Mplus Short Courses
Topic 5

Categorical Latent Variable Modeling Using Mplus: Cross-Sectional Data

Linda K. Muthén
Bengt Muthén

www.statmodel.com
06/29/2009

Table Of Contents

Mplus Background

- Inefficient dissemination of statistical methods:
  - Many good methods contributions from biostatistics, psychometrics, etc are underutilized in practice
- Fragmented presentation of methods:
  - Technical descriptions in many different journals
  - Many different pieces of limited software
- Mplus: Integration of methods in one framework
  - Easy to use: Simple, non-technical language, graphics
  - Powerful: General modeling capabilities
- Mplus versions
  - V1: November 1998
  - V2: February 2001
  - V3: March 2004
  - V4: February 2006
  - V5: November 2007
  - V5.2: November 2008
- Mplus team: Linda & Bengt Muthén, Thuy Nguyen, Tihomir Asparouhov, Michelle Conn, Jean Maninger

Statistical Analysis With Latent Variables A General Modeling Framework

Statistical Concepts Captured By Latent Variables

Continuous Latent Variables

- Measurement errors
- Factors
- Random effects
- Frailties, liabilities
- Variance components
- Missing data

Categorical Latent Variables

- Latent classes
- Clusters
- Finite mixtures
- Missing data

Statistical Analysis With Latent Variables
A General Modeling Framework (Continued)

Models That Use Latent Variables

Continuous Latent Variables

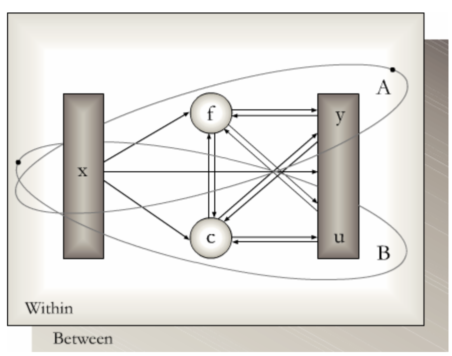
- Factor analysis models
- Structural equation models
- Growth curve models
- Multilevel models

Categorical Latent Variables

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

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

General Latent Variable Modeling Framework



- Observed variables
  - x background variables (no model structure)
  - y continuous and censored outcome variables
  - u categorical (dichotomous, ordinal, nominal) and count outcome variables
- Latent variables
  - f continuous variables
    - – interactions among f's
  - c categorical variables
    - – multiple c's

Mplus

Several programs in one

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

Fully integrated in the general latent variable framework

Overview Of Mplus Courses

- **Topic 1.** August 20, 2009, Johns Hopkins University: Introductory - advanced factor analysis and structural equation modeling with continuous outcomes
- **Topic 2.** August 21, 2009, Johns Hopkins University: Introductory - advanced regression analysis, IRT, factor analysis and structural equation modeling with categorical, censored, and count outcomes
- **Topic 3.** March, 2010, Johns Hopkins University: Introductory and intermediate growth modeling
- **Topic 4.** March, 2010, Johns Hopkins University: Advanced growth modeling, survival analysis, and missing data analysis

Overview Of Mplus Courses (Continued)

- **Topic 5.** August, 2010, Johns Hopkins University: Categorical latent variable modeling with cross-sectional data
- **Topic 6.** August 2010, Johns Hopkins University: Categorical latent variable modeling with longitudinal data

- **Topic 7.** March, 2011, Johns Hopkins University: Multilevel modeling of cross-sectional data
- **Topic 8.** March 2011, Johns Hopkins University: Multilevel modeling of longitudinal data

