General Latent Variable Modeling Using Mplus Version 3

Block 1: Structural Equation Modeling

Bengt Muthén
bmuthen@ucla.edu

Mplus: www.statmodel.com

1

Program 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 language, graphics
  - Powerful: General modeling capabilities

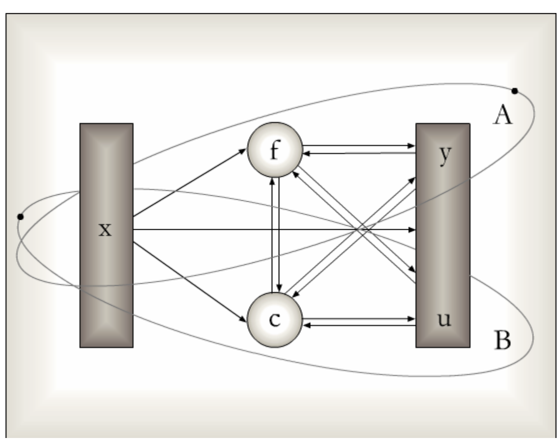
2

The General Modeling Framework of Mplus

- Achieves its flexibility from using a combination of categorical and continuous latent variables
- Special cases:
  - Factor analysis and structural equation modeling
  - Growth modeling
  - Mixture (latent class) modeling
  - Multilevel modeling
  - Combinations

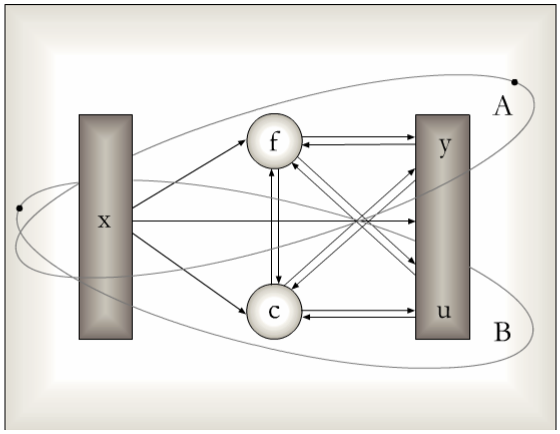
3

General Latent Variable Modeling Framework



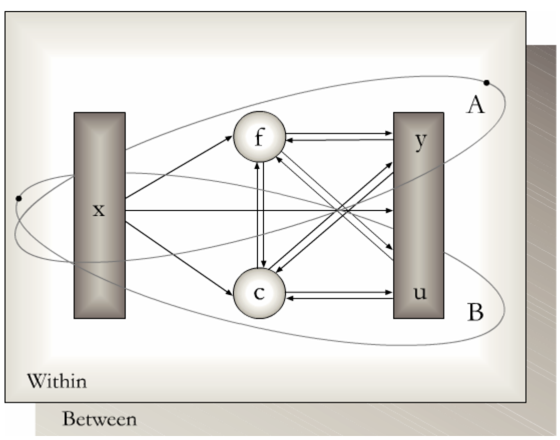
4

General Latent Variable Modeling Framework



5

General Latent Variable Modeling Framework



6

General Latent Variable Modeling Framework

- Muthén, B. (2002). Beyond SEM: General latent variable modeling. Behaviormetrika, 29, 81-117
- Asparouhov & Muthen (2004). Maximum-likelihood estimation in general latent variable modeling
- Muthen & Muthen (1998-2004). Mplus Version 3
- Mplus team: Linda Muthen, Bengt Muthen, Tihomir Asparouhov, Thuy Nguyen, Michelle Conn
  (see www.statmodel.com)

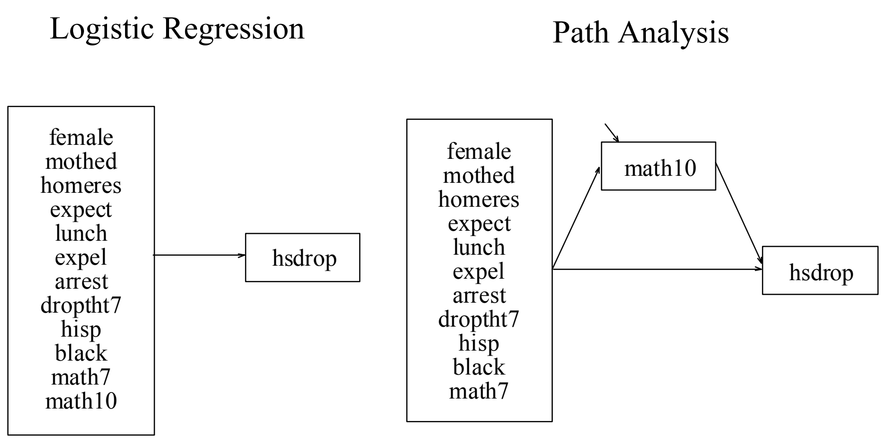
7

Continuous Latent Variables

- Factor analysis, structural equation modeling
  - Constructs measured with multiple indicators
- Growth modeling
  - Growth factors, random effects: random intercepts and random slopes representing individual differences of development over time (unobserved heterogeneity)
- Survival analysis
  - Frailties
- Missing data modeling

8

Path Analysis with a Categorical Outcome and Missing Data on a Mediator



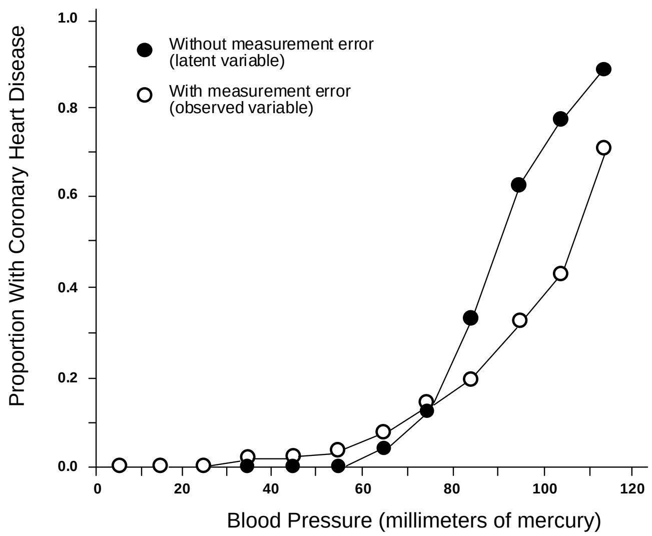
9

Continuous Latent Variables: Two Examples

- Muthen (1992). Latent variable modeling in epidemiology. Alcohol Health & Research World, 16, 286-292
  - Blood pressure predicting coronary heart disease
- Nurses' Health Study (Rosner, Willet & Spiegelman, 1989). Nutritional study of 89,538 women.
  - Dietary fat intake questionnaire for everyone
  - Dietary diary for 173 women for 4 1-week periods at 3-month intervals

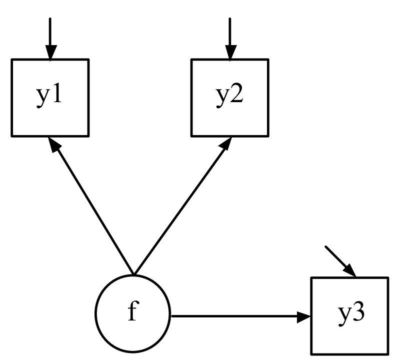
10

Measurement Error in a Covariate



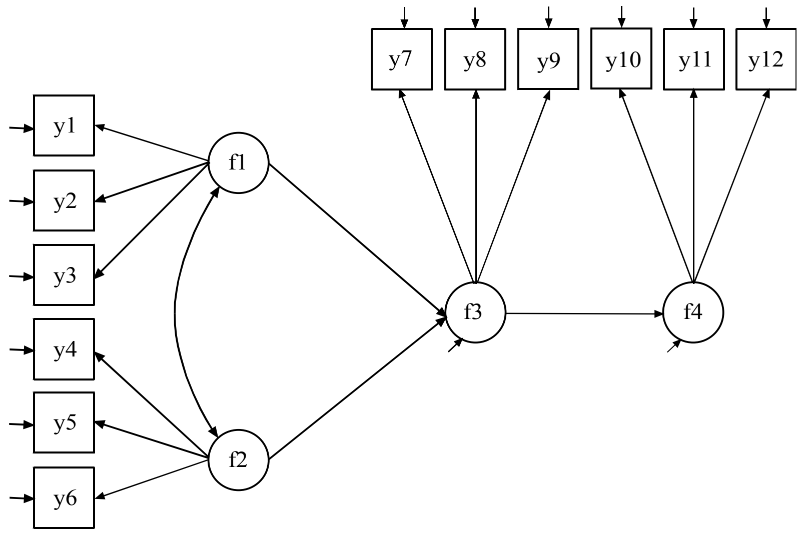
11

Measurement Error in a Covariate



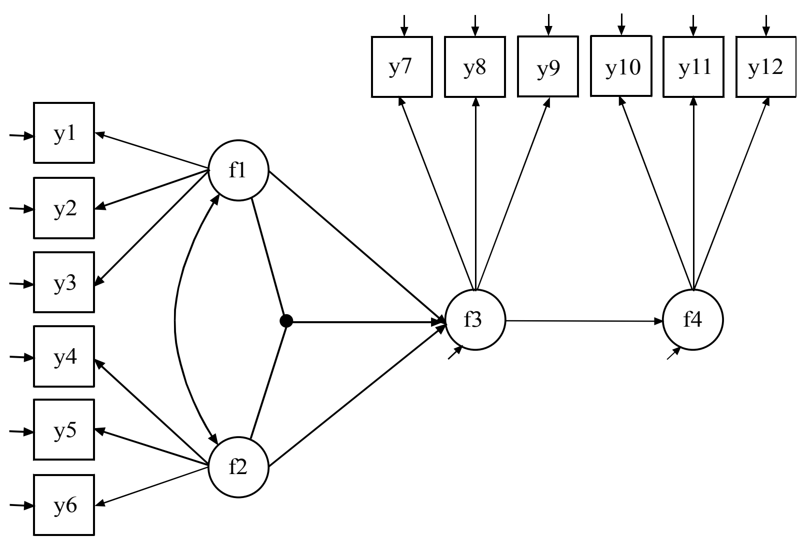
12

Structural Equation Model



13

Structural Equation Model with Interaction between Latent Variables



14

Antisocial Behavior (ASB) Data

The antisocial Behavior (ASB) 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.

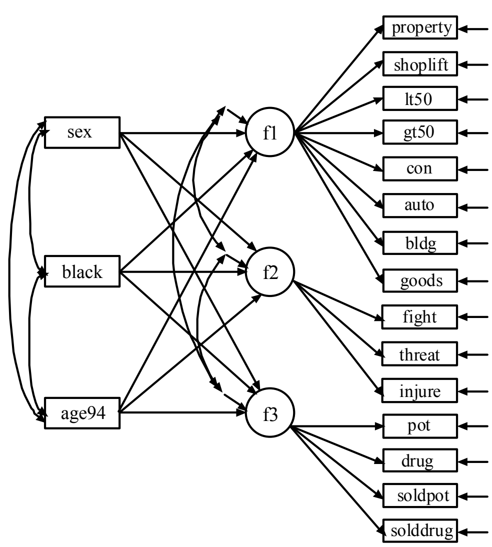
Data for the analysis include 15 of the 17 antisocial behavior items that were collected in 1980 when respondents were between the ages of 16 and 23 and the background variables of age, gender, and ethnicity. The ASB items assessed the frequency of various behaviors during the past year. A sample of 7,326 respondents has complete data on the antisocial behavior items and the background variables of age, gender, and ethnicity. Following is a list of the 15 items:

Damaged property
Fighting
Shoplifting
Stole < \$50
Stole > \$50
Seriously threaten
Intent to injure
Use marijuana
Use other drugs
Sold marijuana
Sold hard drugs
"Con" someone
Take auto
Broken into building
Held stolen goods

These items were dichotomized 0/1 with 0 representing never in the last year. An EFA suggested three factors: property offense, person offense, and drug offense.

15

ASB CFA With Covariates



16

Input For CFA With Covariates With Categorical Outcomes For 15 ASB Items

```
TITLE:      CFA with covariates with categorical outcomes using
            15 antisocial behavior items and 3 covariates

DATA:       FILE IS asb.dat;
            FORMAT IS 34X 54F2.0;

VARIABLE:   NAMES ARE property fight shoplift lt50 gt50 force
            threat injure pot drug soldpot solddrug con auto bldg
            goods gambling dsm1-dsm22 sex black hisp single
            divorce dropout college onset fhist1 fhist2 fhist3
            age94 cohort dep abuse;

            USEV ARE property-gt50 threat-goods sex black age94

            CATEGORICAL ARE property-goods;
```

17

Input For CFA With Covariates With Categorical Outcomes For 15 ASB Items (Continued)

```
ANALYSIS:   TYPE = MEANSTRUCTURE;

MODEL:      f1 BY property shoplift-gt50 con-goods;

            f2 BY fight threat injure;

            f3 BY pot-solddrug;

            f1-f3 ON sex black age94;

            property-goods ON sex-age94@0;

OUTPUT:     STANDARDIZED TECH2;
```

18

Output Excerpts CFA With Covariates With Categorical Outcomes For 15 ASB Items

Model Results

		Estimates	S.E.	Est./S.E.	Std	StdYX
F1 BY						
	PROPERTY	1.000	.000	.000	.791	.760
	SHOPLIFT	.974	.023	42.738	.771	.742
	LT50	.915	.023	39.143	.724	.700
	GT50	1.055	.031	33.658	.835	.799
	CON	.752	.024	31.637	.595	.581
	AUTO	.796	.030	26.462	.629	.613
	BLDG	1.084	.030	35.991	.858	.818
	GOODS	1.071	.025	42.697	.847	.809

19

Output Excerpts CFA With Covariates With Categorical Outcomes For 15 ASB Items (Continued)

F2 BY						
	FIGHT	1.000	.000	.000	.773	.734
	THREAT	1.096	.035	31.382	.847	.797
	INJURE	1.082	.037	28.888	.836	.787
F3 BY						
	POT	1.000	.000	.000	.866	.851
	DRUG	1.031	.023	45.818	.893	.876
	SOLDPOT	1.046	.023	45.844	.905	.888
	SOLDDRUG	.923	.036	25.684	.799	.787

20

Output Excerpts CFA With Covariates With Categorical Outcomes For 15 ASB Items (Continued)

```
F1         ON
   SEX              .516      .024   21.206      .653      .326
   BLACK           -.080      .025   -3.168     -.102     -.047
   AGE94           -.054      .006   -9.856     -.069     -.150

F2         ON
   SEX              .561      .026   21.715      .726      .363
   BLACK            .174      .025    7.087      .225      .103
   AGE94           -.068      .006  -12.286     -.087     -.191

F3         ON
   SEX              .229      .026    8.760      .265      .132
   BLACK           -.272      .029   -9.384     -.315     -.144
   AGE94            .039      .006    6.481      .045      .099
```

21

Output Excerpts CFA With Covariates With Categorical Outcomes For 15 ASB Items (Continued)

Tests Of Model Fit

```
Chi-square Test of Model Fit
            Value                         1225.266*
            Degrees of Freedom                 105**
            P-Value                           0.0000
CFI / TLI
            CFI                                0.945
            TLI                                0.964
RMSEA (Root Mean Square Error Of Approximation)
            Estimate                           0.038
WRMR (Weighted Root Mean Square Residual)
            Value                              2.498
```

22

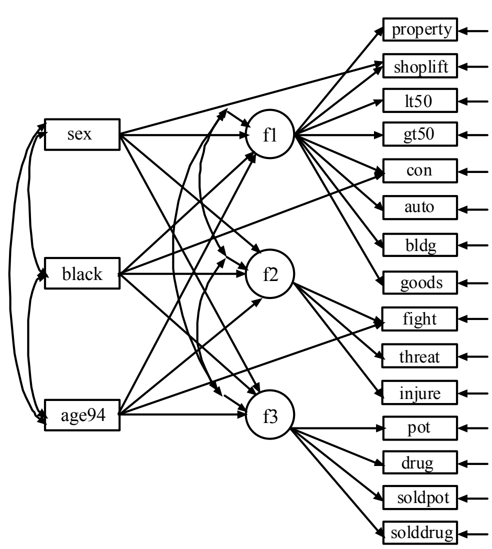
Output Excerpts (Continued)

