Statistical Analysis with Latent Variables

Educ 231E (M231E), Spring 2004

Bengt Muthen, UCLA
bmuthen@ucla.edu

Statistical Analysis with Latent Variables: Logistics

- UCLA lectures: 20 lectures through June 14
- UCLA lab sessions: Evening computer exercises once a week (TA: Karen Nylund)
- Video conferencing: off-campus sites
- Streaming video on the web from UCLA

Web Addresses

- **Course web site:** http://www.gseis.ucla.edu/faculty/muthen/courses.htm
- **Streaming video web site:** http://www.ats.ucla.edu/stat/seminars/default.htm
- **Mplus web site:** http://www.statmodel.com

Statistical Analysis with Latent Variables ED231E, Spring 2004 Syllabus

WEEK 1 (April 5 & 7)
- **Lecture 1:** Overview of course content. A general latent variable modeling framework
- **Lecture 2:** Confirmatory factor analysis

WEEK 2 (April 12 & 14)
- **Lecture 3:** Multiple-group confirmatory factor analysis
- **Lecture 4:** Structural equation modeling

WEEK 3 (April 19 & 21)
- **Lecture 5:** Introductory growth modeling
- **Lecture 6:** Growth modeling, cont'd

WEEK 4 (April 26 & 28)
- **Lecture 7:** Growth modeling, cont'd
- **Lecture 8:** Growth modeling, cont'd

WEEK 5 (May 3 & 5)
- **Lecture 9:** Introduction to modeling with categorical dependent variables
- **Lecture 10:** Modeling with a preponderance of zeros (zero inflation)

WEEK 6 (May 10 & 12)
- **Lecture 11:** Discrete-time survival analysis
- **Lecture 12:** Discrete-time survival analysis

WEEK 7 (May 17 & May 19)
- **Lecture 13:** Cross-sectional mixture modeling - LCA
- **Lecture 14:** Cross-sectional mixture modeling - LCRA

WEEK 8 (May 24 & 26)
- **Lecture 15:** Longitudinal mixture modeling – LTA
- **Lecture 16:** Longitudinal mixture modeling - GMM

WEEK 9 (June 2) May 31 cancelled due to Memorial Day
- **Lecture 17:** Latent variable modeling with missing data

WEEK 10 (June 7 & 9)
- **Lecture 18:** Multilevel latent variable modeling
- **Lecture 19:** Multilevel latent variable modeling cont'd

FINAL's WEEK (June 14)
- **Lecture 20:** Multilevel mixture modeling

Statistical Analysis with Latent Variables: An Example

- Commonalities of biometric and psychometric themes:
  – Random effects
  – Latent group (class) membership
  – Missing data
  – Multilevel data
  – Measurement modeling

BIOMETRICS 59, 897–906
December 2003

Modeling Tumor Growth with Random Onset

Paul S. Albert* and Joanna H. Shih
Biometric Research Branch, National Cancer Institute, Executive Plaza North,
Room 8136, Bethesda, Maryland 20892-7434, U.S.A.
**email:* AlbertP@ctep.NCI.NIH.GOV

SUMMARY. The longitudinal assessment of tumor volume is commonly used as an endpoint in small animal studies in cancer research. Groups of genetically identical mice are injected with mutant cells from clones developed with different mutations. The interest is on comparing tumor onset (i.e., the time of tumor detection) and tumor growth after onset, between mutation groups. This article proposes a class of linear and nonlinear growth models for jointly modeling tumor onset and growth in this situation. Our approach allows for interval-censored time of onset and missing-at-random dropout due to early sacrifice, which are common situations in animal research. We show that our approach has good small-sample properties for testing and is robust to some key unverifiable modeling assumptions. We illustrate this methodology with an application examining the effect of different mutations on tumorigenesis.

KEY WORDS: Animal studies; Discrete survival; Gompertzian growth; Linear mixed models; Nonlinear mixed models; Shared random effect.

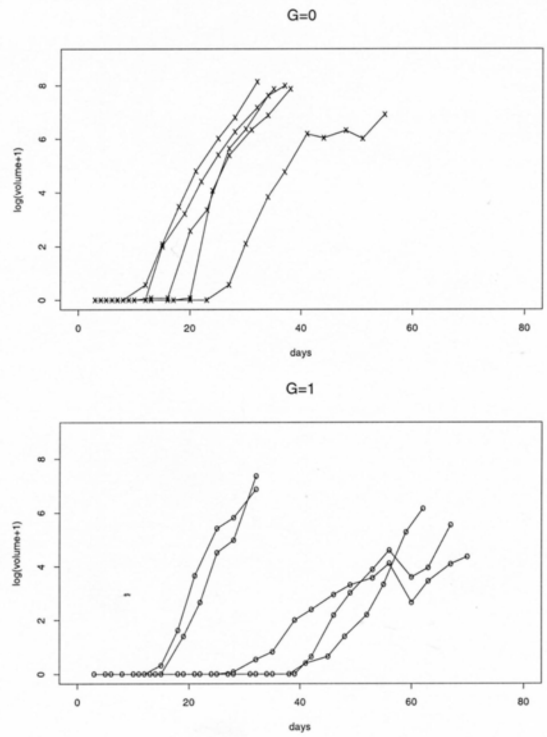


Figure 1. Clone-specific average growth (on log(volume + 1) scale) for Δ1310 group (G = 0) and for $\Delta 1310/S1283A$ group (G = 1).

Statistical Analysis with Latent Variables
A General Modeling Framework

Statistical Concepts Captured by Latent Variables

- Continuous Latent Variables
  - Measurement errors
  - Factors
  - Random effects
  - Variance components
  - Missing data
- Categorical Latent Variables
  - Clusters
  - Latent classes
  - Finite mixtures
  - Missing data

Models That Use Latent Variables

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

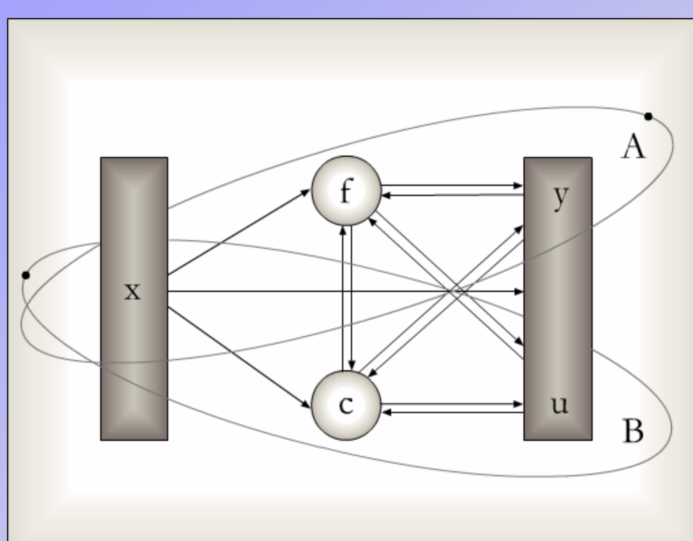
- Latent class models
- Mixture models
- Discrete-time survival models
- Missing data models

