weighted sum over a set of integration points (quadrature nodes) representing values of the latent variable.

139

## Numerical Integration (Continued)

Numerical integration is computationally heavy and thereby time-consuming because the integration must be done at each iteration, both when computing the function value and when computing the derivative values. The computational burden increases as a function of the number of integration points, increases linearly as a function of the number of observations, and increases exponentially as a function of the dimension of integration, that is, the number of latent variables for which numerical integration is needed.

140

## Practical Aspects Of Numerical Integration

- Types of numerical integration available in Mplus with or without adaptive quadrature
  - Standard (rectangular, trapezoid) – default with 15 integration points per dimension
  - Gauss-Hermite
  - Monte Carlo
- Computational burden for latent variables that need numerical integration
  - One or two latent variables      Light
  - Three to five latent variables      Heavy
  - Over five latent variables      Very heavy

141

## Practical Aspects Of Numerical Integration (Continued)

- Suggestions for using numerical integration
  - Start with a model with a small number of random effects and add more one at a time
  - Start with an analysis with TECH8 and MITERATIONS=1 to obtain information from the screen printing on the dimensions of integration and the time required for one iteration and with TECH1 to check model specifications
  - With more than 3 dimensions, reduce the number of integration points to 5 or 10 or use Monte Carlo integration with the default of 500 integration points
  - If the TECH8 output shows large negative values in the column labeled ABS CHANGE, increase the number of integration points to improve the precision of the numerical integration and resolve convergence problems

142

## Technical Aspects Of Numerical Integration

Maximum likelihood estimation using the EM algorithm computes in each iteration the posterior distribution for normally distributed latent variables  $f$ ,

$$[f|y] = [f][y|f] / [y], \quad (97)$$

where the marginal density for  $[y]$  is expressed by integration

$$[y] = \int [f][y|f] df. \quad (98)$$

- Numerical integration is not needed for normally distributed  $y$  - the posterior distribution is normal

143

## Technical Aspects Of Numerical Integration (Continued)

- Numerical integration needed for:
  - Categorical outcomes  $u$  influenced by continuous latent variables  $f$ , because  $[u]$  has no closed form
  - Latent variable interactions  $f \times x, f \times y, f_1 \times f_2$ , where  $[y]$  has no closed form, for example