Tech2 Derivatives With Respect To Gamma

	SEX	BLACK	AGE94
F1	.000	.000	.000
F2	.000	.000	.000
F3	.000	.000	.000
PROPERTY	-.019	.006	.072
FIGHT	-.023	-.015	**.109**
SHOPLIFT	**.039**	.001	.003
LT50	-.001	.014	-.072
GT50	-.007	-.008	-.026
THREAT	.009	.015	-.026
ONJURE	.012	-.001	-.074
POT	.011	-.010	-.058
DRUG	.012	.016	-.016
SOLDPOT	-.019	-.003	.060
SOLDDRUG	-.003	-.004	.013
CON	.020	**-.030**	.051
AUTO	.002	.003	.020
BLDG	-.012	.005	-.003
GOODS	-.013	.003	-.030

23

ASB Model With Direct Effects



24

Summary of Analysis Results For ASB CFA With Covariates And Direct Effects

Input Specification

```
MODEL:      f1 BY property shoplift-gt50 con-goods;
            f2 BY fight threat injure;
            f3 BY pot-solddrug;

            f1-f3 ON sex black age94;

            shoplift ON sex;
            con ON black;
            fight ON age94;
```

25

Summary of Analysis Results For ASB CFA With Covariates And Direct Effects

Test Of Model Fit

```
Chi-square Test of Model Fit
            Value                      946.256*
            Degrees of Freedom             102**
            P-Value                     0.0000

CFI/TLI
            CFI                          0.959
            TLI                          0.972

RMSEA (Root Mean Square Error Of Approximation)
            Estimate                     0.034
WRMR (Weighted Root Mean Square Residual)
            Value                        2.198
```

26

Summary of Analysis Results For ASB CFA With Covariates And Direct Effects (Continued)

```
                 Estimates    S.E.     Est./S.E.     Std       StdYX

F1         BY
   SHOPLIFT       1.002      .024       42.183      .805      .793

F1         ON
   SEX             .596      .026       22.958      .742      .371

SHOPLIFT  ON
   SEX            -.385      .033      -11.594     -.385     -.190
CON        ON
   BLACK           .305      .034        8.929      .305      .136
FIGHT      ON
   AGE94          -.068      .008       -8.467     -.068    - .138

Thresholds

   SHOPLIFT$1       .558     .033       17.015      .558      .558

R-SQUARE

       Observed   Residual
       Variable   Variance   R-Square

       SHOPLIFT       .461       .552
```

27

Interpretation of Direct Effects

- Look at indirect effect of covariate on factor indicator via the factor
- Look at the direct effect of covariate on factor indicator

Shoplift On Gender

- Indirect effect of gender on shoplift
  - F1 has a positive relationship with gender – males have a higher mean than females on the f1 factor
  - Shoplift has a positive loading on the f1 factor
  - Conclusion: males are expected to have a higher probability of shoplifting
- Effect of gender on shoplift
  - Direct effect is negative – for a given factor value, males have a lower probability of shoplifting than females
  - Conclusion – shoplift is not invariant

28

Calculating Item Probabilities

The model with a direct effect from x to item y_j,

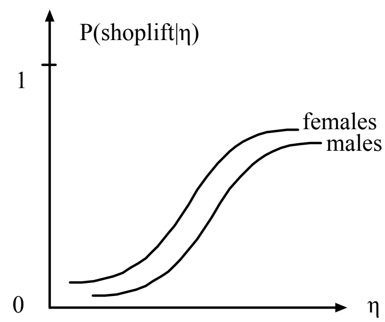
$$y^*_{ij} = \lambda_j \eta_i + \kappa_j x_i + \varepsilon_{ij}, \qquad (45)$$

gives the conditional probability of a $y = 1$ response

$$P(y_{ij} = 1 \mid \eta_{ij}, x_i) = 1 - F[(\tau_j - \lambda_j \eta_i - \kappa_j x_i)\ \theta_{jj}^{-1/2}], \qquad (46)$$

where F is the normal distribution function.

For example, for the item shoplift, $\tau_j = 0.558$, $\kappa_j = -0.385$, $\theta_{jj} = 0.461$. At $\eta = 0$, the probability is 0.21 for females ($x = 0$) and 0.08 for males ($x = 1$).



29

Calculating Item Probabilities Cont'd

Consider

$$P(y_{ij} = 1 \mid \eta_{ij}, x_i) = 1 - \mathrm{F}[(\tau_j - \lambda_j \eta_i - \kappa_j x_i)\ \theta_{jj}^{-1/2}], \qquad (47)$$

Using $\tau_j = 0.558$, $\kappa_j = -0.385$, $\theta_{jj} = 0.461$, and $\eta = 0$.

Here, $\theta_{jj}^{-1/2} = \frac{1}{\sqrt{\theta_{jj}}} = \frac{1}{\sqrt{0.461}} = 1.473$.

For females ($x = 0$):

1. $(\tau_j - \lambda_j \eta_i - \kappa_j x_i) = 0.558 - 1.002 \times 0 - (-0.385) \times 0 = 0.558$.
2. $(\tau_j - \lambda_j \eta_i - \kappa_j x_i)\theta_{jj}^{-1/2} = 0.558 \times 1.473 = 0.822$.
3. $F[0.822] = 0.794$ using a z table
4. $1 - 0.794 = 0.206$.

For males ($x = 1$):

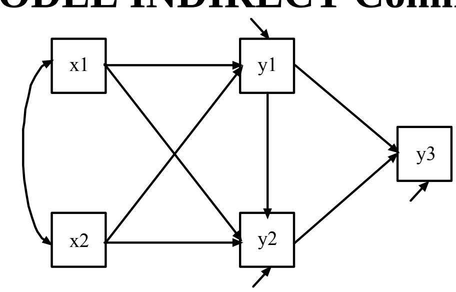
1. $(\tau_j - \lambda_j \eta_i - \kappa_j x_i) = 0.558 - 1.002 \times 0 - (-0.385) \times 1 = 0.943$.
2. $(\tau_j - \lambda_j \eta_i - \kappa_j x_i)\theta_{jj}^{-1/2} = 0.943 \times 1.473 = 1.389$.
3. $F[1.389] = 0.918$ using a z table.
4. $1 - 0.918 = 0.082$.

30

New SEM Features In Mplus Version 3

31

The MODEL INDIRECT Command



MODEL INDIRECT has two options:

- IND – used to request a specific indirect effect or a set of indirect effects
- VIA – used to request a set of indirect effects that includes specific mediators

```
MODEL INDIRECT
    y3 IND y1 x1;        ! x1 -> y1 -> y3
    y3 IND y2 x2;        ! x2 -> y2 -> y3
    y3 IND x1;           ! x1 -> y1 -> y3
                         ! x1 -> y2 -> y3
                         ! x1 -> y1 -> y2 -> y3
    y3 VIA y2 x1;        ! x1 -> y2 -> y3
                         ! x1 -> y1 -> y2 -> y3
```

32

The MODEL INDIRECT Command

MODEL INDIRECT is used to request indirect effects and their standard errors. Delta method standard errors are computed as the default.

The STANDARDIZED option of the OUTPUT command can be used to obtain standardized indirect effects.

The BOOTSTRAP option of the ANALYSIS command can be used to obtain bootstrap standard errors for the indirect effects.

The CINTERVAL option of the OUTPUT command can be used to obtain confidence intervals for the indirect effects and the standardized indirect effects. Three types of 95% and 99% confidence intervals can be obtained: symmetric, bootstrap, or bias-corrected bootstrap confidence intervals. The bootstrapped distribution of each parameter estimate is used to determine the bootstrap and bias-corrected bootstrap confidence intervals. These intervals take non-normality of the parameter estimate distribution into account. As a result, they are not necessarily symmetric around the parameter estimate.

The MODEL CONSTRAINT Command

MODEL CONSTRAINT is used to define linear and non-linear constraints on the parameters in the model. All functions available in the DEFINE command are available for linear and non-linear constraints. Parameters in the model are given labels by placing a name in parentheses after the parameter.

```
MODEL: y ON x1 (p1)
            x2 (p2)
            x3 (p3);

MODEL CONSTRAINT:
        p1 = p2**2 + p3**2;
```

Interaction Modeling Using ML For Observed And Latent Variables

Types of Variables	Interaction Options
observed continuous with observed continuous	DEFINE
observed categorical with observed continuous	DEFINE Multiple Group
observed continuous with continuous latent	XWITH
observed categorical with continuous latent	XWITH Multiple Group
observed continuous with categorical latent	MIXTURE
observed categorical with categorical latent	MIXTURE KNOWNCLASS
continuous latent with continuous latent	XWITH
continuous latent with categorical latent	MIXTURE
categorical latent with categorical latent	MIXTURE

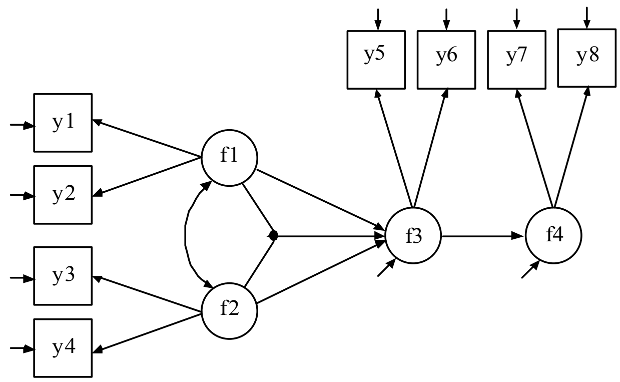
35

The XWITH Option Of The MODEL Command

The XWITH option is used with TYPE=RANDOM to define interactions between continuous latent variables or between continuous latent variables and observed variables. XWITH is short for multiplied with. It is used in conjunction with the | symbol to name and define interaction variables in a model. Following is an example of how to use XWITH and the | symbol to name and define an interaction:

f1f2 | f1 XWITH f2;
f1y | f1 XWITH y;

36



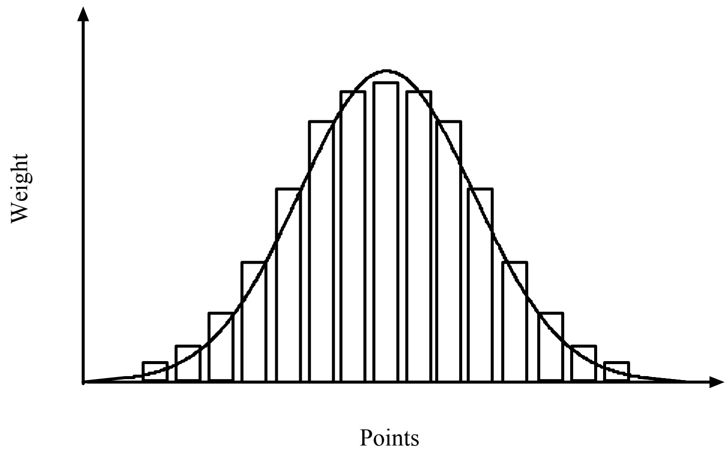
37

Input For An SEM Model With An Interaction Between Two Latent Variables

```
TITLE:      this an example of a structural equation model with
            an
            interaction between two latent variables
DATA:       FILE = firstSEMInter.dat;
VARIABLE:   NAMES = y1-y8;
ANALYSIS:   TYPE = RANDOM;
            ALGORITH = INTEGRATION;
MODEL:      f1 BY y1 y2;
            f2 BY y3 y4;
            f3 BY y5 y6;
            f4 BY y7 y8;

            f4 ON f3;
            f3 ON f1 f2;

            f1f2 | f1 XWITH f2;

            f3 ON f1f2;
OUTPUT:     TECH8;
```

38

Numerical Integration



39

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.

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.

40

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

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 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.

41

Numerical Integration Theory

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

$$[\, f \,|\, y\,] = [\, f\,]\,[\, y \,|\, f\,] \,/\, [\, y\,], \tag{97}$$

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

$$[\, y\,] = \int [\, f\,]\,[\, y \,|\, f\,]\, df. \tag{98}$$

- Numerical integration is not needed: Normally distributed y – the posterior distribution is normal
- Numerical integration is needed:
  - 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

$$[\, y\,] = \int [\, f_1, f_2\,]\,[\, y | f_1, f_2, f_1 f_2\,]\, df_1\, df_2 \tag{99}$$

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

Numerical integration approximates the integral by a sum

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

42

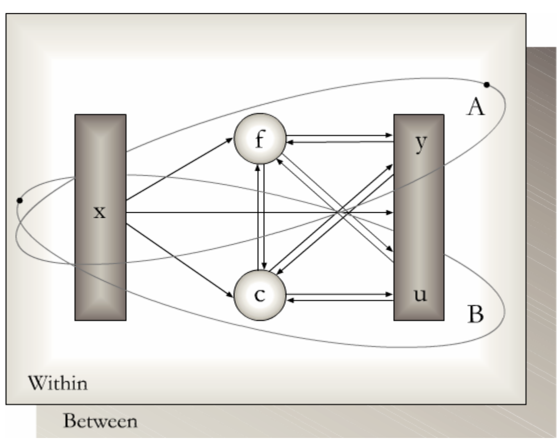
General Latent Variable Modeling
Using Mplus Version 3

Block 2: Growth Modeling

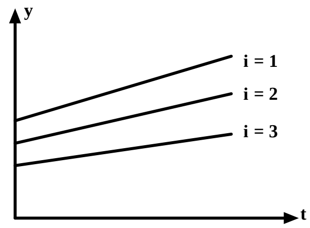
Bengt Muthén
bmuthen@ucla.edu

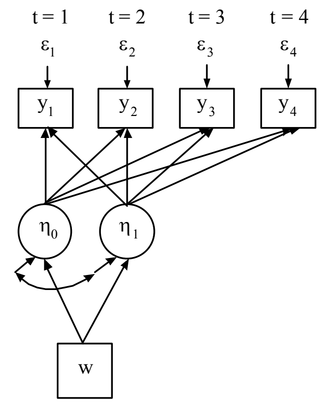
Mplus: www.statmodel.com

General Latent Variable Modeling Framework



Individual Development Over Time





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

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

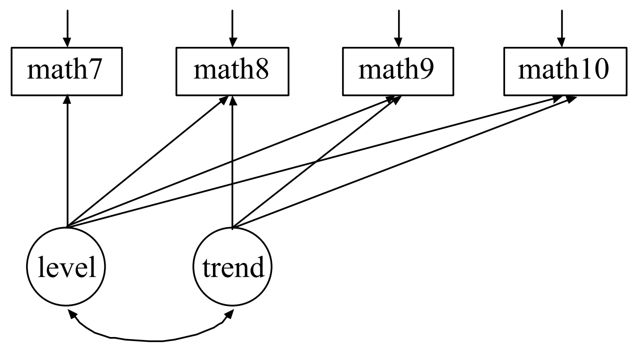
(2b) $\eta_{1i} = \alpha_1 + \gamma_1\, w_i + \zeta_{1i}$

3

Advantages of Growth Modeling in a Latent Variable Framework

- Flexible curve shape
- Individually-varying times of observation
- Random effects (intercepts, slopes) integrated with other latent variables
- Regressions among random effects
- Multiple processes
- Multiple populations
- Multiple indicators
- Embedded growth models
- Categorical latent variables: growth mixtures

4



5

Input For LSAY Linear Growth Model Without Covariates

```
TITLE:    LSAY For Younger Females With Listwise Deletion
          Linear Growth Model Without Covariates

DATA:     FILE IS lsay.dat;
          FORMAT IS 3F8.0 F8.4 8F8.2 3F8.0;

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;
```

6

Input For LSAY Linear Growth Model Without Covariates (Continued)

```
ANALYSIS:      TYPE = MEANSTRUCTURE;

MODEL:    level BY math7-math10@1;
          trend BY math7@0 math8@1 math9@2 math10@3;
          [math7-math10@0];
          [level trend];

OUTPUT:   SAMPSTAT STANDARDIZED MODINDICES (3.84);

!New Version 3 Language For Growth Models
!MODEL: level trend | math7@0 math8@1 math9@2 math10@3
```