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

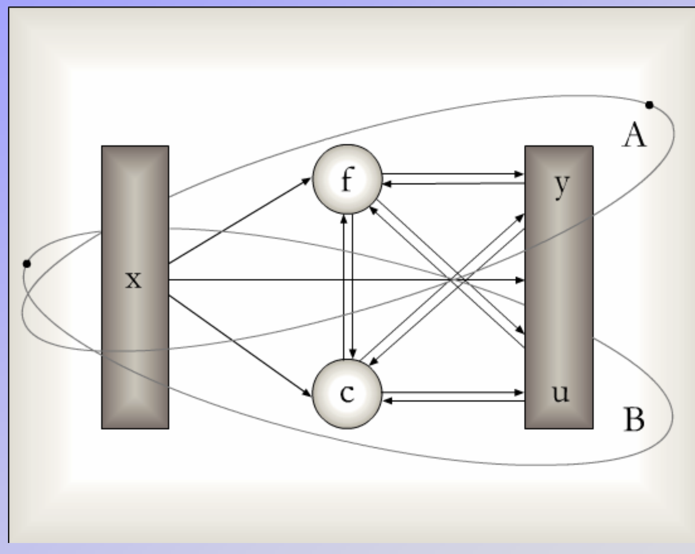
General Latent Variable Modeling Framework

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

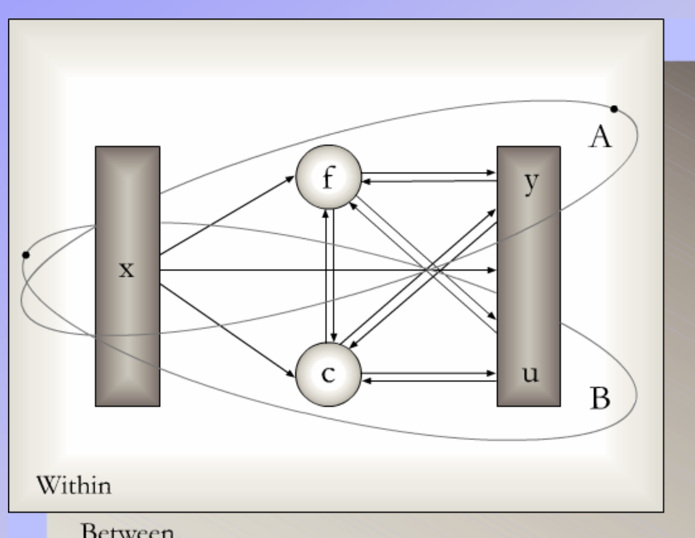
General Latent Variable Modeling Framework



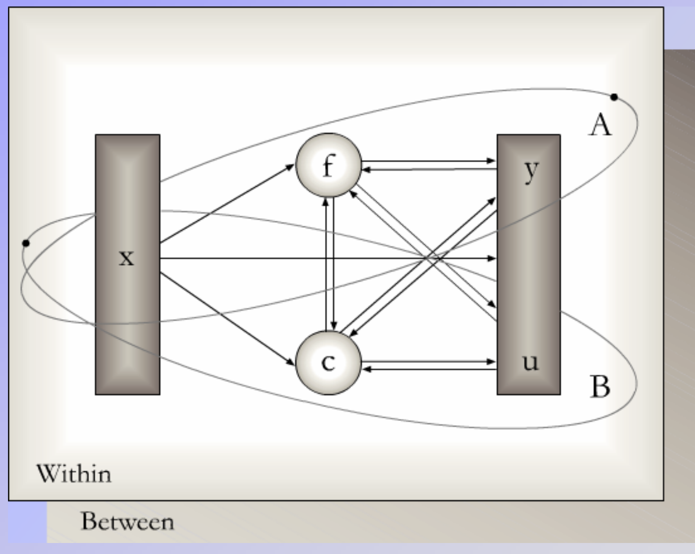
General Latent Variable Modeling Framework



General Latent Variable Modeling Framework



General Latent Variable Modeling Framework

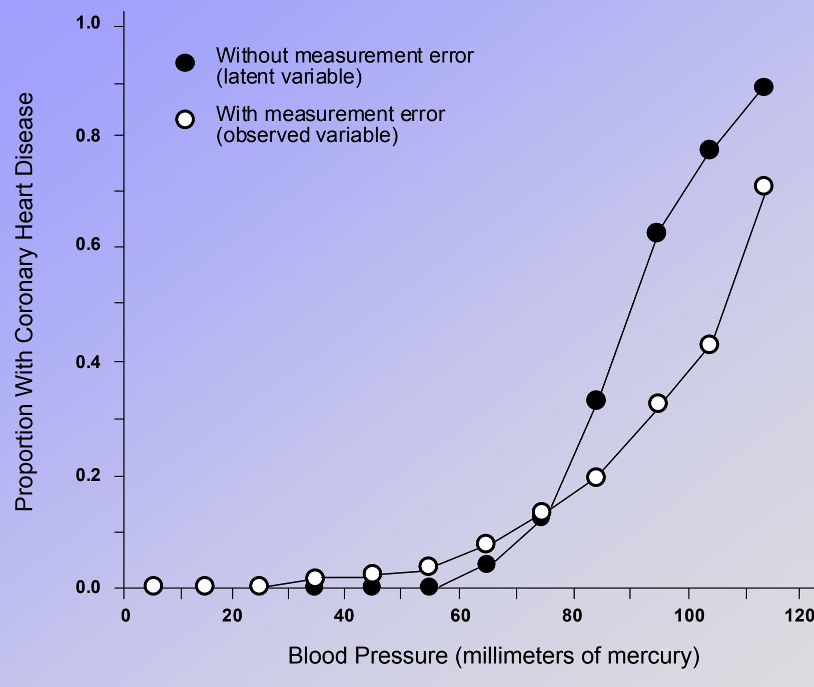


Continuous Latent Variables: Two Examples

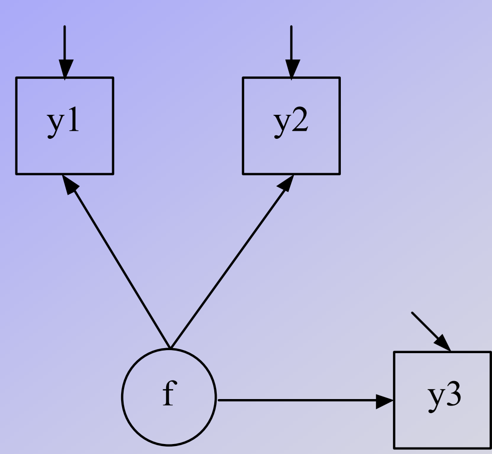
- Muthen (1992). Latent variable modeling in epidemiology. Alcohol Health & Research World, 16, 286-292
  - Blood pressure predicting coronary heart disease

- 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

Measurement Error in a Covariate



Measurement Error in a Covariate



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

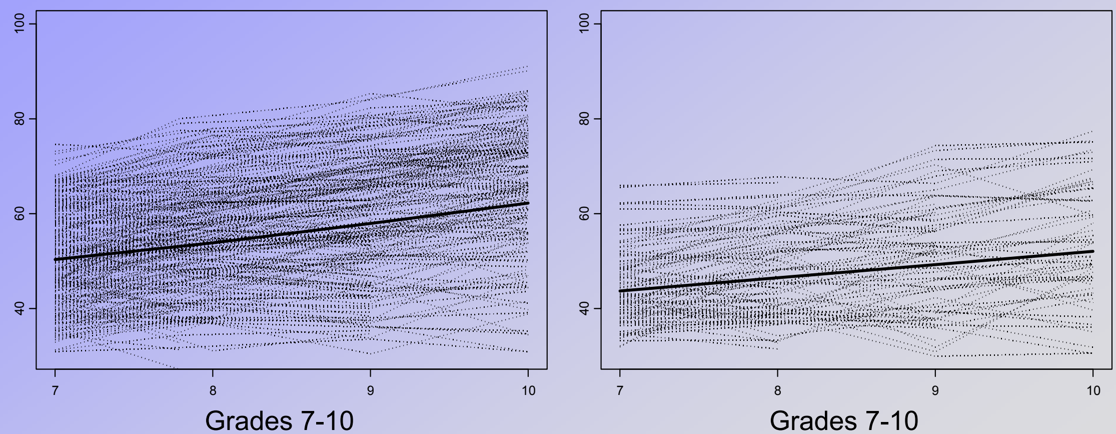
Growth Modeling of LSAY Math Achievement with Random Slopes for Time-Varying Covariates