$$[y] = \int [f_1, f_2][y|f_1, f_2, f_1 f_2] df_1 df_2 \quad (99)$$

- Random slopes, e.g. with two-level mixture modeling

Numerical integration approximates the integral by a sum

$$[y] = \int [f][y|f] df = \sum_{k=1}^K w_k [y|f_k] \quad (100)$$

144

## **Multivariate Approach To Multilevel Modeling**

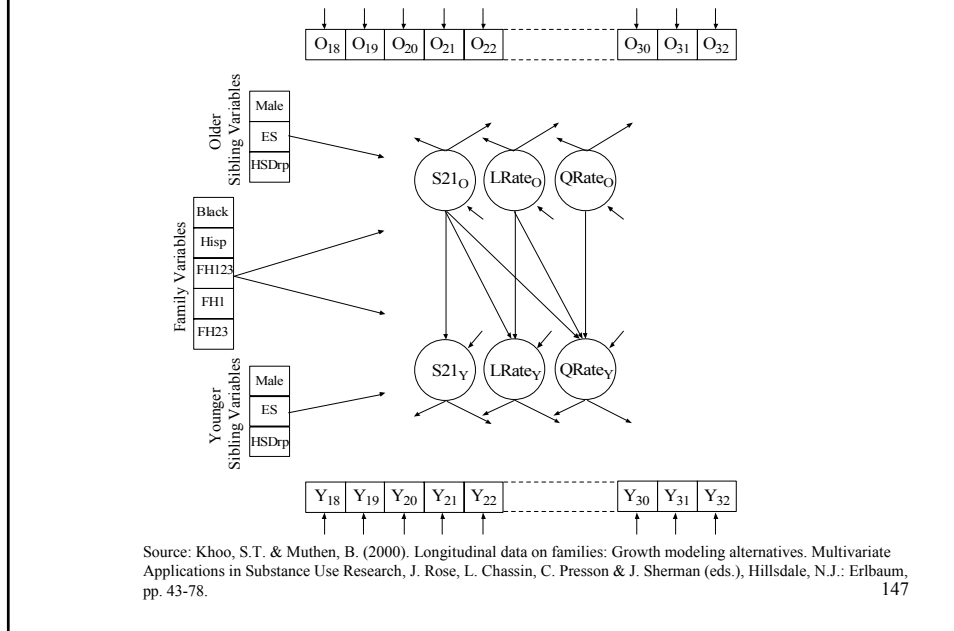
145

## **Multivariate Modeling Of Family Members**

- Multilevel modeling: clusters independent, model for between- and within-cluster variation, units within a cluster statistically equivalent
- Multivariate approach: clusters independent, model for all variables for each cluster unit, different parameters for different cluster units.
  - Used in latent variable growth modeling where the cluster units are the repeated measures over time
  - Allows for different cluster sizes by missing data techniques
  - More flexible than the multilevel approach, but computationally convenient only for applications with small cluster sizes (e.g. twins, spouses)

146

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



### Three-Level Modeling As Single-Level Analysis

Doubly multivariate:

- Repeated measures in wide, multivariate form
- Siblings in wide, multivariate form

It is possible to do four-level by TYPE = TWOLEVEL, for instance families within geographical segments

## Input For Multivariate Modeling Of Family Data

```
TITLE:      Multivariate modeling of family data
            one observation per family

DATA:      FILE IS multi.dat;

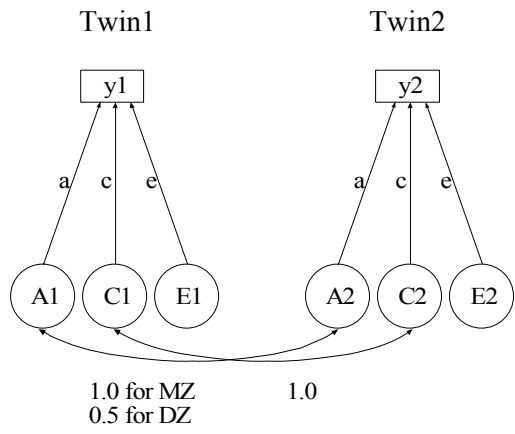
VARIABLE:  NAMES ARE o18-o32 y18-y32 omale oes ohdrop ymale yoes
            yhsdrop black hisp fh123 fh1 fh23;

MODEL:    s21o lrateo grateo | o18@0 o19@1 o20@2 o21@3 o22@4
            o23@5 o24@6 o25@7 o26@8 o27@9 o28@10 o29@11 o30@12
            o31@13 o32@14;
            s21y lratey gratey | y18@0 y19@1 y20@2 y21@3 y22@4 y23@5
            y24@6 y25@7 y26@8 y27@9 y28@10 y29@11 y30@12
            y31@13 y32@14;
            s12o ON omale oes ohdrop black hisp fh123 fh1 fh23;
            221y ON ymale yes yhsdrop black hisp fh123 fh1 fh23;
            s21y ON s21o;
            lratey ON s21o lrateo;
            gratey ON s21o lrateo grateo;
```

149

## Twin Modeling

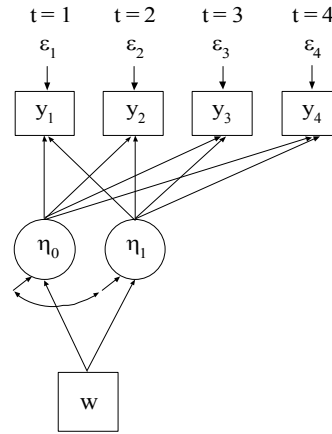
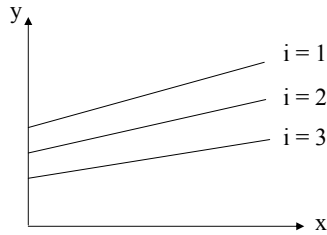
150