Logistic And Probit Regression

Categorical Outcomes: Logit And Probit Regression

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

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

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

Example: Lung cancer and smoking among coal miners

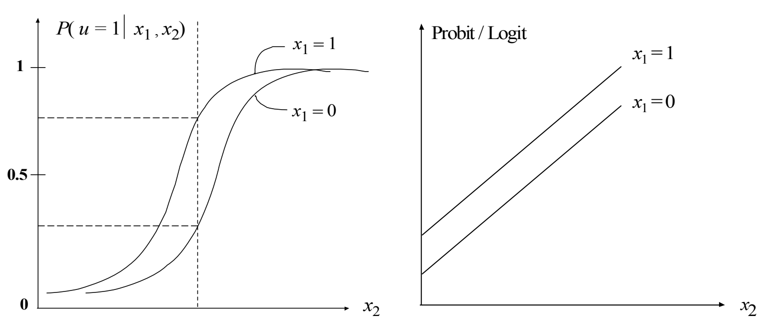
u lung cancer ($u = 1$) or not ($u = 0$)

x_1 smoker ($x_1 = 1$), non-smoker ($x_1 = 0$)

x_2 years spent in coal mine

Categorical Outcomes: Logit And Probit Regression

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



Interpreting Logit And Probit Coefficients

- Sign and significance
- Odds and odds ratios
- Probabilities

Logistic Regression And Log Odds

Odds $(u = 1 \mid x) = P(u = 1 \mid x) / P(u = 0 \mid x)$

$= P(u = 1 \mid x) / (1 - P(u = 1 \mid x))$.

The logistic function

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

gives a log odds linear in x,

$logit = log\,[odds\,(u = 1 \mid x)] = log\,[P(u = 1 \mid x) / (1 - P(u = 1 \mid x))]$

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

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

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

Logistic Regression And Log Odds (Continued)

- *logit* = *log odds* = $\beta_0 + \beta_1 x$
- When x changes one unit, the *logit* (*log odds*) changes β_1 units
- When x changes one unit, the *odds* changes e^{β_1} units

Further Readings On Categorical Variable Analysis

Agresti, A. (2002). Categorical data analysis. Second edition. New York: John Wiley & Sons.

Agresti, A. (1996). An introduction to categorical data analysis. New York: Wiley.

Hosmer, D. W. & Lemeshow, S. (2000). Applied logistic regression. Second edition. New York: John Wiley & Sons.

Long, S. (1997). Regression models for categorical and limited dependent variables. Thousand Oaks: Sage.

Modeling With Categorical Latent Variables

Statistical Analysis With Latent Variables A General Modeling Framework

Statistical Concepts Captured By Latent Variables

Continuous Latent Variables

- Measurement errors
- Factors
- Random effects
- Frailties, liabilities
- Variance components
- Missing data

Categorical Latent Variables

- Latent classes
- Clusters
- Finite mixtures
- Missing data

Statistical Analysis With Latent Variables
A General Modeling Framework (Continued)

Models That Use Latent Variables

Continuous Latent Variables

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

Categorical Latent Variables

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

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

Overview Of Analysis
With Categorical Latent Variables (Mixtures)

Analysis With Categorical Latent Variables

Used to capture heterogeneity when individuals come from different unobserved subpopulations in order to avoid biases in parameter estimates, standard errors, and tests of model fit

Application Areas

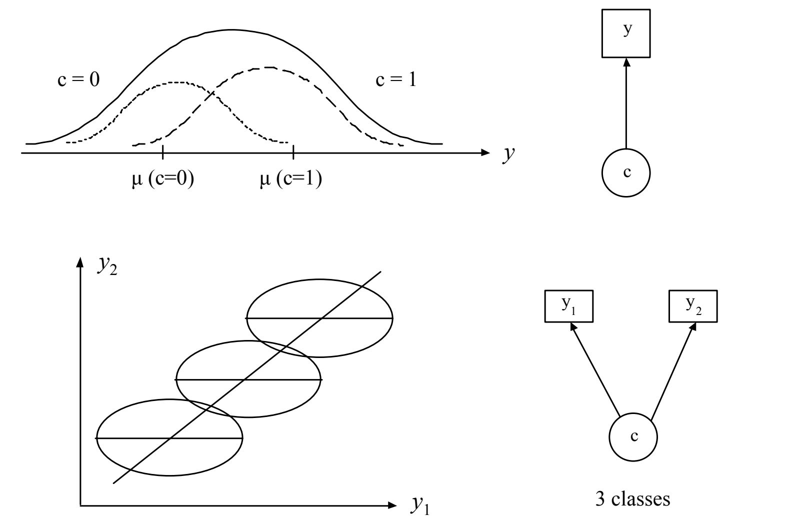
- Cross-sectional data
  - Medical and psychiatric diagnosis – schizophrenia, depression, alcoholism
  - Market segmentation
  - Mastery in educational development
- Longitudinal data
  - Multiple disease processes – prostate-specific antigen development
  - Developmental pathways – adolescent-limited versus life-course persistent antisocial behavior