- **Data source**: LSAY, n = 2,271 students in public schools
  - Clustering of students within schools ignored in this analysis
- **Outcome**: grade 7, 8, 9, 10 math achievement
- **Time-invariant covariates**: female, mother's education, home resources
- **Time-varying covariates**: 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-calculus; 9 = calculus)

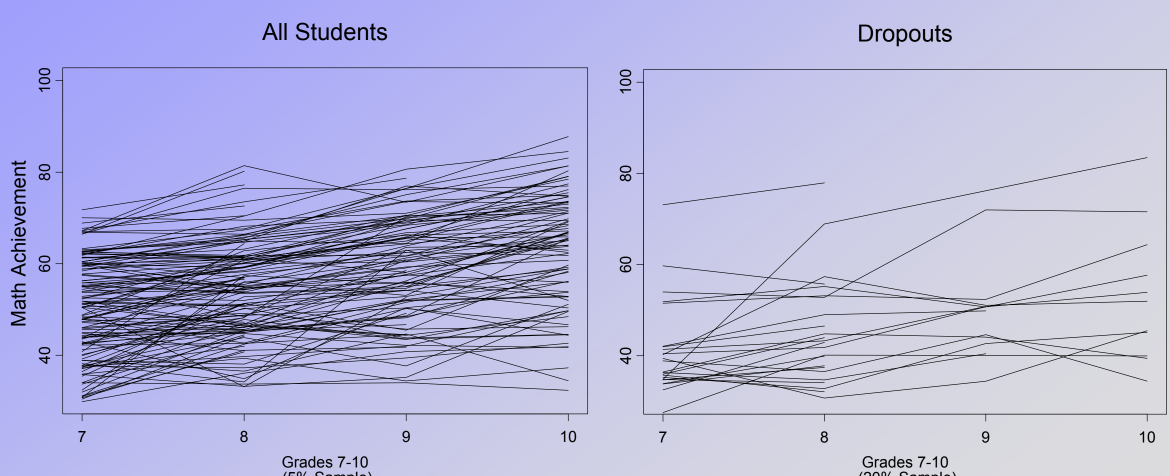
LSAY Math Achievement in Grades 7 – 10

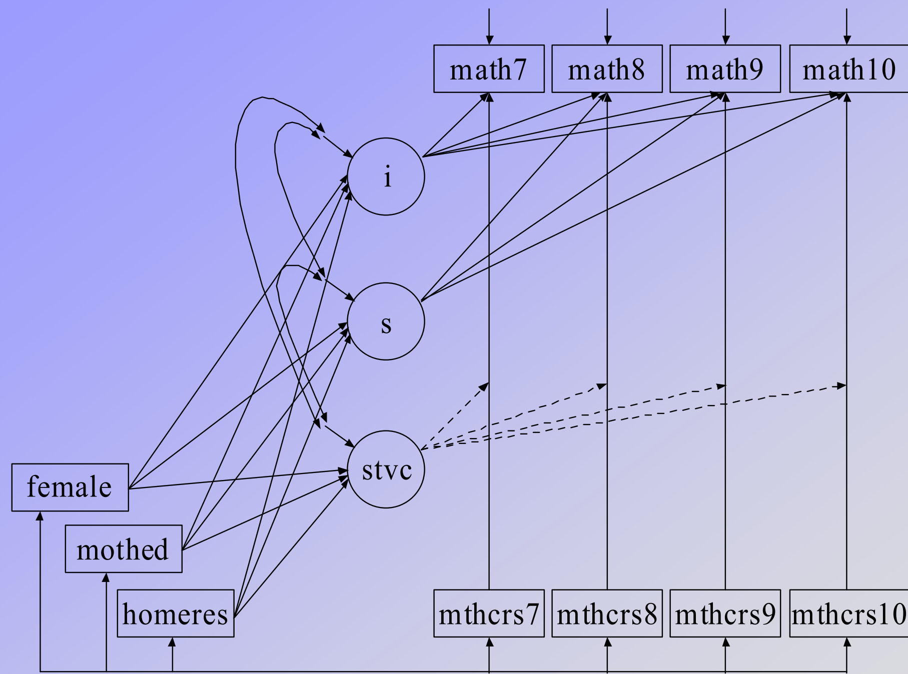
All Students

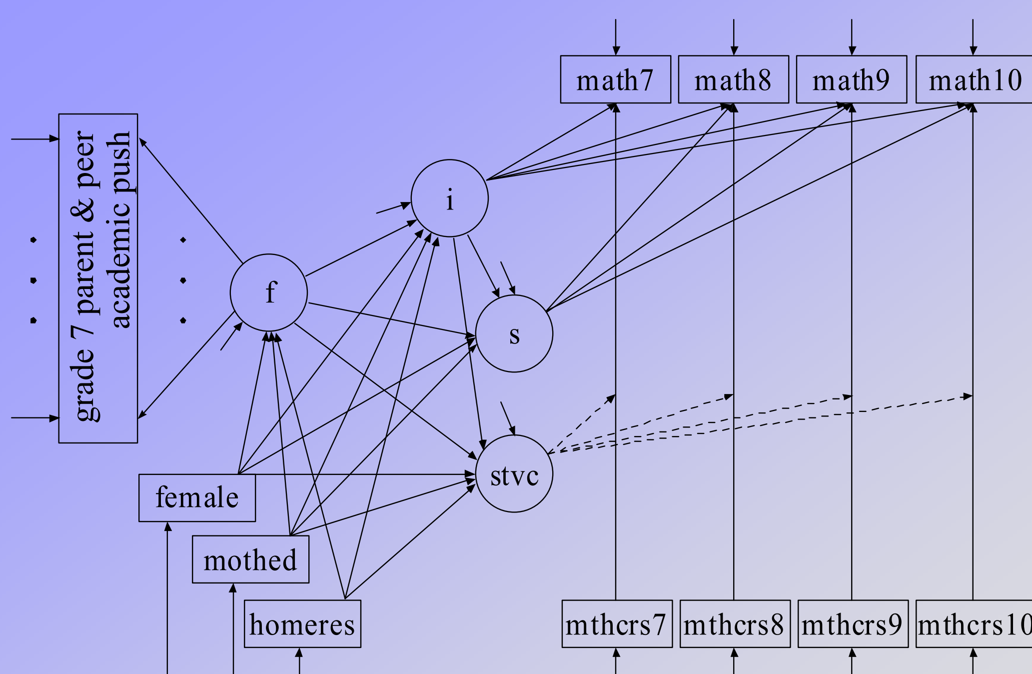
Students with only HS Expectations in G7