Neale & Cardon (1992)  
Prescott (2004)

**Multilevel Growth Models**

## Individual Development Over Time



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

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

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

153

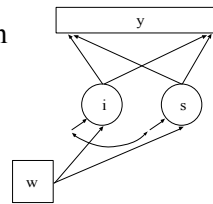
## Growth Modeling Approached In Two Ways: Data Arranged As Wide Versus Long

- Wide: Multivariate, Single-Level Approach

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

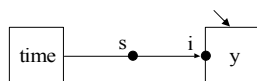
$i_i$  regressed on  $w_i$

$s_i$  regressed on  $w_i$

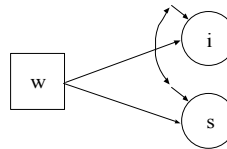


- Long: Univariate, 2-Level Approach (CLUSTER = id)

Within



Between



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

154

## 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

155

## Three-Level Modeling In Multilevel Terms

Time point  $t$ , individual  $i$ , cluster  $j$ .

$y_{ij}$  : individual-level, outcome variable  
 $a_{1ij}$  : individual-level, time-related variable (age, grade)  
 $a_{2ij}$  : 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)

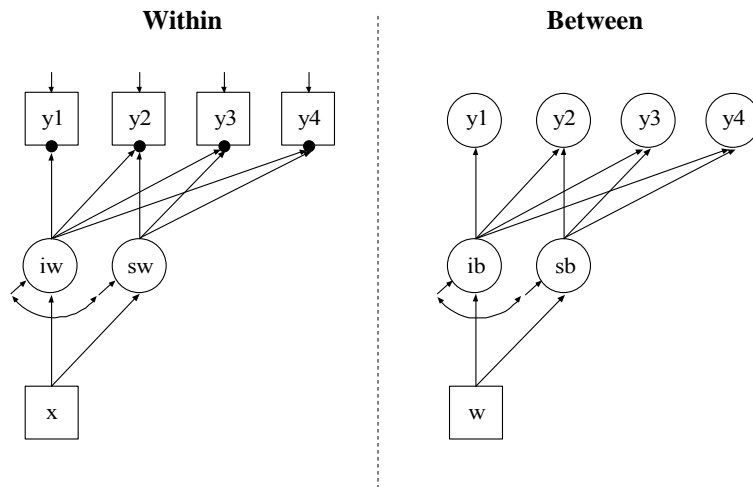
$$\text{Level 1 (Within)} : y_{ij} = \pi_{0ij} + \pi_{1ij} a_{1ij} + \pi_{2ij} a_{2ij} + e_{ij}, \quad (1)$$

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

$$\text{Level 3 (Between)} : \begin{cases} \beta_{00j} = \gamma_{000} + \gamma_{001} w_j + u_{00j}, \\ \beta_{10j} = \gamma_{100} + \gamma_{101} w_j + u_{10j}, \\ \beta_{20j} = \gamma_{200} + \gamma_{201} w_j + u_{20j}, \\ \beta_{01j} = \gamma_{010} + \gamma_{011} w_j + u_{01j}, \\ \beta_{11j} = \gamma_{110} + \gamma_{111} w_j + u_{11j}, \\ \beta_{21j} = \gamma_{210} + \gamma_{211} w_j + u_{21j}. \end{cases} \quad (3)$$