Analysis With Categorical Latent Variables (Continued)

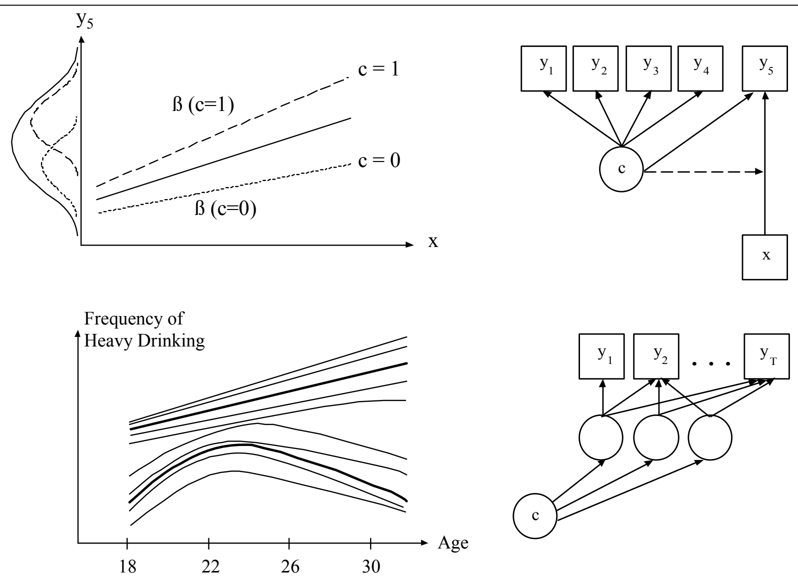
Analysis Methods

- Regression mixture models – CACE intervention
- Latent class analysis with and without covariates
- Latent profile analysis
- Latent transition analysis
- Latent class growth analysis
- Growth mixture modeling
- Discrete-time survival modeling

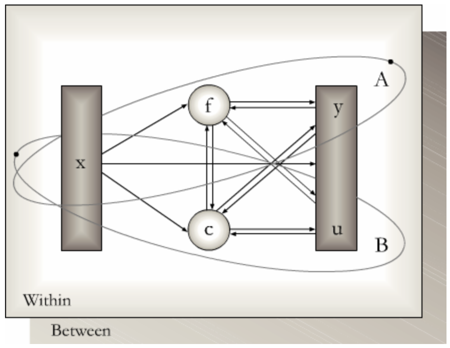
Mixture Modeling



Mixture Modeling (Continued)



General Latent Variable Modeling Framework



- Observed variables
  - x background variables (no model structure)
  - y continuous and censored outcome variables
  - u categorical (dichotomous, ordinal, nominal) and count outcome variables
- Latent variables
  - f continuous variables
    - – interactions among f's
  - c categorical variables
    - – multiple c's

Summary Of Techniques Using Mixtures

	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
FMA	u, y	Single	Multiple	Yes	Yes
LTA	u, y	Multiple	Multiple	No	Yes
LCGA	u, y	Multiple	Single Multiple	No	Yes (GMM)
GMM	u, y	Multiple	Single Multiple	Yes	Yes
DTSMA	u	Multiple	Single Multiple	No	Yes
LLLCA	u, y	Single Multiple	Single Multiple	NA	Yes

Summary Of Techniques Using Mixtures (Continued)

LCA – Latent Class Analysis
LPA – Latent Profile Analysis
LCFA – Latent Class Factor Analysis
FMA – Factor Mixture Analysis
LTA – Latent Transition Analysis
LCGA – Latent Class Growth Analysis
GMM – Growth Mixture Modeling
DTSMA – Discrete-Time Survival Mixture Analysis
LLLCA – Loglinear Latent Class Analysis

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

Further Readings On General Latent Variables

Hagenaars, J.A & McCutcheon, A. (2002). Applied latent class analysis. Cambridge: Cambridge University Press.

McLachlan, G.J. & Peel, D. (2000). Finite mixture models. New York: Wiley & Sons.