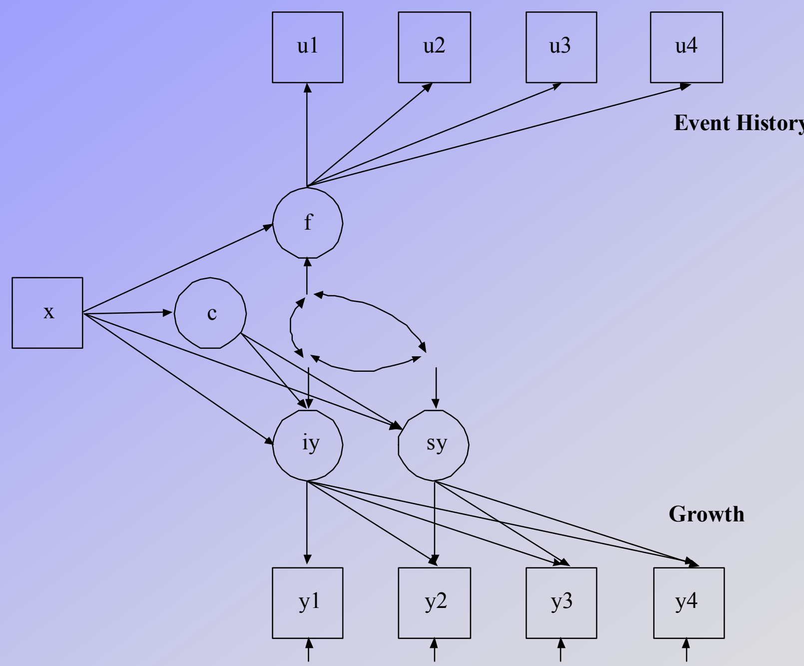
LSAY Math Achievement in Grades 7-10 and High School Dropout




math7
math8
math9
math10
i
s
stvc
female
mothed
homeres
mthcrs7
mthcrs8
mthcrs9
mthcrs10


grade 7 parent & peer academic push
math7
math8
math9
math10
f
i
s
stvc
female
mothed
homeres
mthcrs7
mthcrs8
mthcrs9
mthcrs10

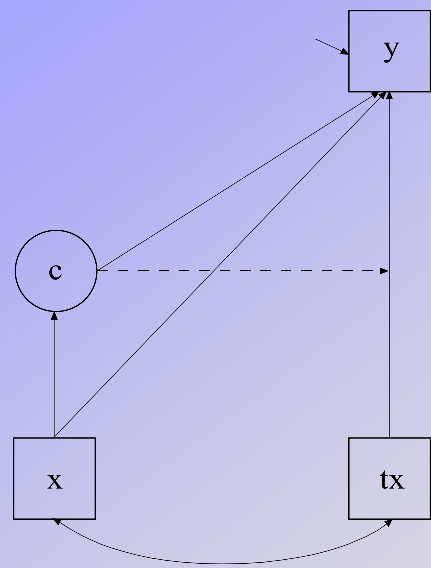
Onset (Survival) Followed by Growth



Categorical Latent Variables

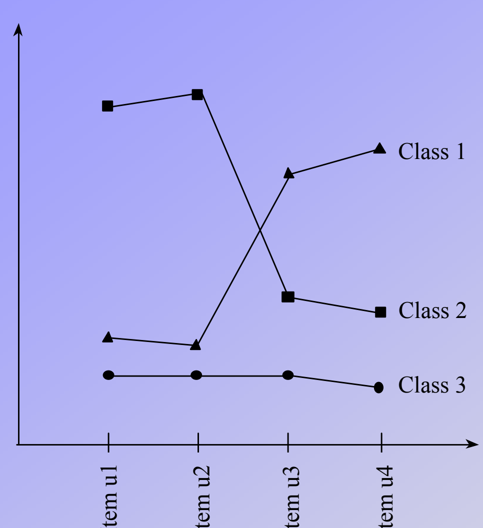
- Mixture regression
- Latent class analysis
- Latent transition analysis

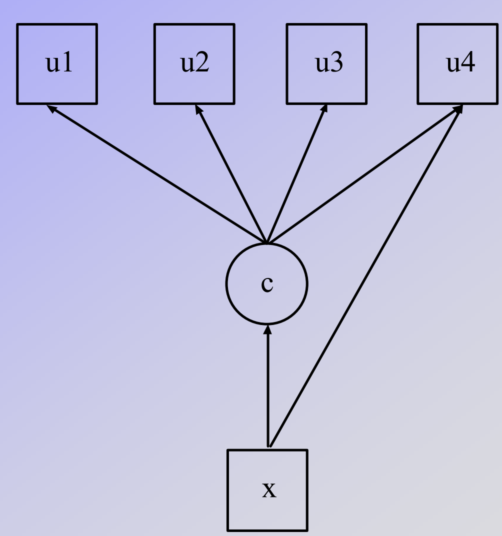
CACE Mixture Modeling



Latent Class Analysis

Item Profiles





Latent Transition Analysis

Transition Probabilities

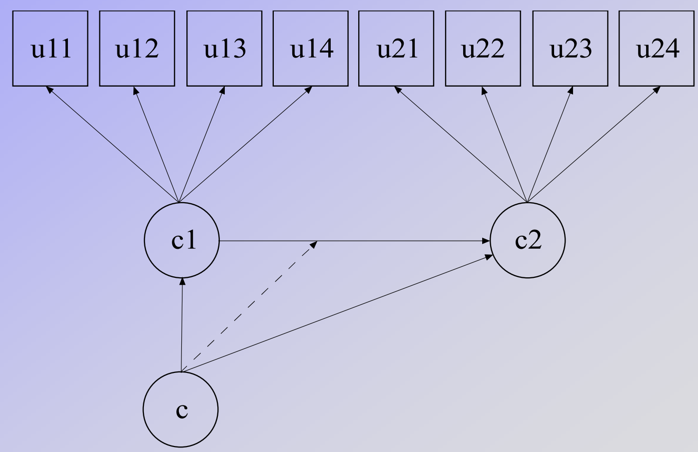
Mover Class (c=1)

c1 \ c2	1	2
1	0.6	0.4
2	0.3	0.7

Stayer Class (c=2)

c1 \ c2	1	2
1	0.90	0.10
2	0.05	0.95

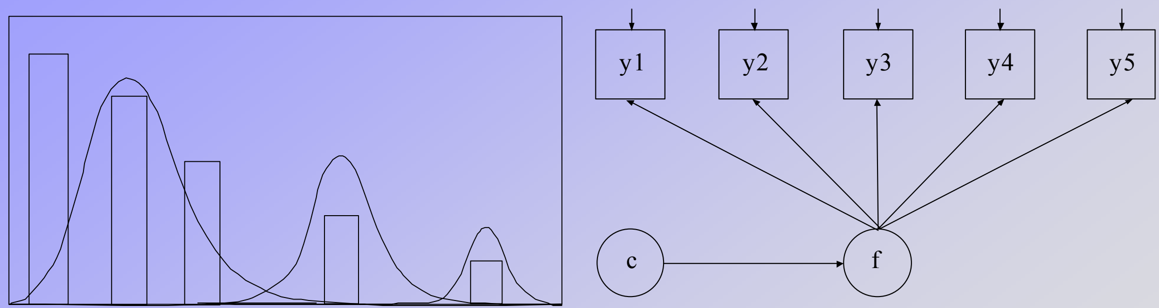
Time Point 1 Time Point 2



Combinations of Continuous and Categorical Latent Variables

- Mixture CFA, SEM
- Growth mixture modeling
- Second-order latent class analysis (twin modeling)
- Longitudinal Complier-Average Causal Effect (CACE) modeling in randomized preventive interventions
- Non-ignorable missing data modeling

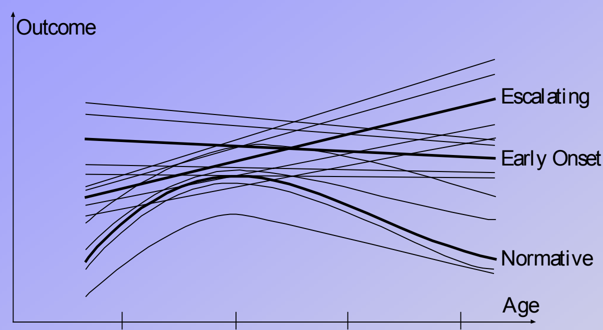
Factor Mixture - Non-Parametric Factor Modeling