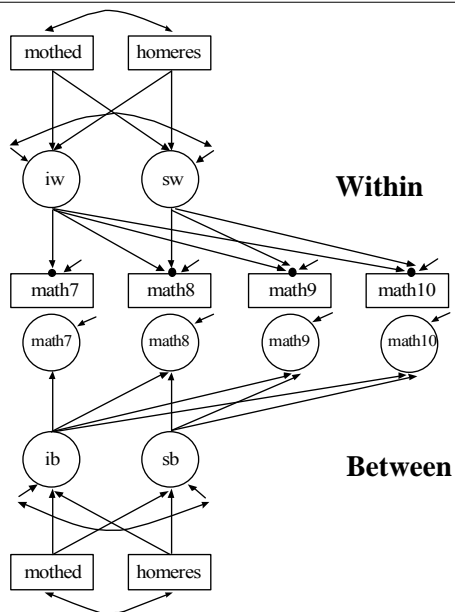
156

## Two-Level Growth Modeling (Three-Level Modeling)



157

## LSAY Two-Level Growth Model



158

## Input For LSAY Two-Level Growth Model With Free Time Scores And Covariates

```
TITLE:      LSAY two-level growth model with free time scores
            and covariates

DATA:      FILE IS lsay98.dat;
            FORMAT IS 3f8 f8.4 8f8.2 3f8 2f8.2;

VARIABLE:  NAMES ARE cohort id school weight math7 math8 math9
            math10 att7 att8 att9 att10 gender mothed homeres;
            USEOBS = (gender EQ 1 AND cohort EQ 2);
            MISSING = ALL (999);
            USEVAR = math7-math10 mothed homeres;
            CLUSTER = school;

ANALYSIS:  TYPE = TWOLEVEL;
            ESTIMATOR = MUML;
```

159

## Input For LSAY Two-Level Growth Model With Free Time Scores And Covariates (Continued)

```
MODEL:     %WITHIN%
            iw sw | math7@0 math8@1
            math9*2 (1)
            math10*3 (2);
            iw sw ON mothed homeres;

            %BETWEEN%
            ib sb | math7@0 math8@1
            math9*2 (1)
            math10*3 (2);
            ib sb ON mothed homeres;

OUTPUT     SAMPSTAT STANDARDIZED RESIDUAL;
```

160

## Output Excerpts LSAY Two-Level Growth Model With Free Time Scores And Covariates

### Summary of Data

Number of clusters            50

Size (s)    Cluster ID with Size s

1	114	
2	136	
6	132	304

34	104
39	309
40	302

Average cluster size 18.627

Estimated Intraclass Correlations for the Y Variables

Variable	Intraclass Correlation	Variable	Intraclass Correlation	Variable	Intraclass Correlation
MATH7	0.199	MATH8	0.149	MATH9	0.168
MATH10	0.165				161

## Output Excerpts LSAY Two-Level Growth Model With Free Time Scores And Covariates (Continued)

### Tests Of Model Fit

Chi-square Test of Model Fit

Value	24.058*
Degrees of Freedom	14
P-Value	0.0451

CFI / TLI

CFI	0.997
TLI	0.995

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.028
----------	-------

SRMR (Standardized Root Mean Square Residual)

Value for Between	0.048
Value for Within	0.007

**Output Excerpts LSAY Two-Level Growth Model  
With Free Time Scores And Covariates (Continued)**

**Model Results**

**Within Level**

SW	BY					
	MATH8	1.000	0.000	0.000	1.073	0.128
	MATH9	2.487	0.163	15.220	2.670	0.288
	MATH10	3.589	0.223	16.076	3.853	0.368
IW	ON					
	MOTHEd	1.780	0.232	7.665	0.246	0.226
	HOMERES	0.892	0.221	4.031	0.124	0.173
SW	ON					
	MOTHEd	0.053	0.063	0.836	0.049	0.045
	HOMERES	0.135	0.044	3.047	0.125	0.176