7

Output Excerpts LSAY Linear Growth Model Without Covariates

Tests of Model Fit

```
Chi-Square Test of Model Fit
          Value                        22.664
          Degrees of Freedom                5
          P-Value                      0.0004

CFI/TLI
          CFI                           0.995
          TLI                           0.994

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                      0.060
          90 Percent C.I.               0.036  0.086
          Probability RMSEA <= .05      0.223

SRMR (Standardized Root Mean Square Residual)
          Value                         0.025
```

8

Output Excerpts LSAY Linear Growth Model Without Covariates (Continued)

Modification Indices

		M.I.	E.P.C.	Std.E.P.C.	StdYX E.P.C.
TREND	BY MATH7	6.793	0.185	0.254	0.029
TREND	BY MATH8	14.694	-0.169	-0.233	-0.025
TREND	BY MATH9	9.766	0.155	0.213	0.021

9

Output Excerpts LSAY Linear Growth Model Without Covariates

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
LEVEL BY					
MATH7	1.000	.000	.000	8.029	.906
MATH8	1.000	.000	.000	8.029	.861
MATH9	1.000	.000	.000	8.029	.800
MATH10	1.000	.000	.000	8.029	.708
TREND BY					
MATH7	.000	.000	.000	.000	.000
MATH8	1.000	.000	.000	1.377	.148
MATH9	2.000	.000	.000	2.753	.274
MATH10	3.000	.000	.000	4.130	.364

10

Output Excerpts LSAY Linear Growth Model Without Covariates (Continued)

LEVEL WITH					
TREND	3.491	.730	4.780	.316	.316
Residual Variances					
MATH7	14.105	1.253	11.259	14.105	.180
MATH8	13.525	.866	15.610	13.525	.156
MATH9	14.726	.989	14.897	14.726	.146
MATH10	25.989	1.870	13.898	25.989	.202
Variances					
LEVEL	64.469	3.428	18.809	1.000	1.000
TREND	1.895	.322	5.894	1.000	1.000

11

Output Excerpts LSAY Linear Growth Model Without Covariates (Continued)

Means					
LEVEL	52.623	.275	191.076	6.554	6.554
TREND	3.105	.075	41.210	2.255	2.255
Intercepts					
MATH7	.000	.000	.000	.000	.000
MATH8	.000	.000	.000	.000	.000
MATH9	.000	.000	.000	.000	.000
MATH10	.000	.000	.000	.000	.000

R-Square

Observed Variable	R-Square
MATH7	0.820
MATH8	0.844
MATH9	0.854
MATH10	0.798

12

Growth Model With Individually Varying Times Of Observation And Random Slopes For Time-Varying Covariates

13

Growth Modeling In Multilevel Terms

Time point t, individual i (two-level modeling, no clustering):

y_{ti} : repeated measures on the outcome, e.g. math achievement
a_{1ti} : time-related variable (time scores); e.g. grade 7-10
a_{2ti} : time-varying covariate, e.g. math course taking
x_i : time-invariant covariate, e.g. grade 7 expectations

Two-level analysis with individually-varying times of observation and random slopes for time-varying covariates:

$$\textit{Level } 1\text{: } y_{ti} = \pi_{0i} + \pi_{1i}\, a_{1ti} + \pi_{2ti}\, a_{2ti} + e_{ti}\,, \qquad (55)$$

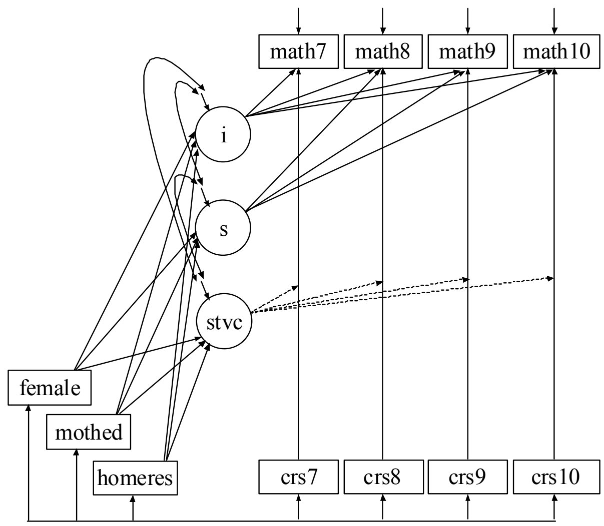
$$\textit{Level } 2\text{: } \begin{cases} \pi_{0i} = \beta_{00} + \beta_{01}\, x_i + r_{0i}\,, \\ \pi_{1i} = \beta_{10} + \beta_{11}\, x_i + r_{1i}\,, \\ \pi_{2i} = \beta_{20} + \beta_{21}\, x_i + r_{2i}\,. \end{cases} \qquad (56)$$

Time scores a_{1ti} read in as data (not loading parameters).

- π_{2ti} possible with time-varying random slope variances
- Flexible correlation structure for $V(e) = \Theta$ (T x T)
- Regressions among random coefficients possible, e.g.

$$\pi_{1i} = \beta_{10} + \gamma_1\, \pi_{0i} + \beta_{11}\, x_i + r_{1i}\,, \qquad (57)$$
$$\pi_{2i} = \beta_{20} + \gamma_2\, \pi_{0i} + \beta_{21}\, x_i + r_{2i}\,. \qquad (58)$$

14



15

Input For Growth Model With Individually Varying Times Of Observation

```
TITLE:    growth model with individually varying times of
          observation and random slopes

DATA:     FILE IS lsaynew.dat;

VARIABLE: NAMES ARE math7 math8 math9 math10 crs7 crs8 crs9
          crs10 female mothed homeres a7-a10;

          ! crs7-crs10 = highest math course taken during each
          ! grade (0-no course, 1=low,basic, 2=average, 3=high,
          ! 4=pre-algebra, 5=algebra I, 6=geometry,
          ! 7=algebra II, 8=pre-calc, 9=calculus)

          MISSING ARE ALL(9999);
          CENTER = GRANDMEAN(crs7-crs10 mothed homeres);
          TSCORES = a7-a10;

DEFINE:   math7 = math7/10;
          math8 = math8/10;
          math9 = math9/10;
          math10 = math10/10;
```

16

Input For Growth Model With Individually Varying Times Of Observation (Continued)

```
ANALYSIS: TYPE = RANDOM MISSING;
          ESTIMATOR = ML;
          MCONVERGENCE = .001;

MODEL:    i  s  | math7-math10 AT a7-a10;

          stvc  | math7 ON crs7;
          stvc  | math8 ON crs8;
          stvc  | math9 ON crs9;
          stvc  | math10 ON crs10;

          i ON female mothed homeres;
          s ON female mothed homeres;
          stvc on female mothed homeres;

          i WITH s;
          stvc WITH i;
          stvc WITH s;

OUTPUT:   TECH8;
```

17

Output Excerpts For Growth Model With Individually Varying Times Of Observation And Random Slopes For Time-Varying Covariates

n = 2271

Tests of Model Fit

```
Loglikelihood

          H0 Value                    -8199.311

Information Criteria

          Number of Free Parameters          22
          Akaike (AIC)                16442.623
          Bayesian (BIC)              16568.638
          Sample-Size Adjusted BIC    16498.740
             (n* = (n + 2) / 24)
```

18

Output Excerpts For Growth Model With Individually Varying Times Of Observation And Random Slopes For Time-Varying Covariates (Continued)

```
                          Estimates      S.E.     Est./S.E.
I          ON
    FEMALE                   0.187       0.036        5.247
    MOTHED                   0.187       0.018       10.231
    HOMERES                  0.159       0.011       14.194
S          ON
    FEMALE                  -0.025       0.012       -2.017
    MOTHED                   0.015       0.006        2.429
    HOMERES                  0.019       0.004        4.835
STVC       ON
    FEMALE                  -0.008       0.013       -0.590
    MOTHED                   0.003       0.007        0.429
    HOMERES                  0.009       0.004        2.167
I          WITH
    S                        0.038       0.006        6.445
STVC       WITH
    I                        0.011       0.005        2.087
    S                        0.004       0.002        2.033
```


Output Excerpts For Growth Model With Individually Varying Times Of Observation And Random Slopes For Time-Varying Covariates (Continued)

```
Intercepts
    MTH7                     0.000       0.000        0.000
    MTH8                     0.000       0.000        0.000
    MTH9                     0.000       0.000        0.000
    MTH10                    0.000       0.000        0.000
    I                        4.992       0.025      198.456
    S                        0.417       0.009       47.275
    STVC                     0.113       0.010       11.416

Residual Variances
    MTH7                     0.185       0.011       16.464
    MTH8                     0.178       0.008       22.232
    MTH9                     0.156       0.008       18.497
    MTH10                    0.169       0.014       12.500
    I                        0.570       0.023       25.087
    S                        0.036       0.003       12.064
    STVC                     0.012       0.002        5.055
```

Random Slopes

- In single-level modeling random slopes β_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

$$V(y_i \mid x_i) = V(\beta_i)\; x_i^2 + \theta \quad (103)$$

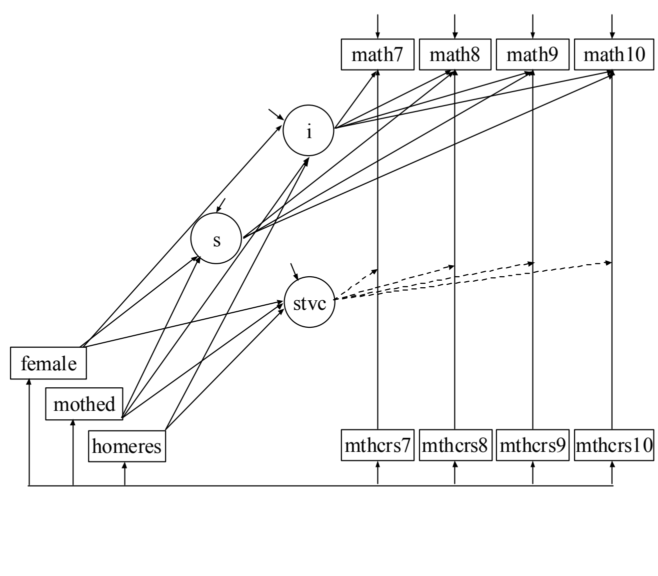
- In two-level modeling random slopes β_j describe variation across clusters *j*

$$y_{ij} = a_j + \beta_j\, x_{ij} + \varepsilon_{ij} , \quad (104)$$
$$a_j = a + \zeta_{0j} , \quad (105)$$
$$\beta_j = \beta + \zeta_{1j} , \quad (106)$$

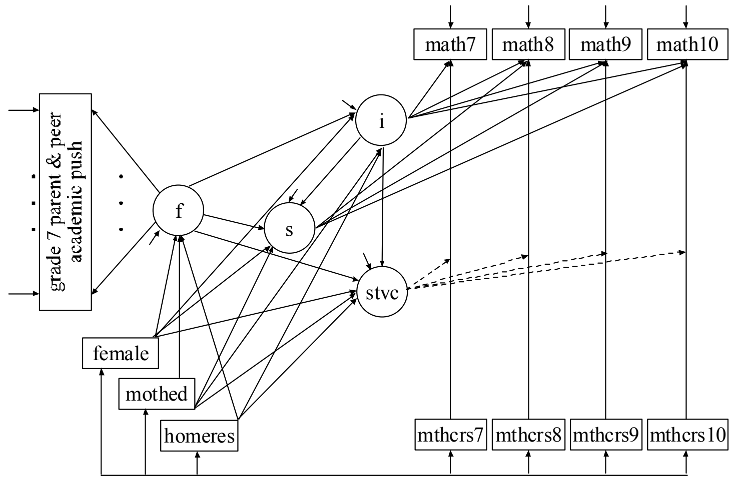
A small variance for a random slope typically leads to slow convergence of the ML-EM iterations. This suggests respecifying the slope as fixed.

Mplus allows random slopes for predictors that are
- Observed covariates
- Observed dependent variables (Version 3)
- Continuous latent variables (Version 3)