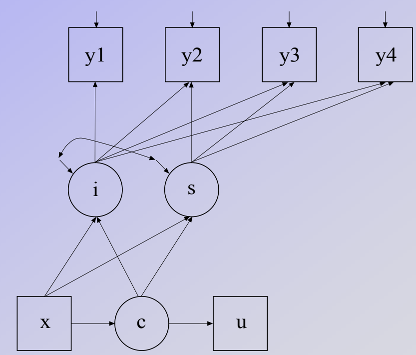


Growth Mixture Modeling

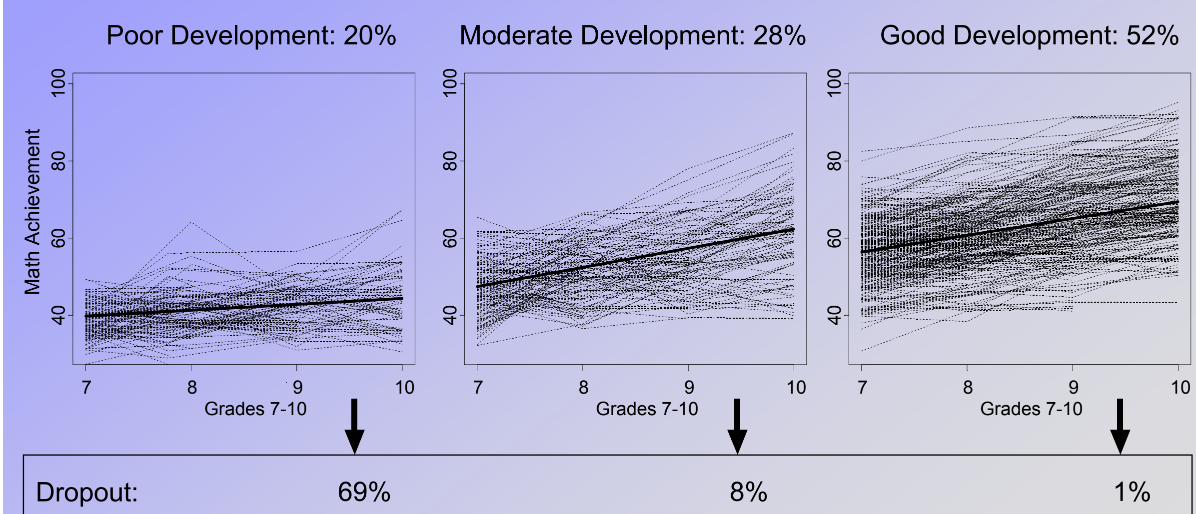
- Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. **Biometrics**, 55, 463-469.
- Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S., Carlin, J., & Liao, J. (2002). General growth mixture modeling for randomized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

Growth Mixture Modeling



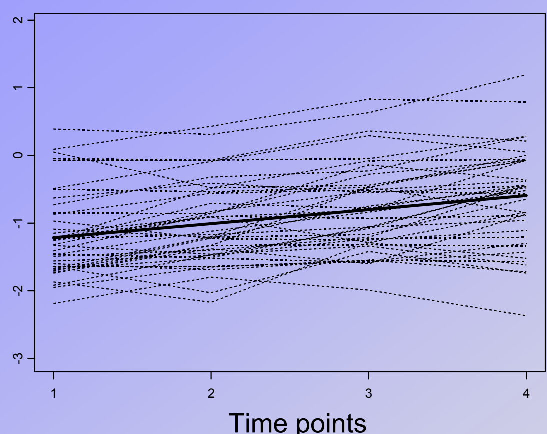


Growth Mixture Modeling: LSAY Math Achievement Trajectory Classes and the Prediction of High School Dropout

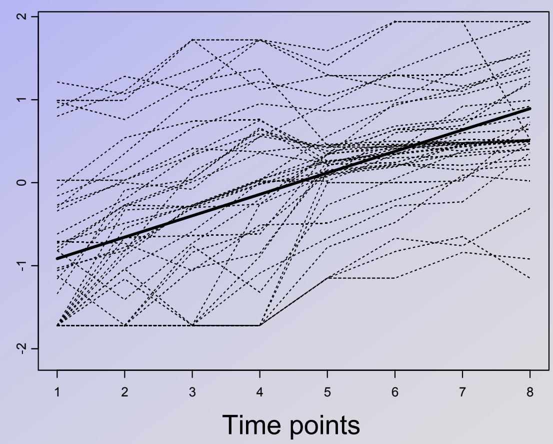


Predicting Reading Failure

Kindergarten Phonemic Awareness

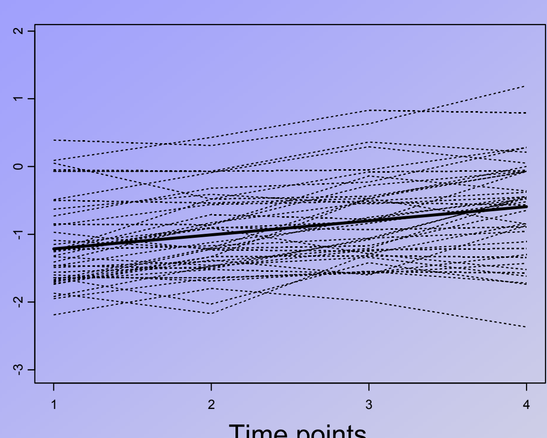


Grade 1-2 Word Recognition

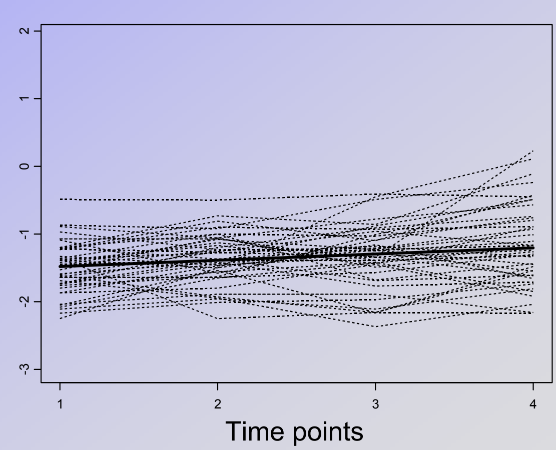


Kindergarten Phonemic Awareness

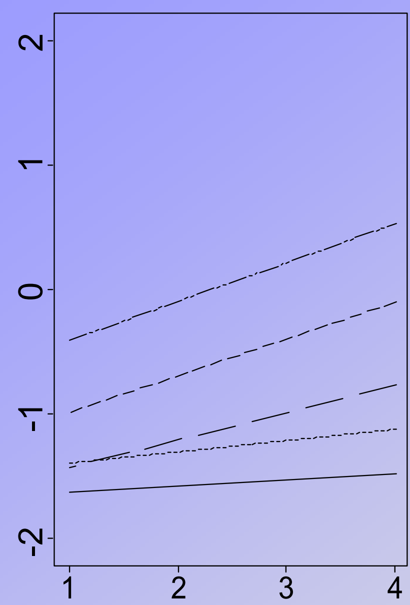
All Children (10% sample)