163

**Output Excerpts LSAY Two-Level Growth Model  
With Free Time Scores And Covariates (Continued)**

SW	WITH					
	IW	2.112	0.522	4.044	0.273	0.273
HOMERES	WITH					
	MOTHEd	0.261	0.039	6.709	0.261	0.203
Residual Variances						
	MATH7	12.748	1.434	8.888	12.748	0.197
	MATH8	12.298	0.893	13.771	12.298	0.174
	MATH9	14.237	1.132	12.578	14.237	0.166
	MATH10	24.829	2.230	11.133	24.829	0.226
	IW	47.060	3.069	15.333	0.903	0.903
	SW	1.110	0.286	3.879	0.964	0.964
Variances						
	MOTHEd	0.841	0.049	17.217	0.841	1.000
	HOMERES	1.970	0.069	28.643	1.970	1.000

164

### Output Excerpts LSAY Two-Level Growth Model With Free Time Scores And Covariates (Continued)

		Estimates	S.E.	Est./S.E.	Std	StdYX
<b>Between Level</b>						
SB	BY					
	MATH8	1.000	0.000	0.000	0.196	0.052
	MATH9	2.487	0.163	15.220	0.488	0.119
	MATH10	3.589	0.223	16.076	0.704	0.115
IB	ON					
	MOTHEd	-1.225	2.587	-0.474	-0.362	-0.107
	HOMERES	7.160	1.847	3.876	2.117	1.011
SB	ON					
	MOTHEd	0.995	0.647	1.538	5.073	1.493
	HOMERES	0.017	0.373	0.045	0.086	0.041
SB	WITH					
	IB	0.382	0.248	1.538	0.575	0.575

165

### Output Excerpts LSAY Two-Level Growth Model With Free Time Scores And Covariates (Continued)

HOMERES	WITH					
	MOTHEd	0.103	0.019	5.488	0.103	0.733
Residual Variances						
	MATH7	2.059	0.552	3.732	2.059	0.153
	MATH8	0.544	0.268	2.033	0.544	0.039
	MATH9	0.105	0.213	0.493	0.105	0.006
	MATH10	1.395	0.504	2.767	1.395	0.067
	IB	1.428	1.690	0.845	0.125	0.125
	SB	-0.051	0.071	-0.713	-1.321	-1.321
Variances						
	MOTHEd	0.087	0.023	3.801	0.087	1.000
	HOMERES	0.228	0.056	4.066	0.228	1.000
Means						
	MOTHEd	2.307	0.043	53.277	2.307	7.838
	HOMERES	3.108	0.062	50.375	3.108	6.509
Intercepts						
	IB	33.510	2.678	12.512	9.909	9.909
	SB	0.163	0.776	0.210	0.830	0.830