Growth Modeling with Time-Varying Covariates

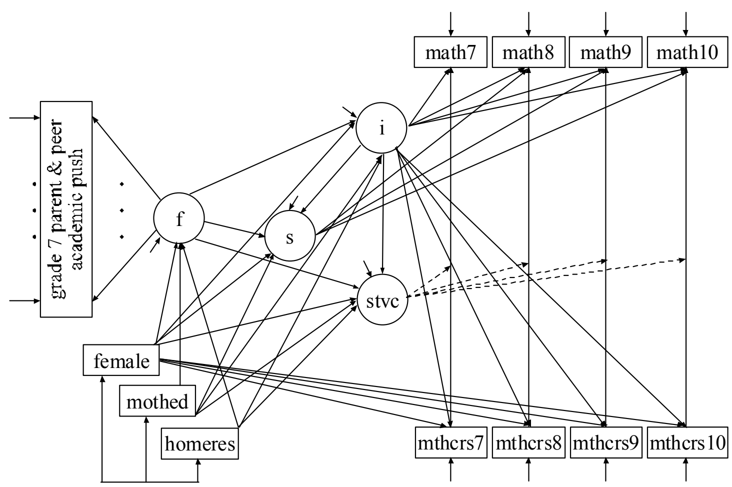


A Generalized Growth Model



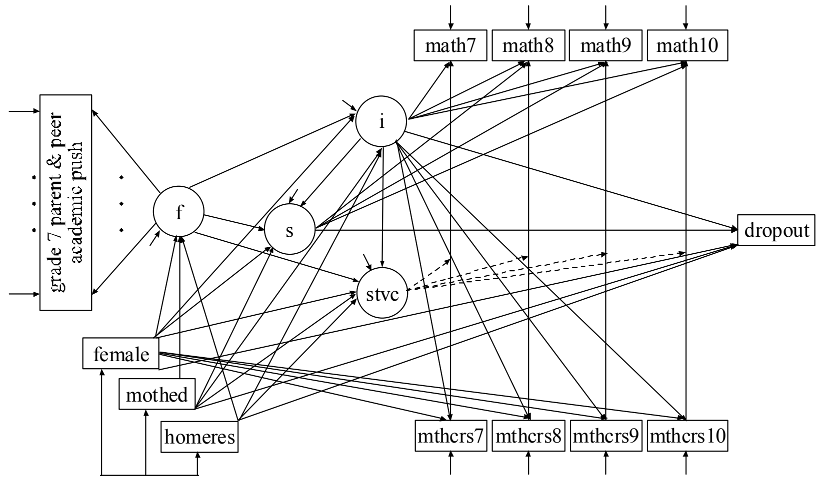
23

A Generalized Growth Model



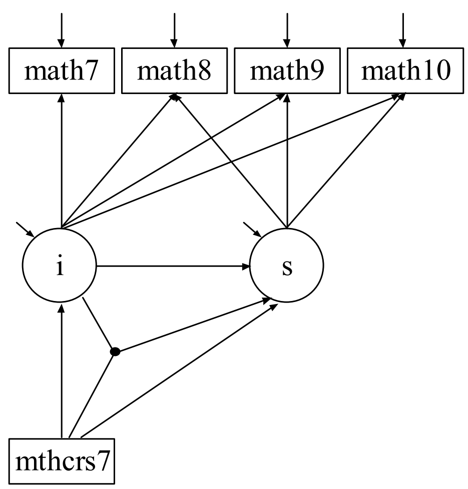
24

A Generalized Growth Model



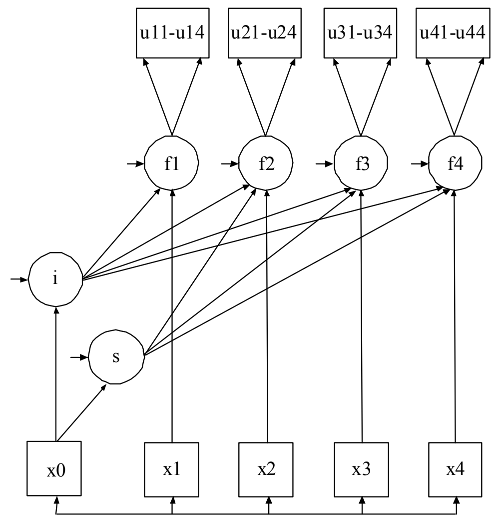
25

Growth Modeling with a Latent Variable Interaction



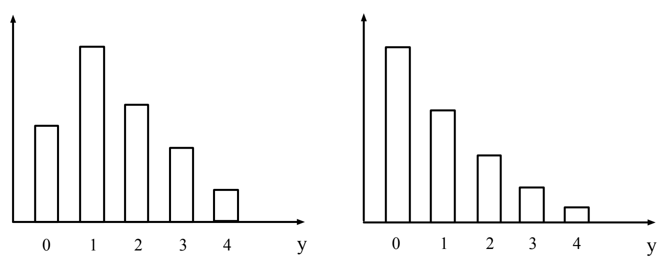
26

Growth in Factors Measured by Multiple Categorical Indicators



27

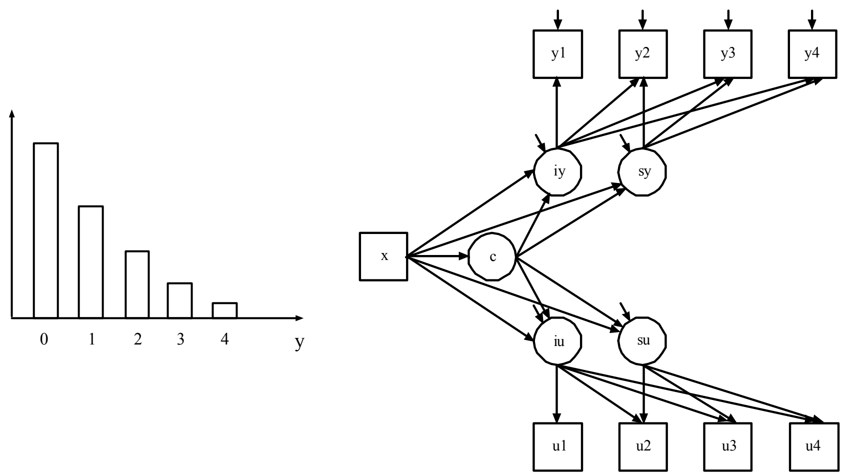
Modeling with a Preponderance of Zeros



- Outcomes: non-normal continuous – count – categorical
- Censored-normal modeling
- Two-part (semicontinuous modeling): Duan et al. (1983), Olsen & Shafer (2001)
- Mixture models, e.g. zero-inflated (mixture) Poisson (Roeder et al., 1999), censored-inflated, mover-stayer latent transition models, growth mixture models
- Onset (survival) followed by growth: Albert & Shih (2003)

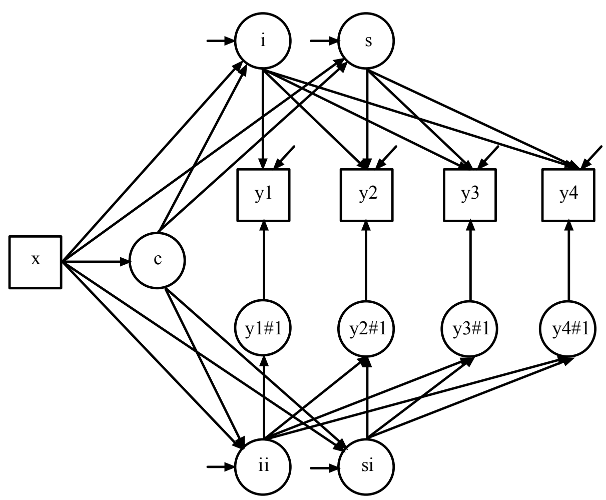
28

Two-Part (Semicontinuous) Growth Modeling



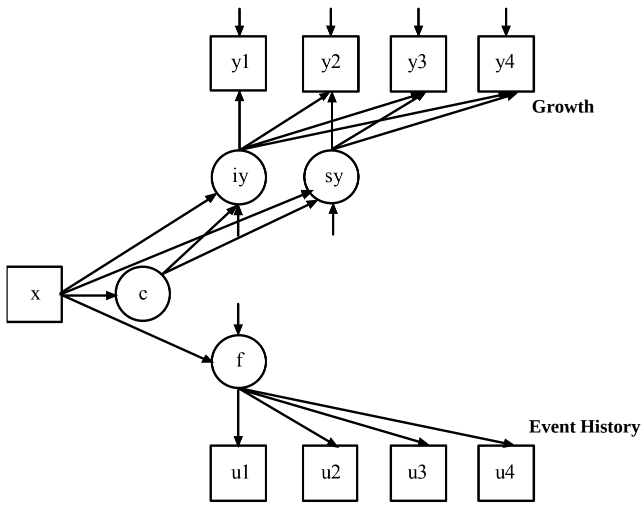
29

Inflated Growth Modeling (Two Classes At Each Time Point)



30

Onset (Survival) Followed by Growth



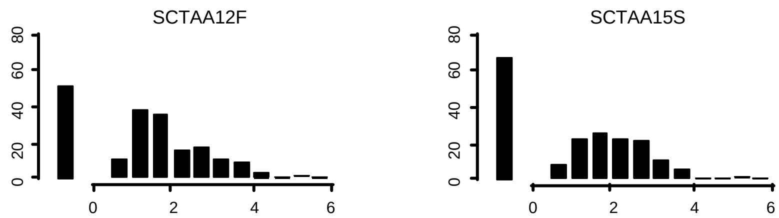
31

Two-Part Growth Modeling

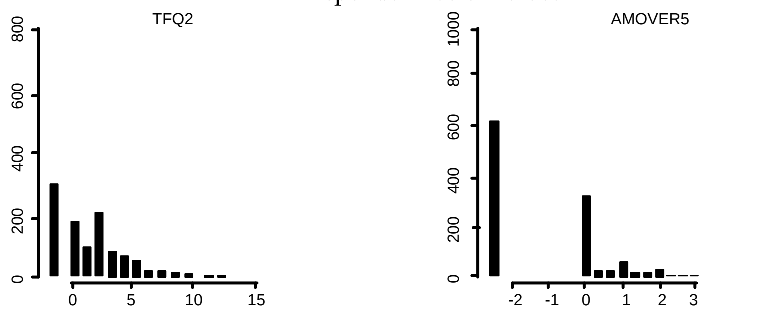
32

Two Types of Skewed Distributions

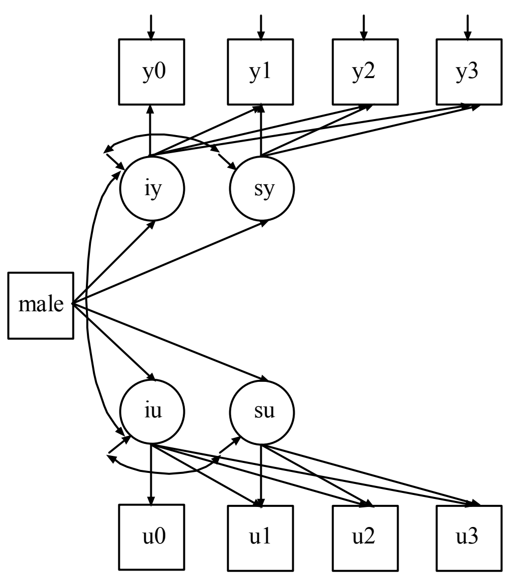
Normal mixture components



Preponderance of zeroes



33



34

Input For Step 1 Of A Two-Part Growth Model

```
TITLE:     step 1 of a two-part growth model
              Amover u    y
               >0    1   >0
                0    0   999
               999  999  999
DATA:      FILE = amp.dat;
VARIABLE:  NAMES ARE caseid
           amover0 ovrdrnk0 illdrnk0 vrydrn0
           amover1 ovrdrnk1 illdrnk1 vrydrn1
           amover2 ovrdrnk2 illdrnk2 vrydrn2
           amover3 ovrdrnk3 illdrnk3 vrydrn3
           amover4 ovrdrnk4 illdrnk4 vrydrn4
           amover5 ovrdrnk5 illdrnk5 vrydrn5
           amover6 ovrdrnk6 illdrnk6 vrydrn6
           tfq0-tfq6 v2 sex race livewith
           agedrnk0-agedrnk6 grades0-grades6;
           USEV = amover0 amover1 amover2 amover3
           sex race u0-u3 y0-y3;
           !MISSING = ALL (999);
```

35

Input For Step 1 Of A Two-Part Growth Model (Continued)

```
DEFINE:      u0 = 1;                  !binary part of variable
             IF(amover0 eq 0) THEN u0 = 0;
             IF(amover0 eq 999) THEN u0 = 999;
             y0 = amover0;            !continuous part of variable
             IF (amover0 eq 0) THEN y0 = 999;
             u1 = 1;
             IF(amover1 eq 0) THEN u1 = 0;
             IF(amover1 eq 999) THEN u1 = 999:
             y1 = amover1;
             IF(amover1 eq 0) THEN y1 = 999;
             u2 = 1;
             IF(amover2 eq 0) THEN u2 = 0;
             IF(amover2 eq 999) THEN u2 = 999;
             y2 = amover2;
             IF(amover2 eq 0) THEN y2 = 999;
             u3 = 1;
             IF(amover3 eq 0) THEN u3 = 0;
             IF(amover3 eq 999) THEN u3 = 999;
             y3 = amover3;
             IF(amover3 eq 0) THEN y3 = 999;
ANALYSIS:    TYPE = BASIC;
SAVEDATA:    FILE = ampyu.dat;
```

36

Output Excerpts Step 1 Of A Two-Part Growth Model

SAVEDATA Information

```
Order and format of variables

   AMOVER0  F10.3
   AMOVER1  F10.3
   AMOVER2  F10.3
   AMOVER3  F10.3
   SEX      F10.3
   RACE     F10.3
   U0       F10.3
   U1       F10.3
   U2       F10.3
   U3       F10.3
   Y0       F10.3
   Y1       F10.3
   Y2       F10.3
   Y3       F10.3

Save file
     ampyu.dat

Save file format
     14F10.3

Save file record length    1000
```

37

Input For Step 2 Of A Two-Part Growth Model

```
TITLE:      two-part growth model with linear growth for both
            parts
DATA:       FILE = ampya.dat;
VARIABLE:   NAMES = amover0-amover3 sex race u0-u3 y0-y3;
            USEV = u0-u3 y0-y3 male;
            USEOBS = u0 NE 999;
            MISSING = ALL (999);
            CATEGORICAL = u0-u3;
DEFINE:     Male = 2-sex;
```

38

Input For Step 2 Of A Two-Part Growth Model (Continued)

```
ANALYSIS:  TYPE = MISSING;
           ESTIMATOR = ML;
           ALGORITHM = INTEGRATION;
           COVERAGE = .09;
MODEL:     iu su | u0@0 u1@0.5 u2@1.5 u3@2.5;
           iy sy | y0@0 y1@0.5 y2@1.5 y3@2.5;
           iu-sy ON male;
           ! estimate the residual covariances
           ! iu with su, iy with sy, and iu with iy
           iu WITH sy@0;
           su WITH iy-sy@0;
OUTPUT:    PATTERNS SAMPSTAT STANDARDIZED TECH1 TECH4 TECH8;
PLOT:      TYPE = PLOT3;
           GROWTH = u0-u3(su) | y0-y3(sy);
```

39

Output Excerpts Step 2 Of A Two-Part Growth Model

Tests of Model Fit

```
Loglikelihood

          H0 Value                     -3277.101

Information Criteria

          Number of Free parameters           19
          Akaike (AIC)                  6592.202
          Bayesian (BIC)                6689.444
          Sample-Size Adjusted BIC      6629.092
             (n* = (n + 2) / 24)
```

40

Output Excerpts Step 2 Of A Two-Part Growth Model (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
IU |					
U0	1.000	0.000	0.000	2.839	0.843
U1	1.000	0.000	0.000	2.839	0.882
U2	1.000	0.000	0.000	2.839	0.926
U3	1.000	0.000	0.000	2.839	0.905
SU |					
U0	0.000	0.000	0.000	0.000	0.000
U1	0.500	0.000	0.000	0.416	0.129
U2	1.500	0.000	0.000	1.249	0.407
U3	2.500	0.000	0.000	2.082	0.664

41

Output Excerpts Step 2 Of A Two-Part Growth Model (Continued)

IY |					
Y0	1.000	0.000	0.000	0.534	0.787
Y1	1.000	0.000	0.000	0.534	0.738
Y2	1.000	0.000	0.000	0.534	0.740
Y3	1.000	0.000	0.000	0.534	0.644
SY |					
Y0	0.000	0.000	0.000	0.000	0.000
Y1	0.500	0.000	0.000	0.117	0.162
Y2	1.500	0.000	0.000	0.351	0.487
Y3	2.500	0.000	0.000	0.586	0.707

42

Output Excerpts Step 2 Of A Two-Part Growth Model (Continued)

```
IU       ON
   MALE          0.569   0.234      2.433    0.200    0.100
SU       ON
   MALE         -0.181   0.119     -1.518   -0.218   -0.109
IY       ON
   MALE          0.149   0.061      2.456    0.279    0.139
SY       ON
   MALE         -0.068   0.038     -1.790   -0.290   -0.145

IU       WITH
   SU           -1.144   0.326     -3.509   -0.484   -0.484
   IY            1.193   0.134      8.897    0.788    0.788
   SY            0.000   0.000      0.000    0.000    0.000
IY       WITH
   SY           -0.039   0.019     -2.109   -0.316   -0.316
SU       WITH
   IY            0.000   0.000      0.000    0.000    0.000
   SY            0.000   0.000      0.000    0.000    0.000
```

43

Output Excerpts Step 2 Of A Two-Part Growth Model (Continued)

```
Intercepts
   Y0                0.000   0.000    0.000     0.000     0.000
   Y1                0.000   0.000    0.000     0.000     0.000
   Y2                0.000   0.000    0.000     0.000     0.000
   Y3                0.000   0.000    0.000     0.000     0.000
   IU                0.000   0.000    0.000     0.000     0.000
   SU                0.855   0.098    8.716     1.027     1.027
   IY                0.232   0.059    3.901     0.435     0.435
   SY                0.240   0.031    7.830     1.025     1.025

Thresholds
   U0$1              2.655   0.206   12.877
   U1$1              2.655   0.206   12.877
   U2$1              2.655   0.206   12.877
   U3$1              2.655   0.206   12.877
```

44

Output Excerpts Step 2 Of A Two-Part Growth Model (Continued)

Residual Variances

Y0	0.175	0.032	5.470	0.175	0.380
Y1	0.266	0.029	9.159	0.266	0.509
Y2	0.238	0.027	8.810	0.238	0.457
Y3	0.269	0.054	5.014	0.269	0.392
IU	7.982	1.086	7.351	0.990	0.990
SU	0.685	0.202	3.400	0.988	0.988
IY	0.279	0.040	7.019	0.981	0.981
SY	0.054	0.017	3.224	0.979	0.979

45

Output Excerpts Step 2 Of A Two-Part Growth Model (Continued)

R-Square

Observed Variable	R-Square
U0	0.710
U1	0.682
U2	0.650
U3	0.666
Y0	0.620
Y1	0.491
Y2	0.543
Y3	0.608

Latent Variable	R-Square
IU	0.010
SU	0.012
IY	0.019
SY	0.021

46

Output Excerpts Step 2 Of A Two-Part Growth Model (Continued)

Technical 4 Output

ESTIMATED MEANS FOR THE LATENT VARIABLES

	IU	SU	IY	SY	MALE
1	0.305	0.758	0.312	0.204	0.536

ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES

	IU	SU	IY	SY	MALE
IU	8.062				
SU	-1.170	0.694			
IY	1.214	-0.007	0.285		
SY	-0.010	0.003	-0.042	0.055	
MALE	0.142	-0.045	0.037	-0.017	0.249

Output Excerpts Step 2 Of A Two-Part Growth Model (Continued)

ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES

	IU	SU	IY	SY	MALE
IU	1.000				
SU	-0.495	1.000			
IY	0.801	-0.015	1.000		
SY	-0.014	0.016	-0.336	1.000	
MALE	0.100	-0.109	0.139	-0.145	1.000

General Latent Variable Modeling
Using Mplus Version 3

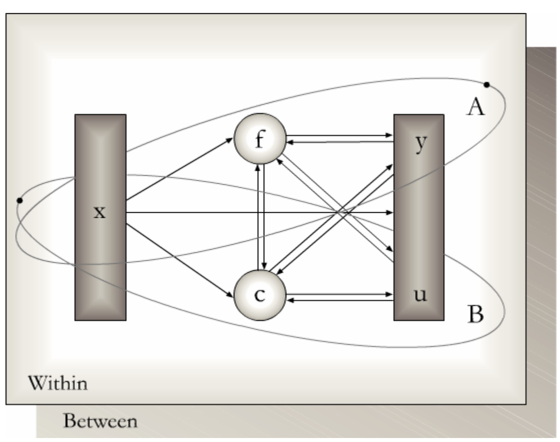
Block 3: Mixture Modeling

Bengt Muthén
bmuthen@ucla.edu

Mplus: www.statmodel.com

1

General Latent Variable Modeling Framework



2

Categorical Latent Variables

- Mixture regression
- Latent class analysis
- Latent transition analysis
- Missing data modeling

Categorical and Continuous Latent Variables

- SEMM
- Growth mixture modeling

3

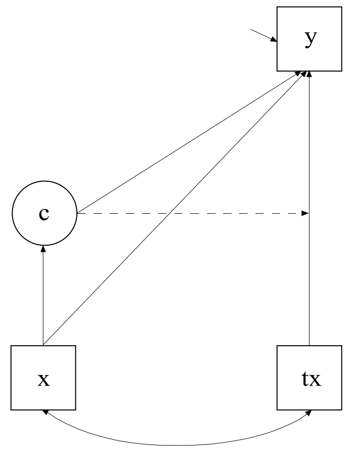
Randomized Preventive Interventions and Complier-Average Causal Effect Estimation (CACE)

- Angrist, Imbens & Rubin (1996)
- Yau & Little (1998, 2001)
- Jo (2002)
- Dunn et al. (2003)

- Compliance status observed for those invited for treatment
- Compliance status unobserved for controls

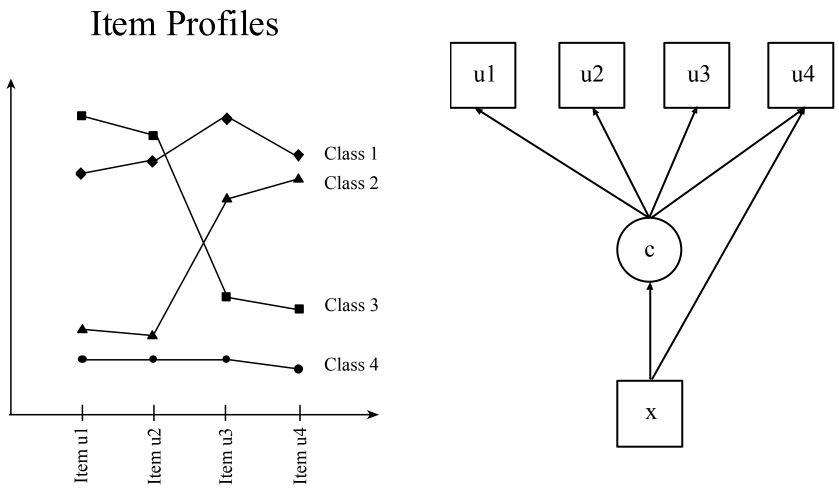
4

CACE Mixture Modeling



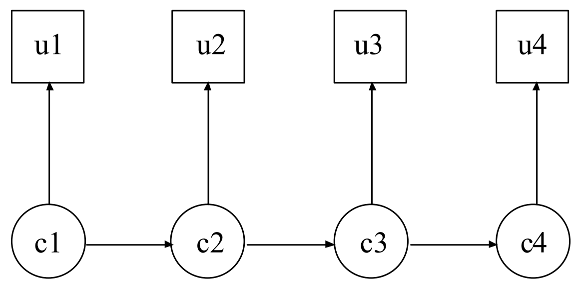
5

Latent Class Analysis



6

Hidden Markov Modeling



Latent Transition Analysis

Transition Probabilities

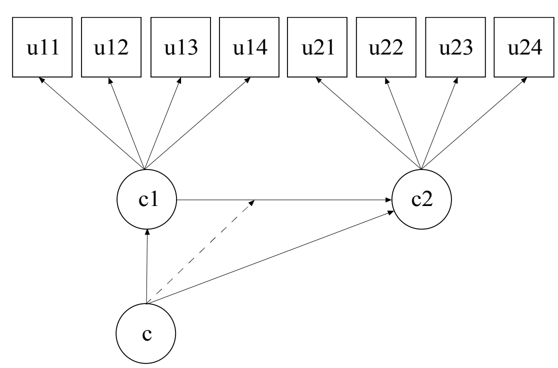
Mover Class (c=1)

		c2: 1	c2: 2
c1	1	0.6	0.4
c1	2	0.3	0.7

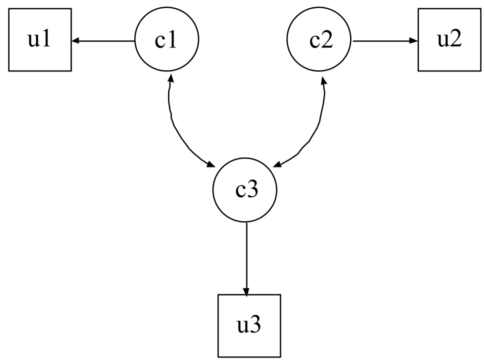
Stayer Class (c=2)

		c2: 1	c2: 2
c1	1	0.90	0.10
c1	2	0.05	0.95

Time Point 1 Time Point 2



Loglinear Modeling of Frequency Tables



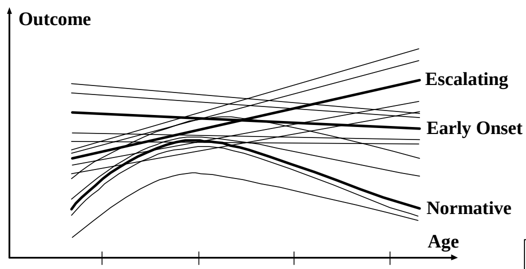
9

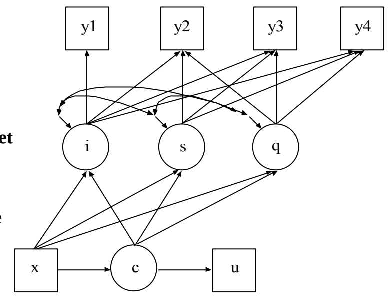
Growth Mixture Modeling

- Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. **Biometrics**, 55, 463-469.
- Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S., Carlin, J., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. **Biostatistics**, 3, 459-475.

10

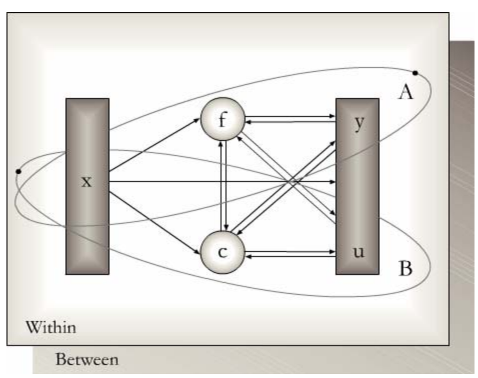
Growth Mixture Modeling





11

General Modeling Framework



x	background variables
y, u	continuous, censored, count, and categorical outcome variables
f	continuous latent variables
c	categorical latent variables

12

Summary Of Techniques Using Latent Classes

	Outcome/ Indicator Scale	Number of Timepoints	Number of Outcomes/ Timepoint	Within-Class Variation Standard	Within-Class Variation Mplus
LCA	u	Single	Multiple	No	Yes
LPA	y	Single	Multiple	No	Yes
LCFA	u, y	Single	Multiple	No	Yes
LCGA	u, y	Multiple	Single Multiple	No	Yes (GMM)
LTA	u, y	Multiple	Multiple	No	Yes

13

Summary Of Techniques Using Latent Classes (Continued)

	Outcome/ Indicator Scale	Number of Timepoints	Number of Outcomes/ Timepoint	Within-Class Variation Standard	Within-Class Variation Mplus
SEMM	u, y	Single	Multiple	Yes	Yes
GMM	u, y	Multiple	Single Multiple	Yes	Yes
GGMM	u, y	Multiple	Single Multiple	Yes	Yes
DTSMA	u	Multiple	Single Multiple	No	Yes
LLLCA	u, y	Single Multiple	Single Multiple	NA	Yes

14

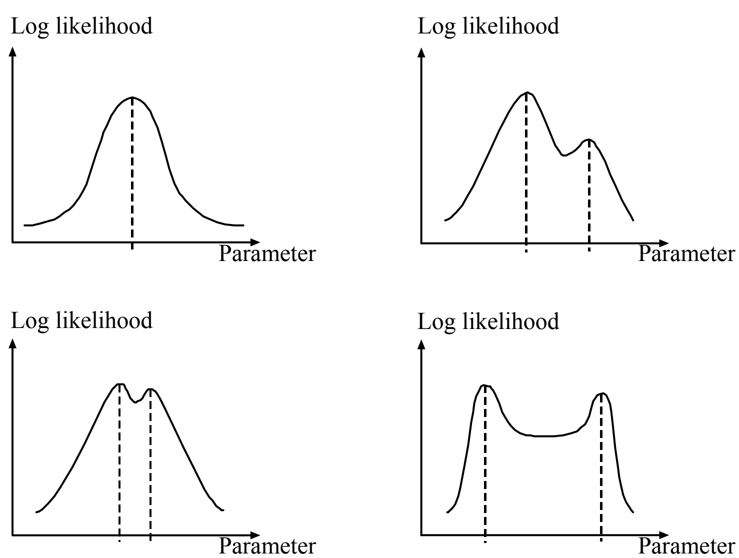
Summary Of Techniques Using Latent Classes (Continued)

LCA – Latent Class Analysis
LPA – Latent Profile Analysis
LCFA – Latent Class Factor Analysis
LCGA – Latent Class Growth Analysis
LTA – Latent Transition Analysis
SEMM – Structural Equation Mixture Modeling
GMM – Growth Mixture Modeling
GGMM – General Growth Mixture Modeling
DTSMA – Discrete-Time Survival Mixture Analysis
LLLCA – Loglinear Latent Class Analysis

u – categorical dependent variables
y – continuous, censored, count dependent variables

15

Global and Local Solutions



16

Random Starts in Version 3

When TYPE=MIXTURE is used, random sets of starting values are generated as the default for all parameters in the model except variances and covariances. These random sets of starting values are random perturbations of either user-specified starting values or default starting values produced by the program. Maximum likelihood optimization is done in two stages. In the initial stage, 10 random sets of starting values are generated. An optimization is carried out for ten iterations using each of the 10 random sets of starting values. The ending values from the optimization with the highest loglikelihood are used as the starting values in the final stage of optimization which is carried out using the default optimization settings for TYPE=MIXTURE. Random starts can be turned off or done more thoroughly.

Recommendations for a more thorough investigation of multiple solutions:

STARTS = 100 10;
or
STARTS = 500 10;
with
STITERATIONS = 20;

17

Antisocial Behavior (ASB) Data

The Antisocial Behavior (ASB) 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 17 antisocial behavior items that were collected in 1980 when respondents were between the ages of 16 and 23 and the background variables of age, gender and ethnicity. The ASB items assessed the frequency of various behaviors during the past year. A sample of 7,326 respondents has complete data on the antisocial behavior items and the background variables of age, gender, and ethnicity. Following is a list of the 17 items:

18

Antisocial Behavior (ASB) Data (Continued)

Damaged property	Use other drugs
Fighting	Sold marijuana
Shoplifting	Sold hard drugs
Stole < $50	"Con" someone
Stole > $50	Take auto
Use of force	Broken into building
Seriously threaten	Held stolen goods
Intent to injure	Gambling operation
Use marijuana	

19

Input For LCA Of 17 Antisocial Behavior (ASB) Items With Random Starts

```
TITLE:    LCA of 17 ASB items

DATA:     FILE IS asb.dat;
          FORMAT IS 34x 42f2;

VARIABLE: NAMES ARE property fight shoplift lt50 gt50 force
          threat injure pot drug soldpot solddrug con auto
          bldg goods gambling dsm1-dsm22 sex black hisp;

          USEVARIABLES ARE property-gambling;

          CLASSES = c(5);

          CATEGORICAL ARE property-gambling;

ANALYSIS: TYPE = MIXTURE;
          STARTS = 500 10;
          STITERATIONS = 20;

OUTPUT:   TECH8 TECH10 TECH11;

SAVEDATA: FILE IS asb.sav;
          SAVE IS CPROB;
```

20

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items With Random Starts

```
Loglikelihood values at local maxima and seeds:

          -40808.314   195353
          -40808.406   783165
          -40808.406   863691
          -40815.960   939709
          -40815.960   303634
          -40815.960   85734
          -40815.960   316165
          -40815.960   458181
          -40815.960   502532
          -40816.006   605161
```

21

Deciding On The Number Of Classes For The ASB Items

Number of Classes	Loglike-lihood	# par.	BIC	AIC	Entropy	LRT p-value for k-1
1	-48,168.475	17	96,488	96,370	NA	NA
2	-42,625.653	35	85,563	85,321	.838	.0000
3	-41,713.142	53	83,898	83,532	.743	.0000
4	-41,007.498	71	82,647	82,157	.742	.0000
5	-40,808.314	89	82,409	81,795	.741	.0000
6	-40,604.231	107	82,161	81,422	.723	.0019

22

Deciding On The Number Of Classes For The ASB Items (Continued)

Four-Class Solution

```
FINAL CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE BASED ON
ESTIMATED POSTERIOR PROBABILITIES
   Class 1         672.41667       0.09178  High
   Class 2        1354.73100       0.18492  Drug
   Class 3        1821.71706       0.24866  Person Offense
   Class 4        3477.13527       0.47463  Normative (Pot)
```

Five-Class Solution

```
FINAL CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE BASED ON
ESTIMATED POSTERIOR PROBABILITIES              Comparison To
                                               Four-Class
                                               Solution
   Class 1         138.06985       0.01888  High
   Class 2         860.41897       0.11771  Property Offense
   Class 3        1257.56652       0.17151  Drug
   Class 4        1909.32749       0.26219  Person Offense
   Class 5        3160.61717       0.42971  Normative (Pot)
```

Six-Class Solution – adds a variation on Class 2 in the 5-class solution

23

Input For LCA Of 17 Antisocial Behavior (ASB) Items

```
TITLE:     LCA of 17 ASB items

DATA:      FILE IS asb.dat;
           FORMAT IS 34x 42f2;

VARIABLE:  NAMES ARE property fight shoplift lt50 gt50 force
           threat injure pot drug soldpot solddrug con auto
           bldg goods gambling dsm1-dsm22  sex black hisp;

           USEVARIABLES ARE property-gambling;

           CLASSES = c(4);

           CATEGORICAL ARE property-gambling;
```

24

Input For LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
ANALYSIS: TYPE = MIXTURE;

MODEL:                                 !Not needed in Version 3
          %OVERALL%                    !Not needed in Version 3
          %c#1%                        !Not needed in Version 3
          [property$1-gambling$1*0];  !Not needed in Version 3
          %c#2%                        !Not needed in Version 3
          [property$1-gambling$1*1];  !Not needed in Version 3
          %c#3%                        !Not needed in Version 3
          [property$1-gambling$1*2];  !Not needed in Version 3
          %c#4%                        !Not needed in Version 3
          [property$1-gambling$1*3];  !Not needed in Version 3

OUTPUT:   TECH8 TECH10 TECH11;

SAVEDATA: FILE IS asb.sav;
          SAVE IS CPROB;
```

25

Output Excerpts For LCA Of 17 Antisocial Behavior (ASB) Items

Tests of Model Fit