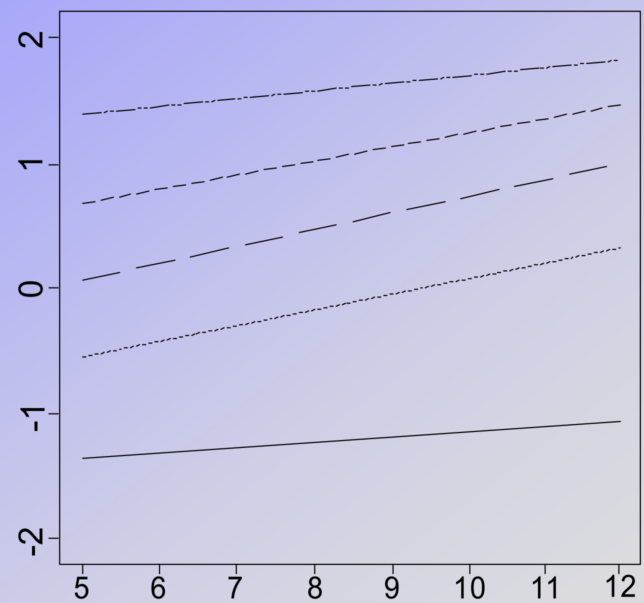
Children in the Lowest Decile of
End of Grade 2 Word Recognition



Kindergarten Growth Phonemic Awareness

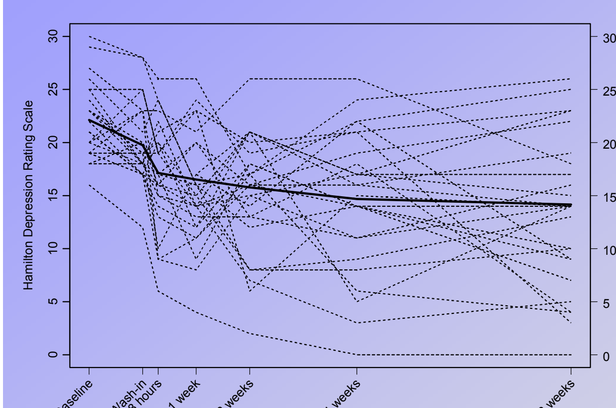


Grade 1 and Grade 2 Growth Word Recognition

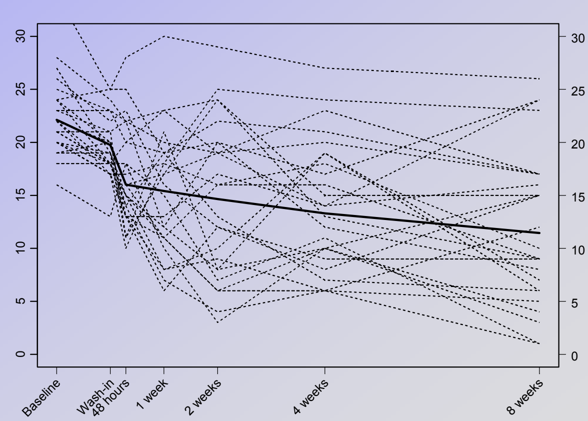


A Clinical Trial of Depression Medication

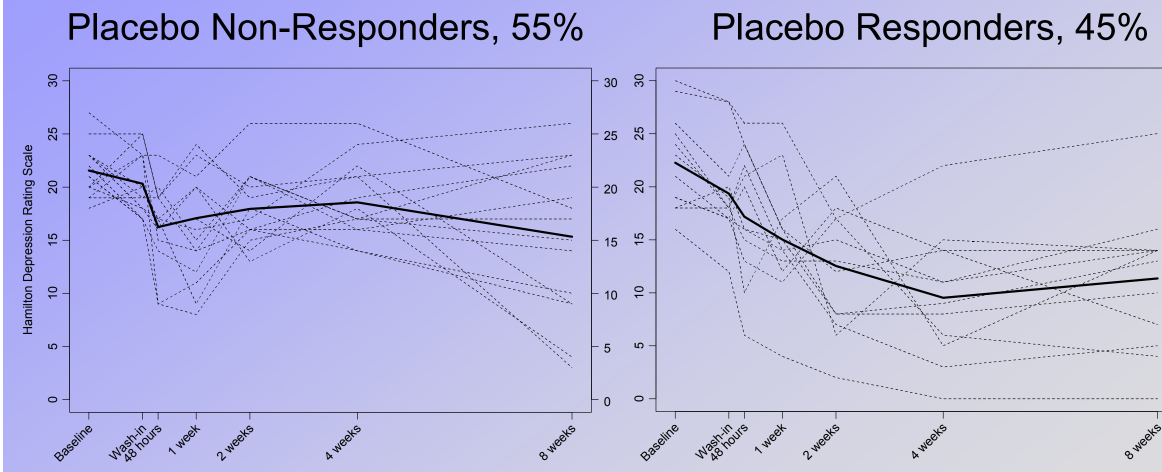
Placebo Group



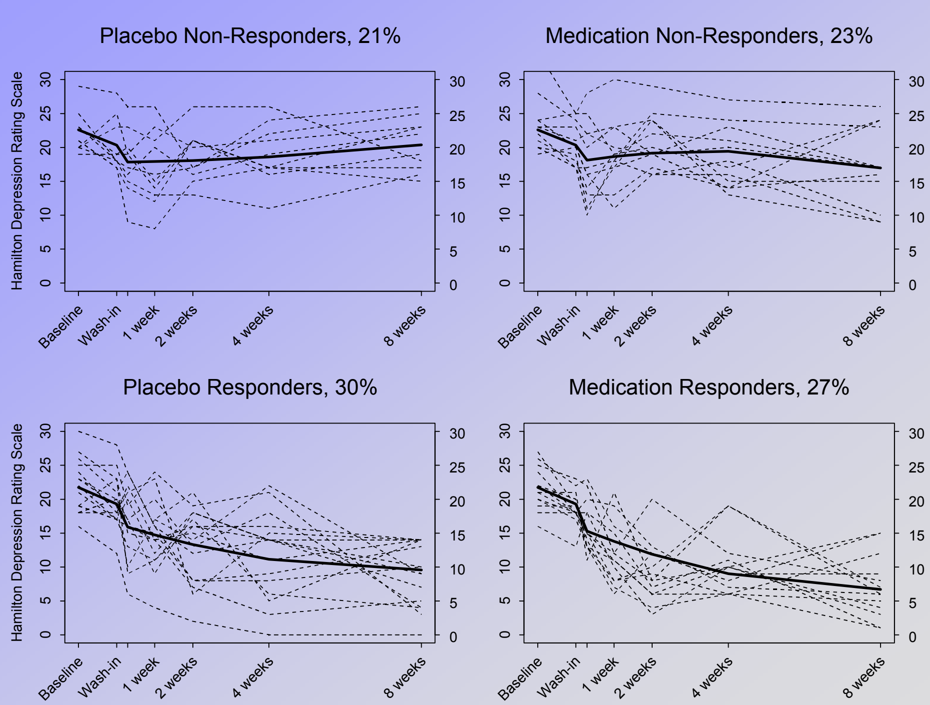
Medication Group



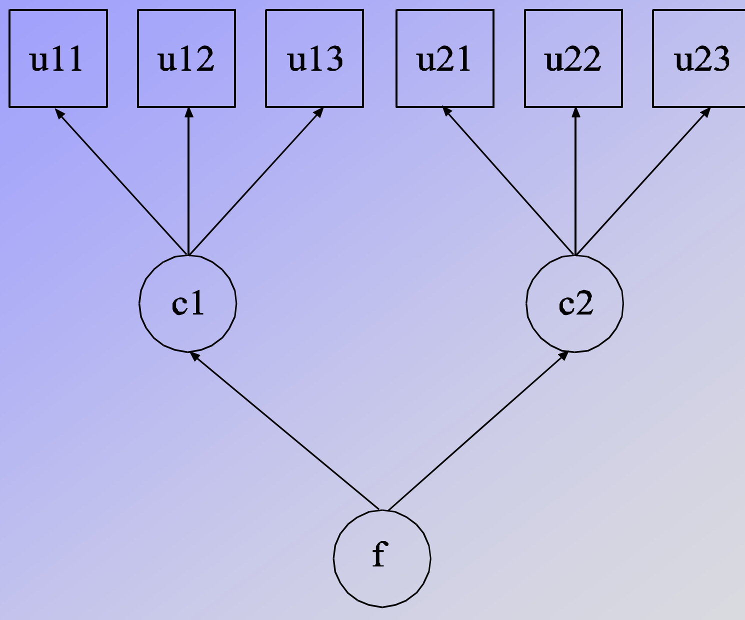
A Clinical Trial of Depression Medication