166

**Output Excerpts LSAY Two-Level Growth Model  
With Free Time Scores And Covariates (Continued)**

**R-Square**

Within Level

Observed  
Variable R-Square

MATH7 0.803  
MATH8 0.826  
MATH9 0.834  
MATH10 0.774

Latent  
Variable R-Square

IW 0.097  
SW 0.036

167

**Output Excerpts LSAY Two-Level Growth Model  
With Free Time Scores And Covariates (Continued)**

**R-Square**

Between Level

Observed  
Variable R-Square

MATH7 0.847  
MATH8 0.961  
MATH9 0.994  
MATH10 0.933

Latent  
Variable R-Square

IW 0.875  
SW Undefined 0.23207E+01

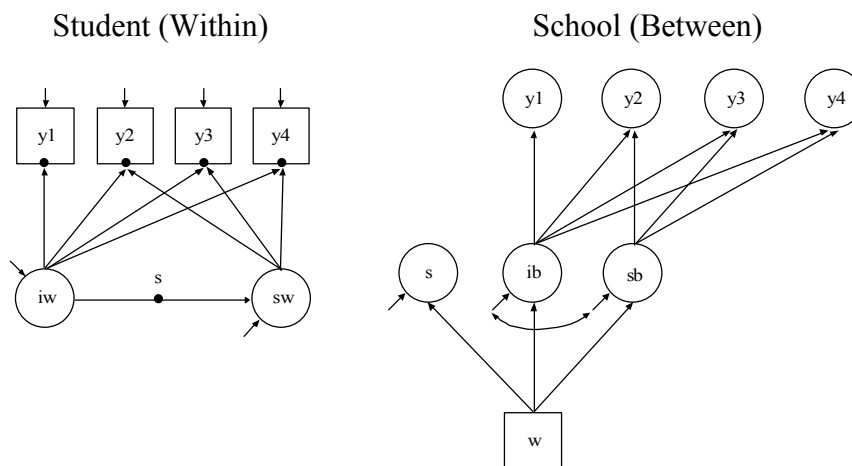
168

## Further Readings On Three-Level Growth Modeling

- Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (ed), *Sociological Methodology* (pp. 453-480). Boston: Blackwell Publishers. (#73)
- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Second edition. Newbury Park, CA: Sage Publications.
- Snijders, T. & Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Thousand Oakes, CA: Sage Publications.

169

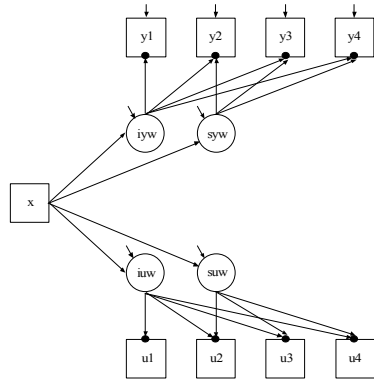
## Multilevel Modeling With A Random Slope For Latent Variables