```
Loglikelihood

          H0 Value                      -41007.498

Information Criteria

          Number of Free parameters          71
          Akaike (AIC)                 82156.996
          Bayesian (BIC)               82646.838
          Sample-Size Adjusted BIC     82421.215
             (n* = (n + 2) / 24)
          Entropy                          0.742
```

26

Output Excerpts For LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
Chi-Square Test of Model Fit for the Latent Class Indicator Model
Part**

          Pearson Chi-Square

          Value                         20827.381
          Degrees of freedom               130834
          P-Value                          1.0000

          Likelihood Ratio Chi-Square

          Value                          6426.411
          Degrees of Freedom               130834
          P-Value                          1.0000


**Of the 131072 cells in the latent class indicator table, 166
were deleted in the calculation of chi-square due to extreme
values.
```

27

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

Classification Information

```
FINAL CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE
   Class 1         672.41594       0.09178
   Class 2        1354.72999       0.18492
   Class 3        1821.73064       0.24867
   Class 4        3477.12344       0.47463

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY CLASS
MEMBERSHIP

Class Counts and Proportions
   Class 1               664       0.09064
   Class 2              1237       0.16885
   Class 3              1772       0.24188
   Class 4              3653       0.49863

Average Class Probabilities by Class
                  1        2       3      4
   Class 1    0.896   0.057   0.046  0.000
   Class 2    0.032   0.835   0.090  0.043
   Class 3    0.021   0.072   0.803  0.104
   Class 4    0.000   0.043   0.070  0.887
```

28

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

Number of Classes

TECHNICAL 10

UNIVARIATE MODEL FIT INFORMATION

Estimated Probabilities

Variable	H1	H0	Residual
PROPERTY			
Category 1	0.815	0.815	0.000
Category 2	0.185	0.185	0.000
FIGHT			
Category 1	0.719	0.719	0.000
Category 2	0.281	0.281	0.000
SHOPLIFT			
Category 1	0.736	0.736	0.000
Category 2	0.264	0.264	0.000

29

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

BIVARIATE MODEL FIT INFORMATION

Estimated Probabilities

Variable	Variable	HI	H0	Residual
PROPERTY	FIGHT			
Category 1	Category 1	0.635	0.631	0.004
Category 1	Category 2	0.180	0.184	-0.004
Category 2	Category 1	0.084	0.088	-0.004
Category 2	Category 2	0.101	0.097	0.004
PROPERTY	SHOPLIFT			
Category 1	Category 1	0.656	0.646	0.010
Category 1	Category 2	0.159	0.169	-0.010
Category 2	Category 1	0.080	0.090	-0.010
Category 2	Category 2	0.105	0.095	0.010

30

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
TECHNICAL 11- check that the H0 loglikelihood value is the same
as the k-1 class H0 loglikelihood value to be certain a local
solution has not been reached.

VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 3 (H0) VERSUS 4
CLASSES

         H0 Loglikelihood Value                         -41713.142
         2 Times the Loglikelihood Difference             1411.288
         Difference in the Number of Parameters                 19
         Mean                                               -0.960
         Standard Deviation                                  43.222
         P-Value                                            0.0000

      LO-MENDELL-RUBIN ADJUSTED LRT TEST
         Value                                            1402.991
         P-Value                                            0.0000
```

31

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
Class 1
Thresholds
   PROPERTY$1      -1.267    0.142    -8.911
   FIGHT$1         -1.047    0.117    -8.972
   SHOPLIFT$1      -1.491    0.125   -11.927
   LT50$1          -0.839    0.114    -7.377
   GT50$1           0.523    0.117     4.477
   FORCE$1          1.027    0.113     9.113
   THREAT$1        -1.495    0.125   -11.996
   INJURE$1         0.394    0.096     4.125
   POT1            -2.220    0.193   -11.496
   DRUG$1          -0.394    0.122    -3.234
   SOLDPOT$1       -0.053    0.116    -0.455
   SOLDDRUG$1       1.784    0.135    13.233
   CONS$1          -0.585    0.109    -5.388
   AUTO$1           0.591    0.102     5.796
   BLDG$1           0.290    0.112     2.591
   GOODS$1         -0.697    0.112    -5.699
   GAMBLING$1       1.722    0.125    13.774
```

32

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
Class 2
Thresholds
   PROPERTY$1       1.533     0.113   13.550
   FIGHT$1          1.403     0.118   11.857
   SHOPLIFT$1       0.310     0.083    3.755
   LT50$1           0.988     0.085   11.561
   GT50$1           3.543     0.218   16.252
   FORCE$1          4.058     0.319   12.718
   THREAT$1         0.499     0.097    5.153
   INJURE$1         2.462     0.165   14.881
   POT1            -3.232     0.311  -10.403
   DRUG$1          -0.336     0.118   -2.853
   SOLDPOT$1        1.033     0.109    9.457
   SOLDDRUG$1       3.189     0.180   17.691
   CONS$1           1.386     0.093   14.918
   AUTO$1           2.473     0.144   17.195
   BLDG$1           3.381     0.223   15.186
   GOODS$1          2.167     0.148   14.632
   GAMBLING$1       4.078     0.269   15.158
```


Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
Class 3
Thresholds
   PROPERTY$1       0.962     0.104    9.267
   FIGHT$1         -0.134     0.089   -1.508
   SHOPLIFT$1       0.780     0.096    8.084
   LT50$1           1.350     0.108   12.470
   GT50$1           3.360     0.197   17.067
   FORCE$1          2.456     0.116   21.213
   THREAT$1        -0.747     0.105   -7.131
   INJURE$1         1.465     0.102   14.420
   POT1             0.567     0.088    6.467
   DRUG$1           3.649     0.298   12.258
   SOLDPOT$1        5.393     0.737    7.320
   SOLDDRUG$1       6.263     0.752    8.325
   CONS$1           0.508     0.079    6.467
   AUTO$1           2.121     0.102   20.809
   BLDG$1           3.100     0.193   16.099
   GOODS$1          1.969     0.130   15.122
   GAMBLING$1       3.514     0.182   19.260
```

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

Class 4

Thresholds

PROPERTY$1	3.687	0.176	20.891
FIGHT$1	2.281	0.107	21.345
SHOPLIFT$1	2.609	0.114	22.923
LT50$1	3.046	0.119	25.566
GT50$1	5.796	0.403	14.386
FORCE$1	5.276	0.343	15.395
THREAT$1	2.171	0.136	15.985
INJURE$1	5.765	0.664	8.682
POT1	1.290	0.065	19.888
DRUG$1	4.430	0.305	14.502
SOLDPOT$1	6.367	0.589	10.801
SOLDDRUG$1	6.499	0.573	11.342
CONS$1	2.525	0.106	23.928
AUTO$1	4.314	0.208	20.784
BLDG$1	6.741	0.739	9.120
GOODS$1	5.880	0.611	9.627
GAMBLING$1	6.816	0.954	7.144

35

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

LATENT CLASS INDICATOR MODEL PART IN PROBABILITY SCALE

Class 3

PROPERTY			
Category 2	0.277	0.021	13.321
FIGHT			
Category 2	0.533	0.022	24.193
SHOPLIFT			
Category 2	0.314	0.021	15.120
LT50			
Category 2	0.206	0.018	11.635
GT50			
Category 2	0.034	0.006	5.256
FORCE			
Category 2	0.079	0.008	9.379
THREAT			
Category 2	0.678	0.023	29.703
INJURE			
Category 2	0.188	0.015	12.118

36

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
POT
   Category 2              0.362          0.020       17.887
DRUG
   Category 2              0.025          0.007        3.447
SOLDPOT
   Category 2              0.005          0.003        1.364
SOLDDRUG
   Category 2              0.002          0.001        1.332
CON
   Category 2              0.376          0.018       20.388
AUTO
   Category 2              0.107          0.010       10.989
BLDG
   Category 2              0.043          0.008        5.427
GOODS
   Category 2              0.122          0.014        8.752
GAMBLING
   Category 2              0.029          0.005        5.645
```

37

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

Technical 8 Output

```
E STEP ITER  LOGLIKELIHOOD     ABS CHANGE REL CHANGE   CLASS COUNTS


      1   -0.50814249D+05      0.0000000  0.0000000    888.234   1659.562
                                                      2208.576   2569.628
      2   -0.41810482D+05   9003.7666995  0.1771898    831.905   1722.366
                                                      2165.174   2606.555
      3   -0.41706620D+05    103.8616123  0.0024841    767.588   1807.168
                                                      2146.235   2605.009
      4   -0.41657122D+05     49.4986699  0.0011868    714.379   1867.660
                                                      2146.792   2597.170
      5   -0.41623995D+05     33.1269450  0.0007952    671.621   1905.257
                                                      2162.382   2586.740
```

38

Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
 96   -0.41007499D+05  0.0002095  0.0000000    673.025  1354.398
                                              1824.820  3473.758
 97   -0.41007499D+05  0.0001814  0.0000000    672.982  1354.419
                                              1824.606  3473.993
 98   -0.41007499D+05  0.0001572  0.0000000    672.943  1354.439
                                              1824.408  3474.211
 99   -0.41007499D+05  0.0001362  0.0000000    672.906  1354.457
                                              1824.222  3474.414
100   -0.41007499D+05  0.0001180  0.0000000    672.872  1354.475
                                              1824.050  3474.604
```


Output Excerpts LCA Of 17 Antisocial Behavior (ASB) Items (Continued)

```
153  -0.41007498D+05    0.0000001  0.0000000    672.424  1354.725
                                               1821.771  3477.081
154  -0.41007498D+05    0.0000001  0.0000000    672.423  1354.726
                                               1821.767  3477.085
155  -0.41007498D+05    0.0000000  0.0000000    672.422  1354.726
                                               1821.764  3477.088

171  -0.41007498D+05    0.0000000  0.0000000    672.416  1354.730
                                               1821.733  3477.121
172  -0.41007498D+05    0.0000000  0.0000000    672.416  1354.730
                                               1821.732  3477.122
173  -0.41007498D+05    0.0000000  0.0000000    672.416  1354.730
                                               1821.731  3477.123
```

Class Probability Excerpts LCA Of 17 Antisocial Behavior (ASB) Items

Saved Data And Posterior Class Probabilities

```
0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.
.000        .001        .013        .987          4.000
1.  0.  0.  1.  0.  0.  0.  0.  1.  1.  1.  0.  0.  0.  0.  0.  0.
.005        .995        .000        .000          2.000
0.  1.  0.  0.  0.  0.  1.  0.  0.  0.  0.  0.  1.  1.  0.  1.  0.
.003        .001        .996        .000          3.000
0.  0.  0.  0.  0.  0.  1.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.
.000        .004        .191        .805          4.000
0.  1.  0.  0.  0.  0.  1.  0.  1.  0.  0.  0.  0.  0.  1.  0.  0.
.004        .121        .871        .004          3.000
```

41

Class Table 1. LCA and EFA for Antisocial Behavior (n=7326)

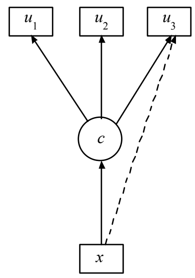
	LCA Solution Categorical Factors				EFA Solution Continuous Factors		
	C#1	C#2	C#3	C#4	Factor 1	Factor 2	Factor 3
Property	**0.78**	0.18	0.28	0.02	**0.65**	0.19	-0.04
Fighting	0.74	0.20	**0.53**	0.09	0.19	**0.60**	-0.13
Shoplifting	**0.82**	0.42	0.31	0.07	**0.61**	-0.03	0.18
Stole < $50	**0.70**	0.27	0.21	0.05	**0.85**	-0.21	0.05
Stole > $50	**0.37**	0.03	0.03	0.00	**0.81**	0.00	0.01
Use of force	0.26	0.02	0.08	0.01	0.34	0.37	-0.01
Seriously threaten	0.82	0.38	**0.68**	0.10	-0.11	**0.89**	0.03
Intent to injure	0.40	0.08	0.19	0.00	-0.11	**0.83**	0.08
Use marijuana	0.90	**0.96**	0.36	**0.22**	-0.02	0.00	**0.88**
Use other drugs	0.60	**0.58**	0.03	0.01	0.01	-0.02	**0.88**
Sold marijuana	0.51	0.26	0.01	0.00	0.15	0.07	**0.74**
Sold hard drugs	0.14	0.04	0.00	0.00	0.19	0.09	**0.59**
"Con" someone	0.64	0.20	**0.38**	0.07	**0.43**	0.25	-0.07
Take auto	0.36	0.08	0.11	0.01	**0.45**	0.15	0.07
Broken into bldg.	**0.43**	0.03	0.04	0.00	**0.80**	0.03	0.01
Held stolen goods	**0.67**	0.10	0.12	0.00	**0.69**	0.11	0.06
Gambling operation	0.15	0.02	0.03	0.00	0.28	0.36	0.08
Class Prob.	0.09	0.18	0.25	0.47			

42

LCA with Covariates

Dichotomous indicators u: $u_1, u_2, \ldots, u_r$. Categorical latent variable c: $c = k$; $k = 1, 2, \ldots, K$. Marginal probability for item $u_j = 1$,

$$P(u_j = 1) = \sum_{k=1}^{K} P(c = k) \; P(u_j = 1 \mid c = k). \tag{5}$$



With a covariate x, consider $P(u_j = 1 \mid c = k, x)$, $P(c = k \mid x)$,

$$logit\,[P(u_j = 1 \mid c = k, x)] = \lambda_{jk} + \kappa_j\, x, \tag{6}$$

$$logit\,[P(c = k \mid x)] = \alpha_k + \gamma_k\, x. \tag{7}$$

43

Multinomial Logistic Regression Of c On x

The multinomial logistic regression model expresses the probability that individual i falls in class k of the latent class variable c as a function of the covariate x,

$$P(c_i = k \mid x_i) = \frac{e^{\alpha_k + \gamma_k\, x_i}}{\sum_{s=1}^{K} e^{\alpha_s + \gamma_s\, x_i}}, \tag{87}$$

where $\alpha_K = 0$, $\gamma_K = 0$ so that $e^{\alpha_K + \gamma_K\, x_i} = 1$.