A Clinical Trial on Depression Medication



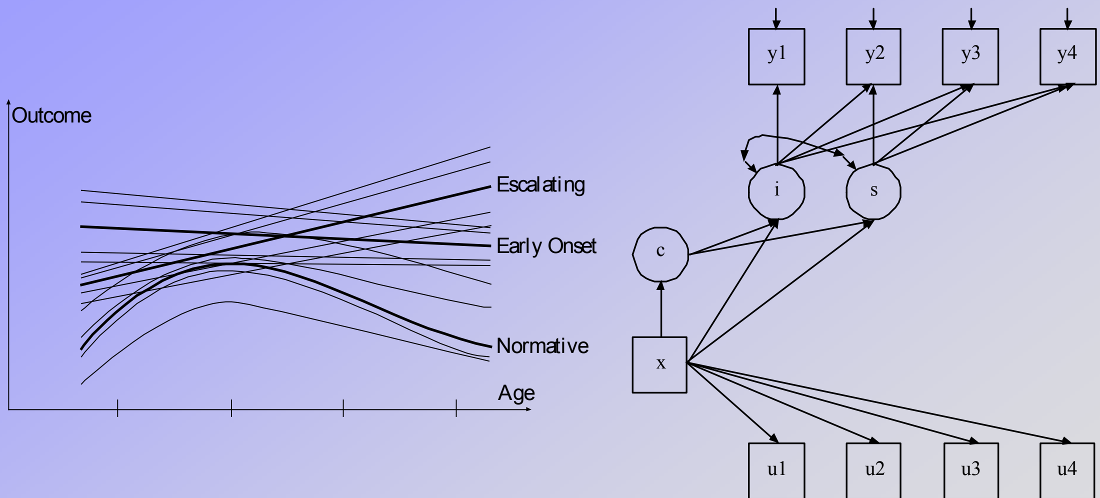
Twin Modeling



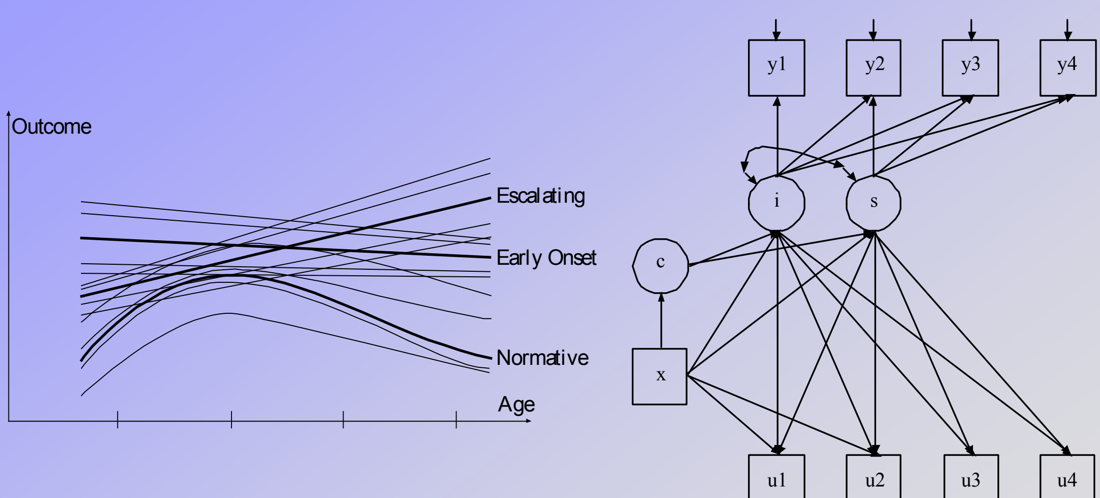
Longitudinal CACE, Non-Ignorable Missing Data

- Yau & Little (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. **Journal of the American Statistical Association**, 96, 1232-1244.
- Frangakis & Rubin (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. **Biometrika**, 86, 365-379.
- Muthén, Jo, & Brown (2003). Comment on the Barnard, Frangakis, Hill & Rubin article, Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. **Journal of the American Statistical Association**, 98, 311-314.

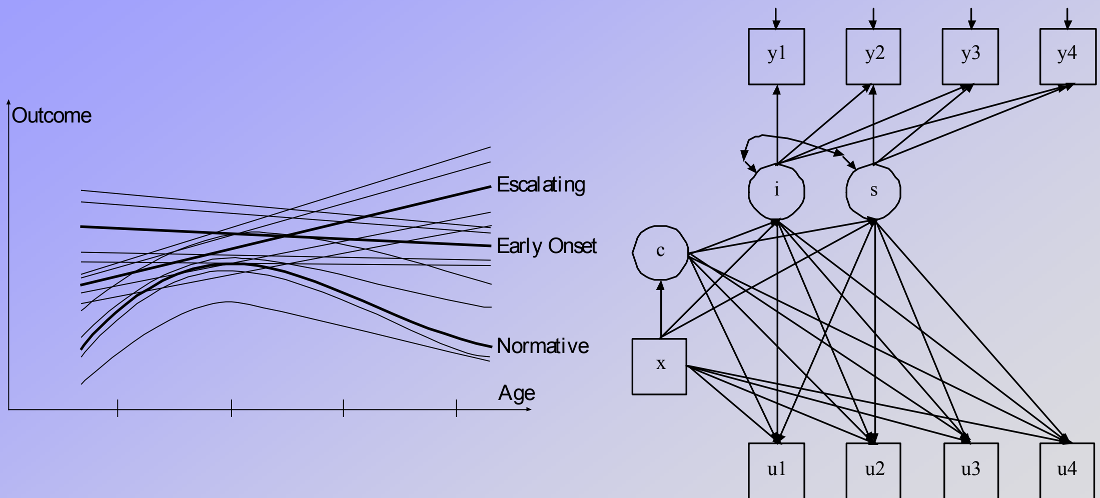
Growth Mixture Modeling with Non-Ignorable Missingness as a Function of Latent Variables



Growth Mixture Modeling with Non-Ignorable Missingness as a Function of Latent Variables



Growth Mixture Modeling with Non-Ignorable Missingness as a Function of Latent Variables