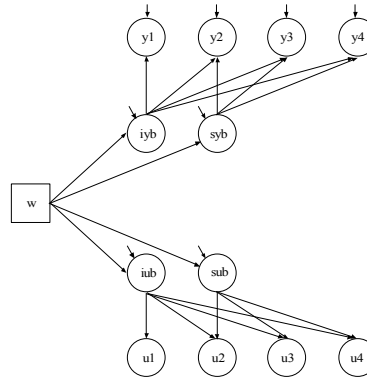
170

## Two-Level, Two-Part Growth Modeling

Within



Between

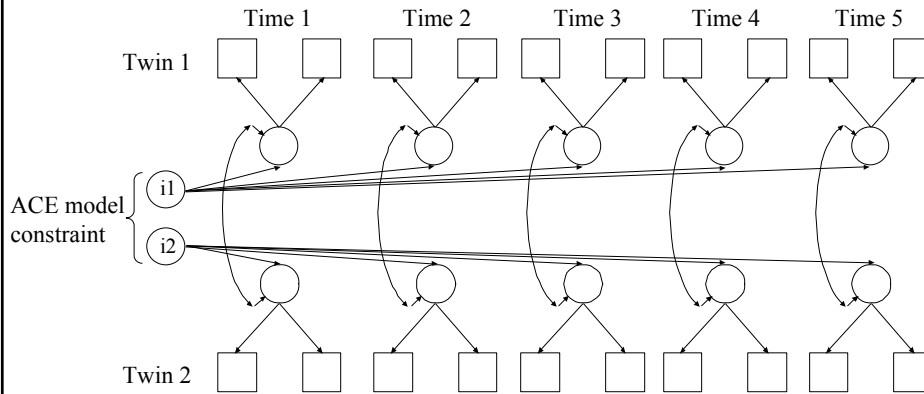


171

## Multiple Indicator Growth Modeling As Two-Level Analysis

172

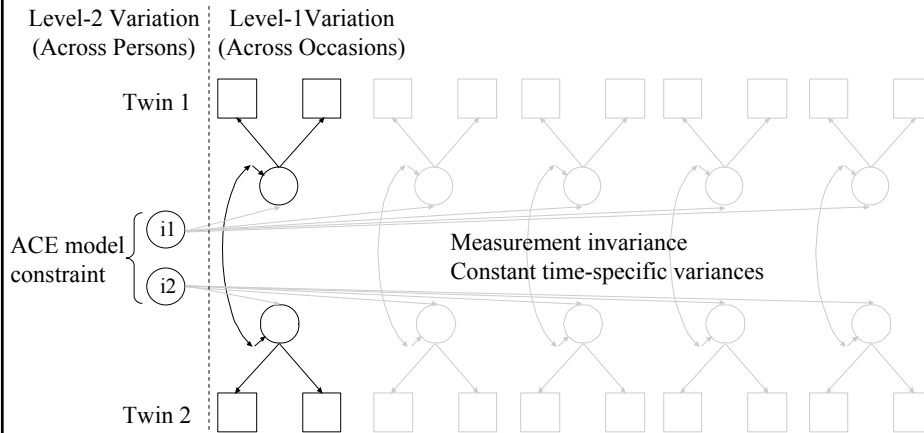
## Wide Data Format, Single-Level Approach



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

173

## 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

174

## **Multilevel Discrete-Time Survival Analysis**

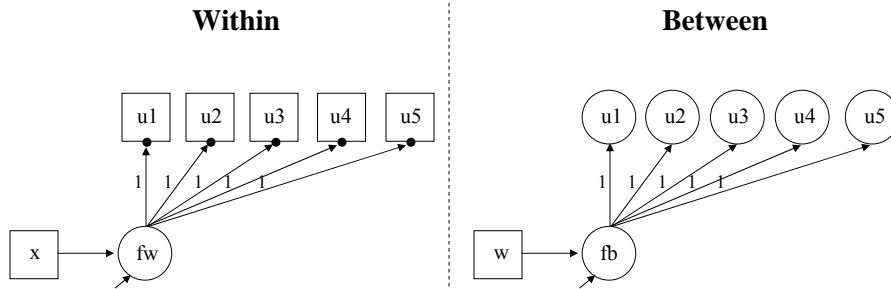
175

## **Multilevel Discrete-Time Survival Analysis**

- Muthén and Masyn (2005) in Journal of Educational and Behavioral Statistics
- Masyn dissertation
- Asparouhov and Muthén

176

## Multilevel Discrete-Time Survival Frailty Modeling



Vermunt (2003)

177

## References

(To request a Muthén paper, please email [bmuthen@ucla.edu](mailto:bmuthen@ucla.edu).)

### Cross-sectional Data

- Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling*, 12, 411-434.
- Chambers, R.L. & Skinner, C.J. (2003). *Analysis of survey data*. Chichester: John Wiley & Sons.
- Harnqvist, K., Gustafsson, J.E., Muthén, B. & Nelson, G. (1994). Hierarchical models of ability at class and individual levels. *Intelligence*, 18, 165-187. (#53)
- Heck, R.H. (2001). Multilevel modeling with SEM. In G.A. Marcoulides & R.E. Schumacker (eds.), *New developments and techniques in structural equation modeling* (pp. 89-127). Lawrence Erlbaum Associates.
- Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Kaplan, D. & Elliott, P.R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling: A Multidisciplinary Journal*, 4, 1-24.
- Kaplan, D. & Ferguson, A.J. (1999). On the utilization of sample weights in latent variable models. *Structural Equation Modeling*, 6, 305-321.

178

## References (Continued)

- Kaplan, D. & Kresiman, M.B. (2000). On the validation of indicators of mathematics education using TIMSS: An application of multilevel covariance structure modeling. International Journal of Educational Policy, Research, and Practice, 1, 217-242.
- Korn, E.L. & Graubard, B.I. (1999). Analysis of health surveys. New York: John Wiley & Sons.
- Kreft, I. & de Leeuw, J. (1998). Introducing multilevel modeling. Thousand Oakes, CA: Sage Publications.
- Larsen & Merlo (2005). Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. American Journal of Epidemiology, 161, 81-88.
- Longford, N.T., & Muthén, B. (1992). Factor analysis for clustered observations. Psychometrika, 57, 581-597. (#41)
- Ludtke Marsh, Robitzsch, Trautwein, Asparouhov, Muthen (2007). Analysis of group level effects using multilevel modeling: Probing a latent covariate approach. Submitted for publication.
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585. (#24)