This implies that the log odds comparing class k to the last class K is

$$log[P(c_i = k \mid x) \,/\, P(c_i = K \mid x)] = \alpha_k + \gamma_k\, x_i. \tag{88}$$

44

Input For LCA Of 9 Antisocial Behavior (ASB Items With Covariates)

```
TITLE:    LCA of 9 ASB items with three covariates

DATA:     FILE IS asb.dat;
          FORMAT IS 34x 51f2;

VARIABLE: NAMES ARE property fight shoplift lt50 gt50 force
          threat injure pot drug soldpot solddrug con auto
          bldg goods gambling dsm1-dsm22 male black hisp
single
          divorce dropout college onset f1 f2 f3 age94;

          USEVARIABLES ARE property fight shoplift lt50
          threat pot drug con goods age94 male black;

          CLASSES = c(4);

          CATEGORICAL ARE property-goods;

ANALYSIS: TYPE = MIXTURE;
```


Input For LCA Of 9 Antisocial Behavior (ASB Items With Covariates) (Continued)

```
MODEL:
          %OVERALL%

          c#1-c#3 ON age94 male black;

          %c#1%                          !Not needed in Version 3
          [property$1-gambling$1*0];     !Not needed in Version 3

          %c#2%                          !Not needed in Version 3
          [property$1-gambling$1*1];     !Not needed in Version 3

          %c#3%                          !Not needed in Version 3
          [property$1-gambling$1*2];     !Not needed in Version 3

          %c#4%                          !Not needed in Version 3
          [property$1-gambling$1*3];     !Not needed in Version 3

OUTPUT:   TECH8;
```

Output Excerpts LCA Of 9 Antisocial Behavior (ASB) Items With Covariates

Tests of Model Fit

```
Loglikelihood

          H0 Value                  -30416.942
Information Criteria

          Number of Free parameters          48
          Akaike (AIC)                60929.884
          Bayesian (BIC)              61261.045
          Sample-Size Adjusted BIC    61108.512
             (n* = (n + 2) / 24)
          Entropy                         0.690
```

47

Output Excerpts LCA Of 9 Antisocial Behavior (ASB) Items With Covariates (Continued)

```
LATENT CLASS REGRESSION MODEL PART

C#1        ON
   AGE94              -.285          .028        -10.045
   MALE               2.578          .151         17.086
   BLACK               .158          .139          1.141
C#2        ON
   AGE94               .069          .022          3.182
   MALE                .187          .110          1.702
   BLACK              -.606          .139         -4.357
C#3        ON
   AGE94              -.317          .028        -11.311
   MALE               1.459          .101         14.431
   BLACK               .999          .117          8.513
Intercepts
   C#1               -1.822          .174        -10.485
   C#2                -.748          .103         -7.258
   C#3                -.324          .125         -2.600
```

48

Output Excerpts LCA Of 9 Antisocial Behavior (ASB) Items With Covariates (Continued)

Classification Information

```
FINAL CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE
   Class 1         928.40043       0.12673
   Class 2        1499.08913       0.20463
   Class 3        2249.50562       0.30706
   Class 4        2649.00482       0.36159
CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY CLASS
MEMBERSHIP
Class Counts and Proportions
   Class 1              920       0.12558
   Class 2             1433       0.19560
   Class 3             2154       0.29402
   Class 4             2819       0.38479
Average Class Probabilities by Class
                 1        2       3      4
   Class 1    0.859   0.065   0.076  0.000
   Class 2    0.047   0.808   0.087  0.058
   Class 3    0.033   0.067   0.816  0.084
   Class 4    0.000   0.048   0.105  0.846
```


Output Excerpts LCA Of 9 Antisocial Behavior (ASB) Items With Covariates (Continued)

```
LATENT CLASS INDICATOR MODEL PART
Class 1
   PROPERTY$1          -1.185      0.133     -8.883
   FIGHT$1             -0.967      0.108     -8.958
   SHOPLIFT$1          -1.307      0.130    -10.078
   LT50$1              -0.647      0.108     -5.998
   THREAT$1            -1.383      0.108    -12.847
   POT$1               -1.656      0.151    -10.935
   DRUG$1               0.093      0.106      0.878
   CON$1               -0.384      0.091     -4.201
   GOODS$1             -0.299      0.106     -2.816
Class 2
   PROPERTY$1           1.834      0.126     14.583
   FIGHT$1              1.700      0.144     11.769
   SHOPLIFT$1           0.425      0.079      5.418
   LT50$1               1.113      0.082     13.504
   THREAT$1             0.549      0.095      5.782
   POT$1               -2.561      0.246    -10.415
   DRUG$1              -0.127      0.107     -1.185
   CON$1                1.346      0.092     14.595
   GOODS$1              2.272      0.137     16.592
```

Output Excerpts LCA Of 9 Antisocial Behavior (ASB) Items With Covariates (Continued)

```
LATENT CLASS INDICATOR MODEL PART
Class 3
   PROPERTY$1          1.494      0.110     13.632
   FIGHT$1             0.032      0.085      0.378
   SHOPLIFT$1          1.312      0.100     13.083
   LT50$1              1.929      0.122     15.858
   THREAT$1           -0.184      0.090     -2.049
   POT$1               0.916      0.097      9.484
   DRUG$1              4.484      0.480      9.345
   CON$1               0.937      0.074     12.626
   GOODS$1             2.668      0.152     17.502
Class 4
   PROPERTY$1          4.699      0.433     10.840
   FIGHT$1             3.988      0.512      7.787
   SHOPLIFT$1          2.943      0.153     19.222
   LT50$1              3.192      0.158     20.193
   THREAT$1            2.929      0.219     13.360
   POT$1               1.443      0.082     17.624
   DRUG$1              5.236      0.637      8.225
   CON$1               2.814      0.145     19.402
   GOODS$1             7.307      1.901      3.844
```

51

Calculating Latent Class Probabilities For Different Covariate Values

For a class,

logit = intercept + b1*age94 + b2*male + b3*black

Example 1: For age94 = 0, male = 0, black = 0
where age94 = 0 is age 16
male = 1 is female
black = 0 is not black

	exp	probability (exp/sum)
logitc1 = -1.822	0.162	0.069
logitc2 = -0.748	0.473	0.201
logitc3 = -0.324	0.723	0.307
logitc4 = 0	1.0	0.424
sum	2.358	1.001

52

Calculating Latent Class Probabilities For Different Covariate Values (Continued)

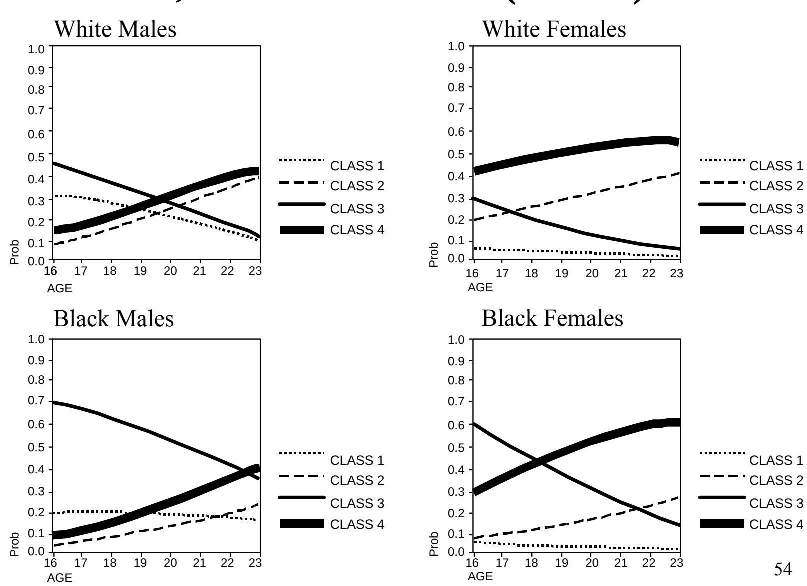
Example 2: For age94 = 1, male = 1, black = 1
where age94 = 1 is age 17
male = 1 is male
black = 1 is black

logitc1 = -1.822 + (-0.285*1) + (2.578*1) + (0.158*1)
= 0.629
logitc2 = -0.748 + 0.069*1 +0.187*1 + (-0.606*1)
= -1.098
logitc3 = -0.324 + (-0.317*1) + 1.459*1 + 0.999*1
= 1.817

	exp	probability (exp/sum)
logitc1 = 0.629	1.876	0.200
logitc2 = -1.098	0.334	0.036
logitc3 = 1.817	6.153	0.657
logitc4 = 0	1.0	0.107
sum	9.363	1.000

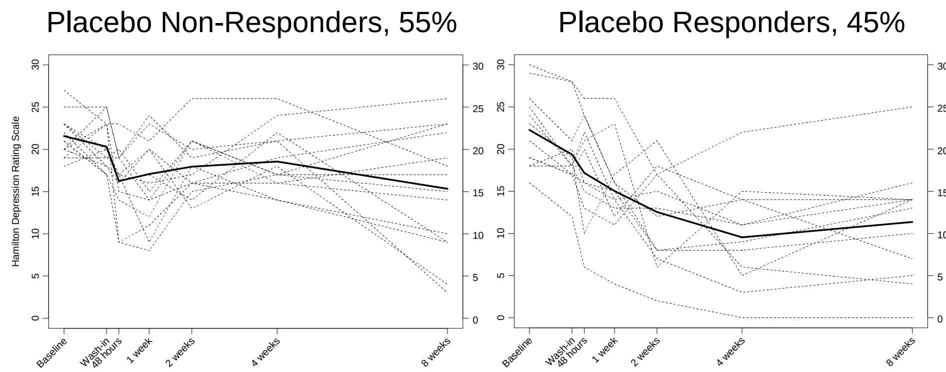
53

ASB Classes Regressed on Age, Male, Black in the NLSY (n=7326)



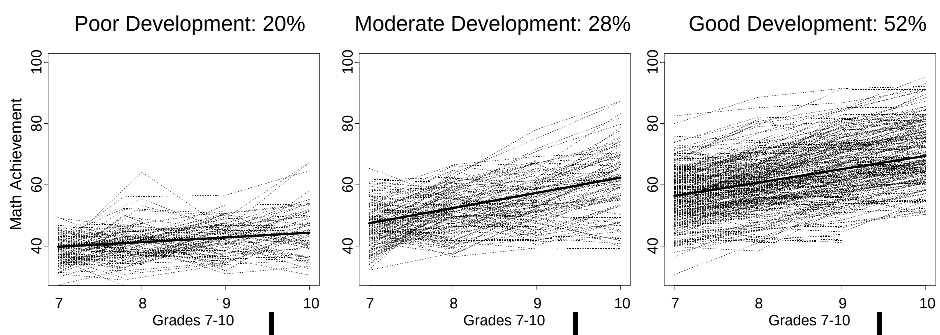
54

A Clinical Trial of Depression Medication: 2-Class Growth Mixture Modeling



55

Growth Mixture Modeling: LSAY Math Achievement Trajectory Classes and the Prediction of High School Dropout



Dropout:	69%	8%	1%

56

General Latent Variable Modeling Using Mplus Version 3

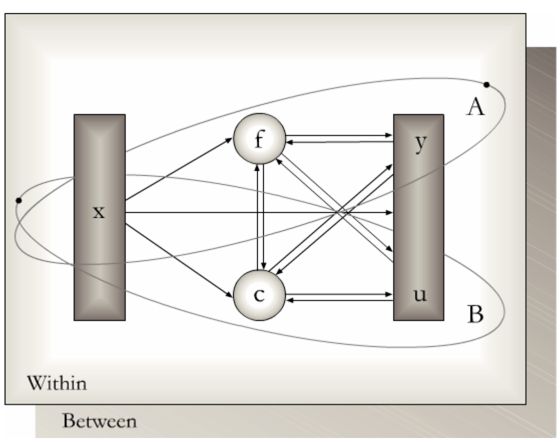
Block 4: Multilevel Modeling

Bengt Muthén
bmuthen@ucla.edu

Mplus: www.statmodel.com

1

General Latent Variable Modeling Framework



2

Multilevel Modeling with Continuous and Categorical Latent Variables

- Multilevel regression
- Multilevel CFA, SEM
- Multilevel growth modeling
- Multilevel discrete-time survival analysis

- Multilevel regression mixture analysis (CACE)
- Multilevel latent class analysis
- Multilevel growth mixture modeling

3

Multilevel Regression Analysis With Random Intercepts And Random Slopes In Multilevel Terms

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

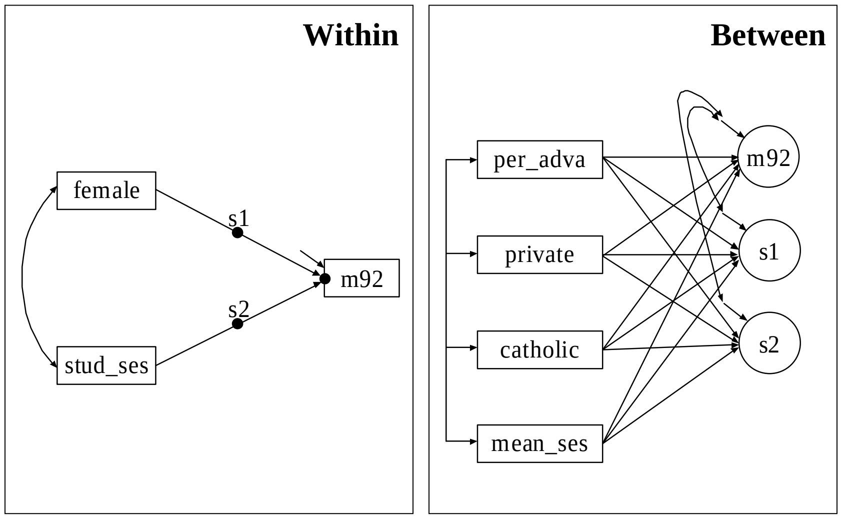
y_{ij} : individual-level outcome variable
x_{ij} : individual-level covariate
w_j : cluster-level covariate

Random intercepts, random slopes:

Level 1 (*Within*) : $y_{ij} = \beta_{0j} + \beta_{1j}\, x_{ij} + r_{ij}$, (8)
Level 2 (*Between*) : $\beta_{0j} = \gamma_{00} + \gamma_{01}\, w_j + u_{0j}$, (9)
Level 2 (*Between*) : $\beta_{1j} = \gamma_{10} + \gamma_{11}\, w_j + u_{1j}$, (10)

- Mplus gives the same estimates as HLM/MLwiN ML (not REML): $V(r)$ (residual variance for level 1), γ_{00}, γ_{01}, γ_{10}, γ_{11}, $V(u_0)$, $V(u_1)$, $Cov(u_0, u_1)$
- Centering of x: subtracting grand mean or group (cluster) mean
- Model testing with varying covariance structure, marginal covariance matrix for y

4



5

Input For Multilevel Regression Model

```
TITLE:    multilevel 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_teac 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 = school;
          CLUSTER = school;
          CENTERING = GRANDMEAN (stud_ses);
```

6

Input For Multilevel Regression Model

```
ANALYSIS: TYPE = TWOLEVEL RANDOM MISSING;

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

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

OUTPUT:   TECH8 SAMPSTAT;
```

7

Output Excerpts For Multilevel Regression Model

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   7400  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
```

8

Output Excerpts For Multilevel Regression Model

```
          3459    917  58687  81919  37741  63302  63143
   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
```

9

Output Excerpts For Multilevel Regression Model (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.
Within Level
Residual
Variances
   M92               70.577      1.149    61.442
Between Level
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
```

10

Output Excerpts For Multilevel Regression Model

```
 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                      54.886        0.428      128.231
    S1                       -0.856        0.507       -1.688
    S2                        4.075        0.309       13.208
  Residual Variances
    M92                       8.679        1.003        8.649
    S1                        5.740        1.411        4.066
    S2                        0.307        0.527        0.583
```

11

Random Slopes

- In single-level modeling random slopes β_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

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

- In two-level modeling random slopes β_j describe variation across clusters *j*

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

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

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

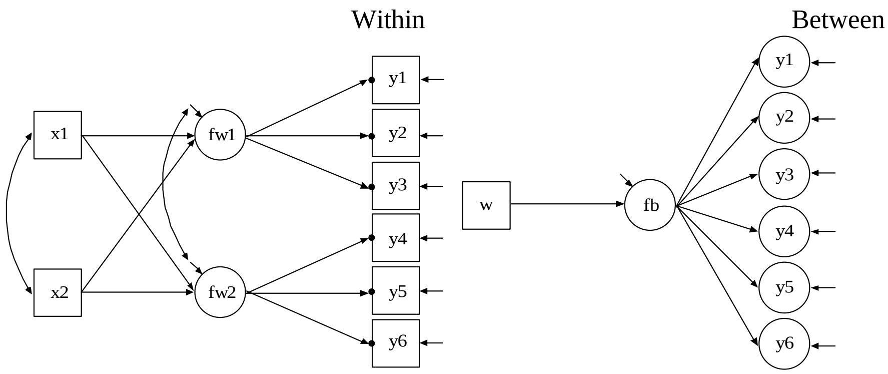
A small variance for a random slope typically leads to slow convergence of the ML-EM iterations. This suggests respecifying the slope as fixed.