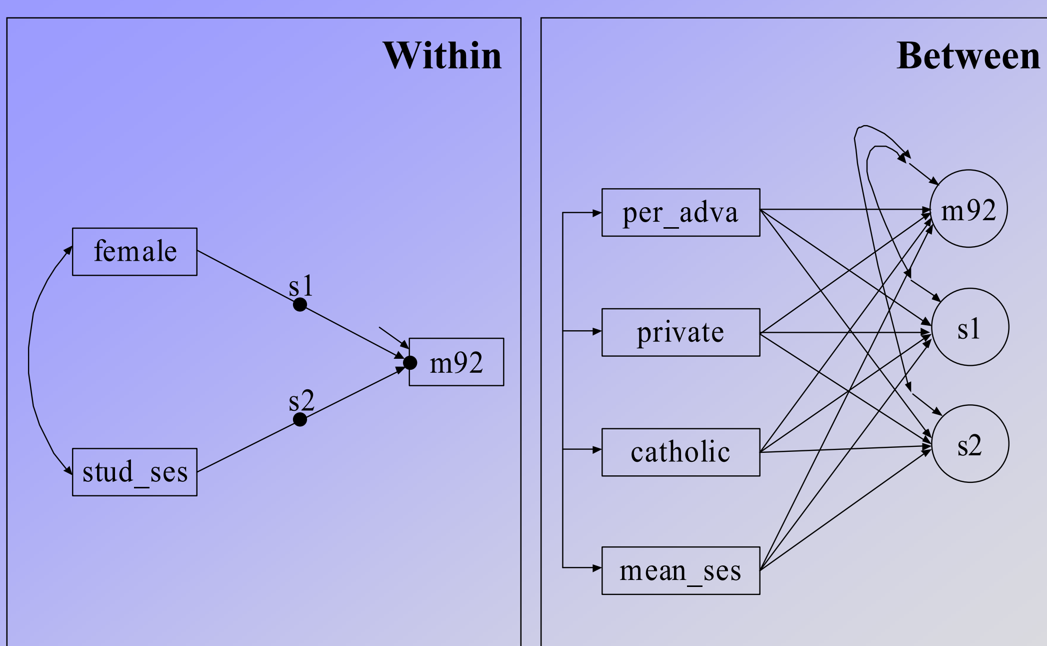
Multilevel Modeling with Continuous and Categorical Latent Variables

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

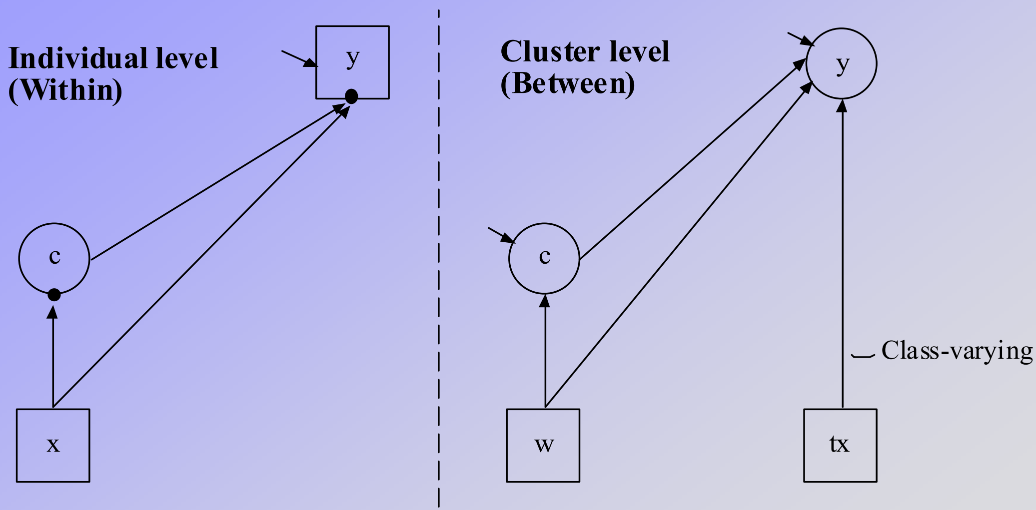
- Multilevel regression mixture analysis (CACE)
- Multilevel latent class analysis
- Multilevel growth mixture modeling

2-Level Regression of NELS Math Achievement

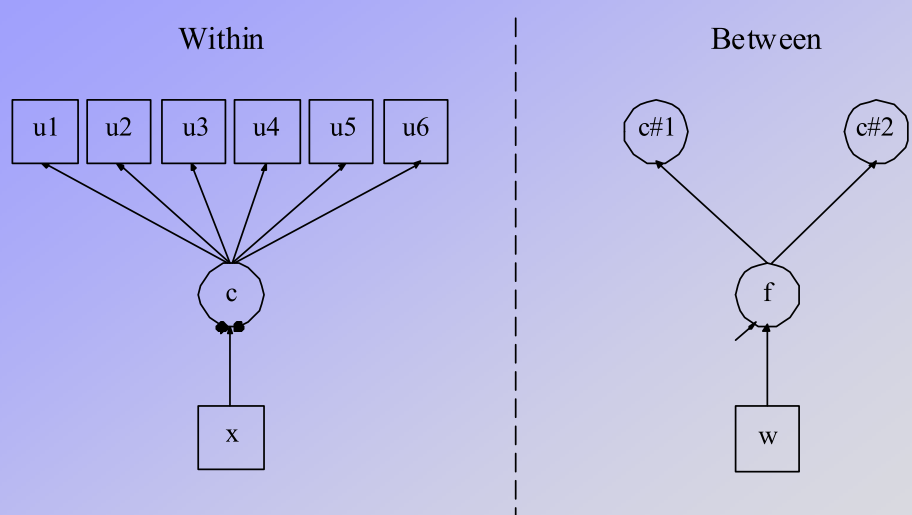
- **Data source**: NELS, n = 14,217 students in 913 schools
- **Outcome**: math achievement in grade 12
- **Individual-level covariates**: female, stud_ses
- **School-level covariates**: per_adva (percent teachers with an MA or higher), school type (public, private, catholic), family mean ses

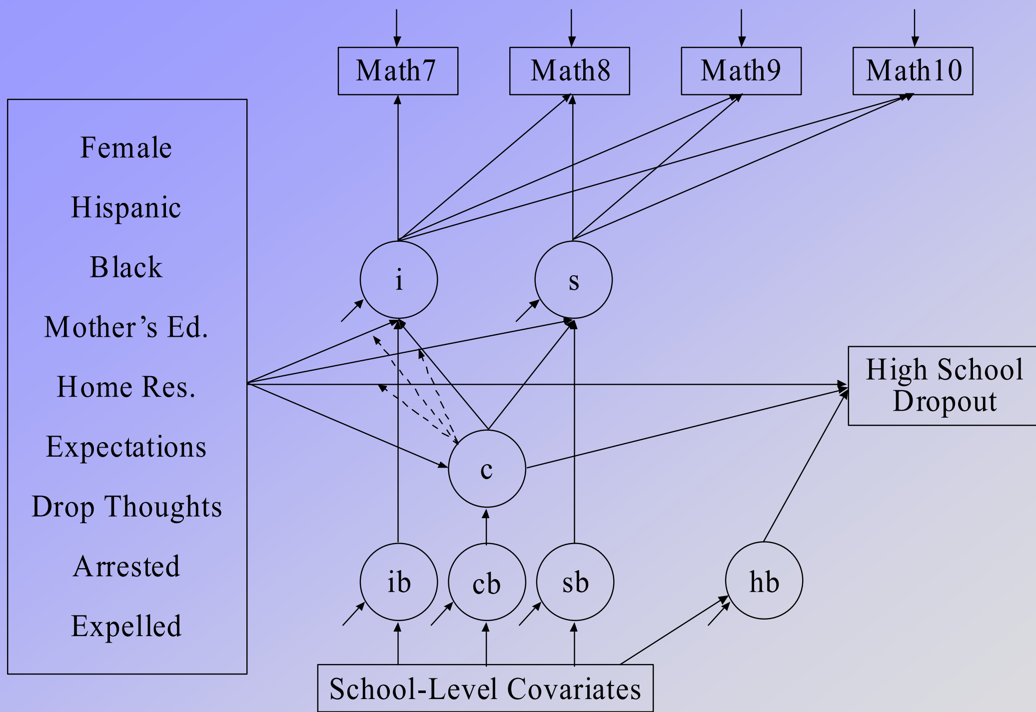


Two-Level CACE Mixture Modeling



Two-Level Latent Class Analysis





References

- See course and Mplus web sites