179

## References (Continued)

- Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society meeting in Princeton, N.J., June 1990. UCLA Statistics Series 62. (#32)
- Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. Journal of Educational Measurement, 28, 338-354. (#37)
- Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox & I. Kreft (eds.), Multilevel Modeling, a special issue of Sociological Methods & Research, 22, 376-398. (#55)
- Muthén, B., Khoo, S.T. & Gustafsson, J.E. (1997). Multilevel latent variable modeling in multiple populations. (#74)
- Muthén, B. & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (ed.), Sociological Methodology 1995, 216-316. (#59)
- Neale, M.C. & Cardon, L.R. (1992). Methodology for genetic studies of twins and families. Dordrecht, The Netherlands: Kluwer.
- Patterson, B.H., Dayton, C.M. & Graubard, B.I. (2002). Latent class analysis of complex sample survey data: application to dietary data. Journal of the American Statistical Association, 97, 721-741.
- Prescott, C.A. (2004). Using the Mplus computer program to estimate models for continuous and categorical data from twins. Behavior Genetics, 34, 17- 40.

180

## References (Continued)

- Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. Second edition. Newbury Park, CA: Sage Publications.
- Skinner, C.J., Holt, D. & Smith, T.M.F. (1989). Analysis of complex surveys. West Sussex, England, Wiley.
- Snijders, T. & Bosker, R. (1999). Multilevel analysis. An introduction to basic and advanced multilevel modeling. Thousand Oakes, CA: Sage Publications.
- Stapleton, L. (2002). The incorporation of sample weights into multilevel structural equation models. Structural Equation Modeling, 9, 475-502.
- Vermunt, J.K. (2003). Multilevel latent class models. In Stolzenberg, R.M. (Ed.), Sociological Methodology (pp. 213-239). New York: American Sociological Association.

### Longitudinal Data

- Choi, K.C. (2002). Latent variable regression in a three-level hierarchical modeling framework: A fully Bayesian approach. Doctoral dissertation, University of California, Los Angeles.

181

## References (Continued)

- Khoo, S.T. & Muthén, 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. (#79)
- Masyn, K. E. (2003). Discrete-time survival mixture analysis for single and recurrent events using latent variables. Doctoral dissertation, University of California, Los Angeles.
- Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (ed.) Sociological Methodology (pp. 453-480). Boston: Blackwell Publishers.
- Muthén, B. (1997). Latent variable growth modeling with multilevel data. In M. Berkane (ed.), Latent variable modeling with application to causality (149-161), New York: Springer Verlag.
- Muthén, B. & Masyn, K. (in press). Discrete-time survival mixture analysis. Journal of Educational and Behavioral Statistics, Spring 2005.
- Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. Second edition. Newbury Park, CA: Sage Publications. Snijders, T. & Bosker, R. (1999). Multilevel analysis. An introduction to basic and advanced multilevel modeling. Thousand Oakes, CA: Sage Publications.

182

## References (Continued)

Seltzer, M., Choi, K., Thum, Y.M. (2002). Examining relationships between where students start and how rapidly they progress: Implications for conducting analyses that help illuminate the distribution of achievement within schools. CSE Technical Report 560. CRESST, University of California, Los Angeles.

### General

### Mplus Analysis

Asparouhov, T. & Muthén, B. (2003a). Full-information maximum-likelihood estimation of general two-level latent variable models. In preparation.

Asparouhov, T. & Muthén, B. (2003b). Maximum-likelihood estimation in general latent variable modeling. In preparation.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. Behaviormetrika, 29, 81-117.

Muthén, B. & Asparouhov, T. (2003b). Advances in latent variable modeling, part II: Integrating continuous and categorical latent variable modeling using Mplus. In preparation.

183

## References (Continued)

### Numerical Integration

Aitkin, M. A general maximum likelihood analysis of variance components in generalized linear models. Biometrics, 1999, 55, 117-128.

Bock, R.D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46, 443-459.

184