Mplus allows random slopes for predictors that are

- Observed covariates
- Observed dependent variables (Version 3)
- Continuous latent variables (Version 3)

12

Two-Level Factor Analysis with Covariates



13

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)

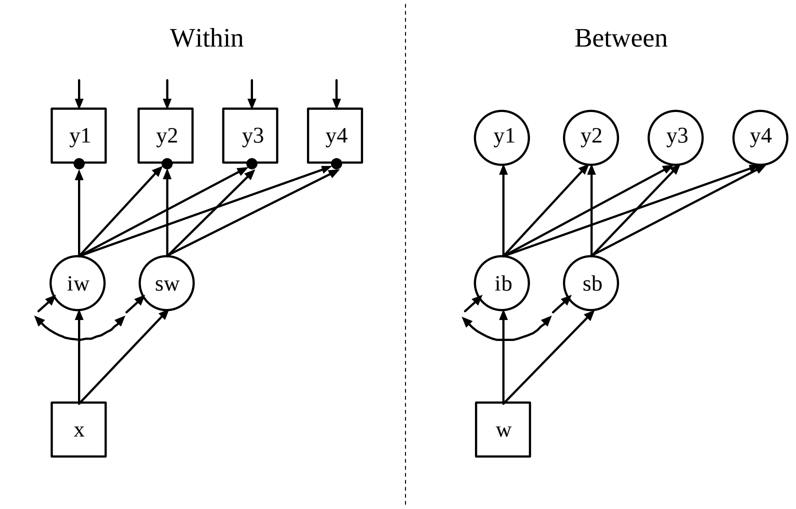
$$\textit{Level } 1 \ (\textit{Within}) \ : \ y_{tij} = \pi_{0ij} + \pi_{1ij} \, a_{1tij} + \pi_{2tij} \, a_{2tij} + e_{tij} \, , \qquad (1)$$

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

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

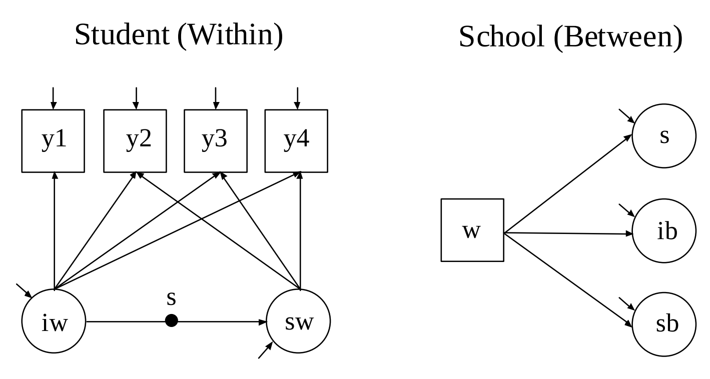
14

Two-Level Growth Modeling (3-Level Modeling)



15

Multilevel Modeling with a Random Slope for Latent Variables



16

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.

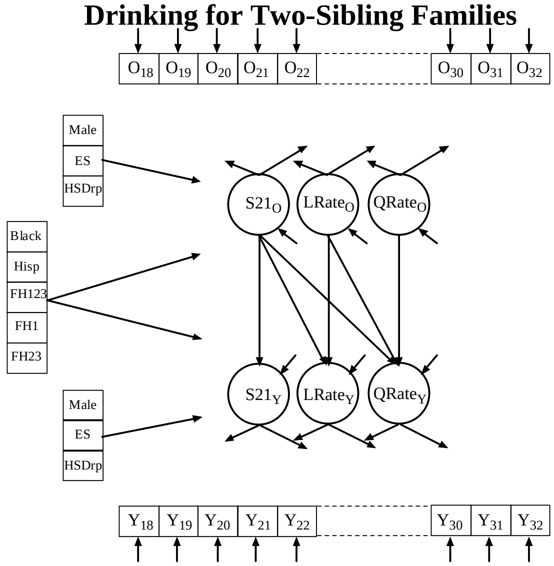
- used in the latent variable growth modeling, where the cluster units are the repeated measures over time

- 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)

17

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

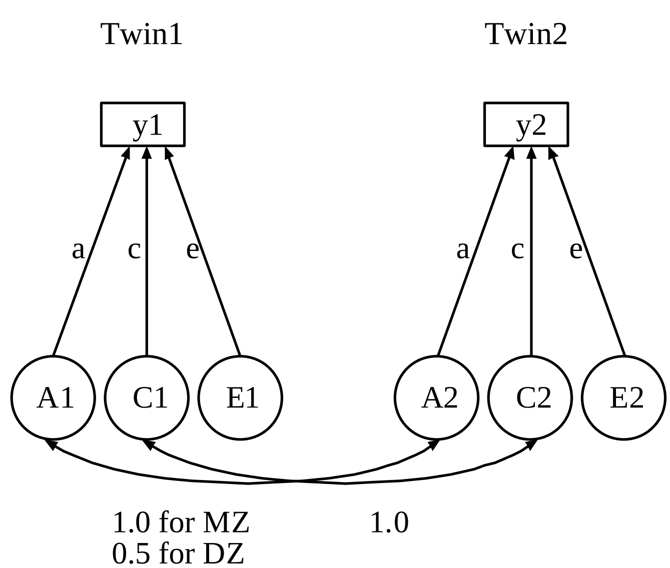


Source: Khoo, S.T. & Muthen, B. (2000). Longitudinal data on families: Growth modeling alternatives. Multivariate Applications in Substance Use Research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds.), Hillsdale, N.J.: Erlbaum, pp. 43-78.

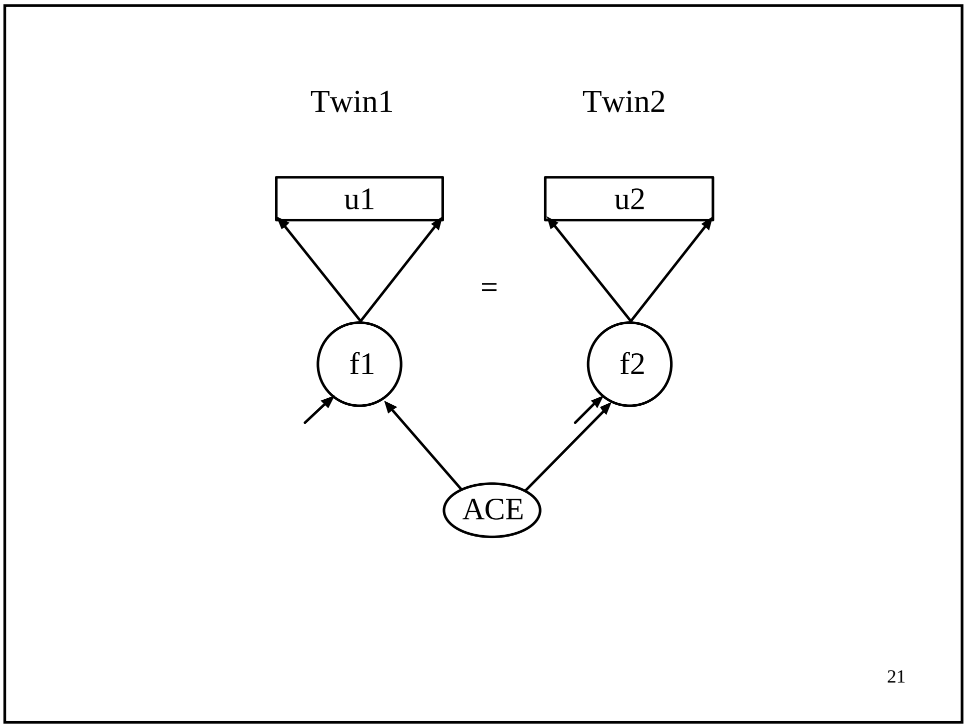
18

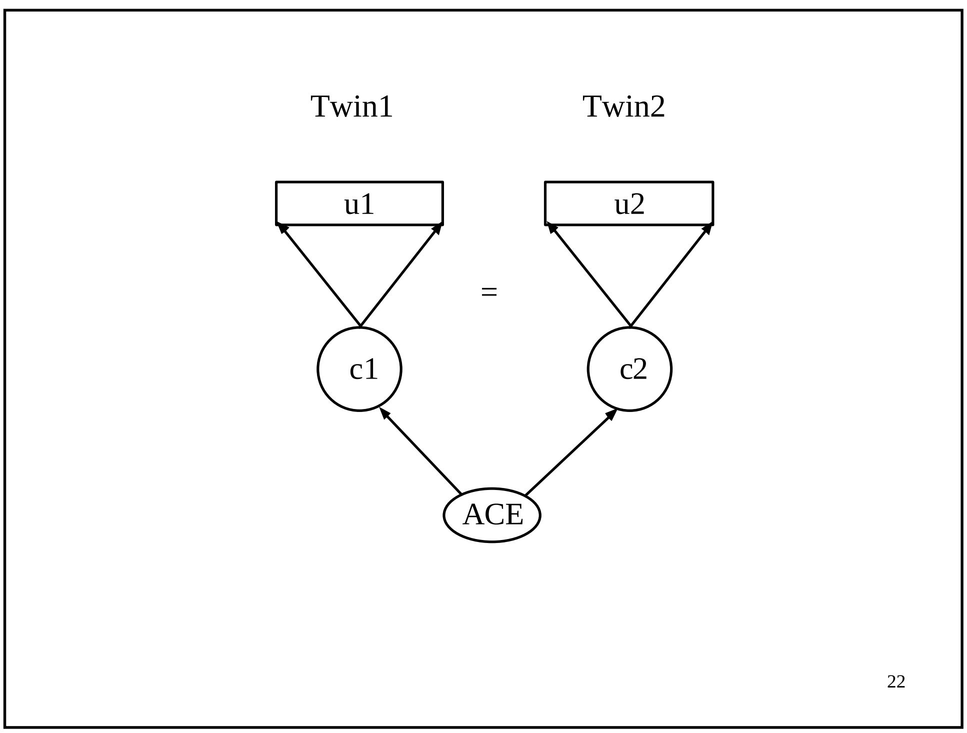
Twin Modeling

19

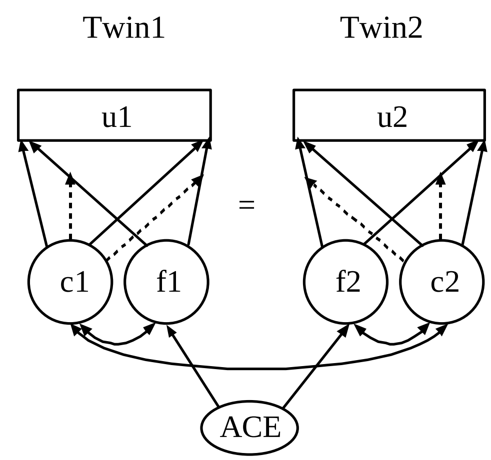


20

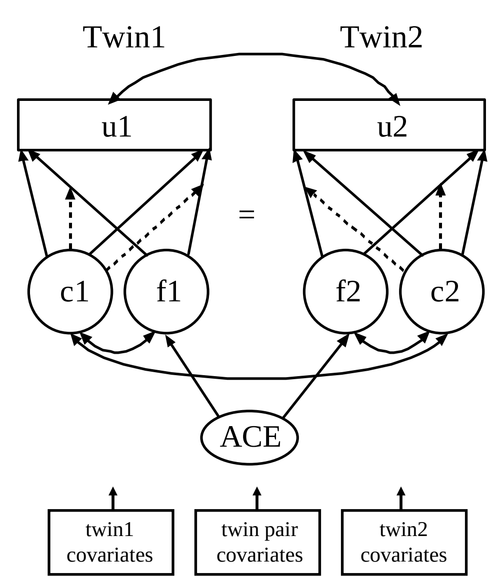

Twin1
Twin2
u1
u2
=
f1
f2
ACE
21


Twin1
Twin2
u1
u2
=
c1
c2
ACE
22

Hybrid Model (Severity LCA or Three-Part Modeling



23

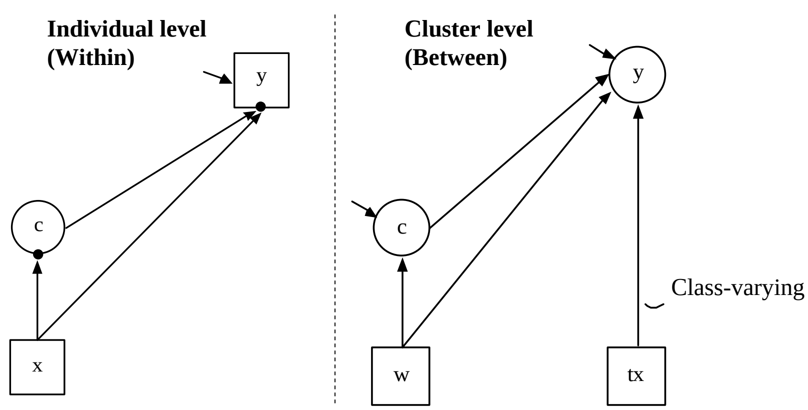


24

Multilevel Mixture Modeling

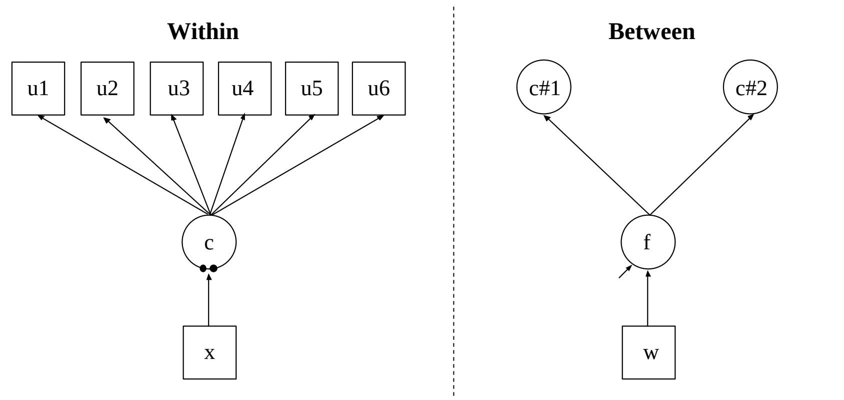
25

Two-Level CACE Mixture Modeling



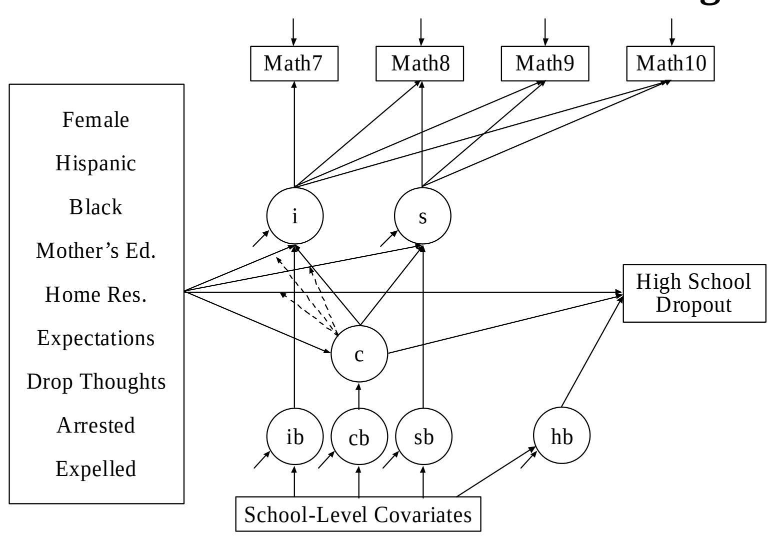
26

Two-Level Latent Class Analysis



27

Multilevel Growth Mixture Modeling



28

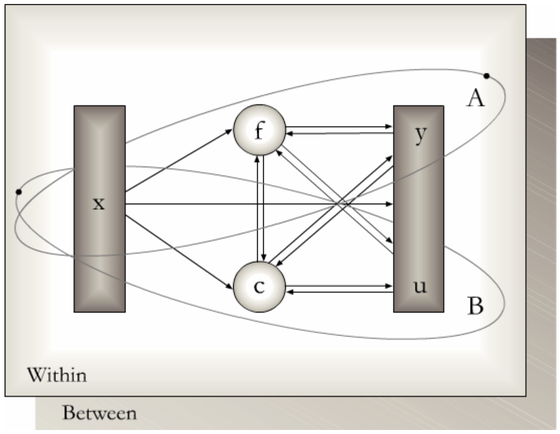
Monte Carlo Simulations in Mplus

- Data generation, analysis, and results summaries across replications
- Studies of tests of model fit, parameter estimation, standard errors, coverage, and power as a function of model variations, parameter values, and sample size
- Model Population, Model Missing, Model for analysis
- Full modeling framework available: continuous and categorical latent variables, multilevel data, different types of outcomes

References

- See the Mplus web site www.statmodel.com

General Latent Variable Modeling Framework



31