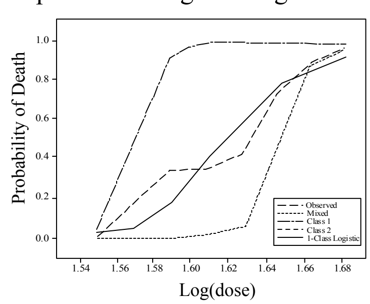
Muthén, B. (2002). Beyond SEM: General latent variable modeling. Behaviormetrika, 29, 81-117. (#96)

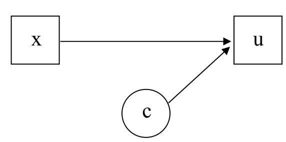
Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 1-24. Charlotte, NC: Information Age Publishing, Inc.

Regression Mixture Analysis

Logistic Regression Mixture Analysis

Comparison of Logistic Regression Curves*





Mixed Logistic Regression Equation

$$P(x) = \frac{.34}{1 + e^{-(-196.2 + 124.8x)}} + \frac{.66}{1 + e^{-(-205.7 + 124.8x)}}$$

* Reproduced from data analyzed by Follmann and Lambert, “Generalizing Logistic Regression by Nonparametric Mixing,” *Journal of American Statistical Association*, 295-300: March, 1989.

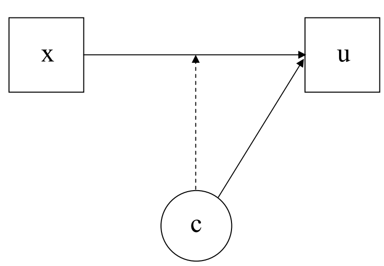
Randomized Response Modeling Of Sensitive Questions

"Did you take hard drugs last year?"

- 2 dice:
  - 5 - 10: answer truthfully
  - 2, 3, or 4: answer yes (regardless)
  - 11 or 12: answer no (regardless)
- Latent classes:
  - Class 1: people who answer truthfully
  - Class 2: people giving forced answers determined by the dice (random response class)

Source: Hox & Lensvelt-Mulders (2004)

Randomized Response Modeling Of Sensitive Questions (Continued)



- Class probability fixed to reflect design
- Class 2: intercept fixed to reflect forced yes proportion and slope fixed at zero (no relation to x)
- Class 1: Truthful class gives estimates of intercept and slope for u regressed on x

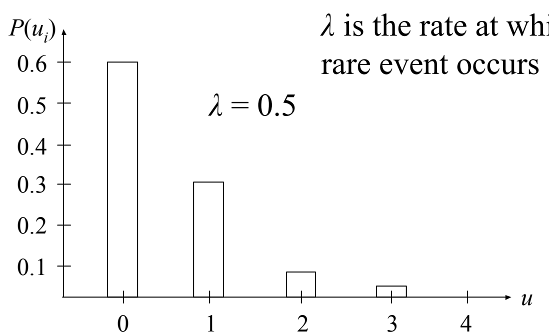
Regression With A Count Dependent Variable

Poisson Regression

A Poisson distribution for a count variable u_i has

$$P(u_i = r) = \frac{\lambda_i^r e^{-\lambda_i}}{r!}, \text{ where } u_i = 0, 1, 2, \ldots$$

λ is the rate at which a rare event occurs



Regression equation for the log rate:

$$^{e}\log \lambda_i = \ln \lambda_i = \beta_0 + \beta_1 x_i$$

Zero-Inflated Poisson (ZIP) Regression

A Poisson variable has mean = variance.

Data often have variance > mean due to preponderance of zeros.

$\pi = P$ (being in the zero class where only $u = 0$ is seen)

$1 - \pi = P$ (not being in the zero class with u following a Poisson distribution)

A mixture at zero:

$$P(u = 0) = \pi + (1 - \pi) \underbrace{e^{-\lambda}}_{\text{Poisson part}}$$

The ZIP model implies two regressions:

$$\text{logit}(\pi_i) = \gamma_0 + \gamma_1 x_i,$$

$$\ln \lambda_i = \beta_0 + \beta_1 x_i$$

35

Negative Binomial Regression

Unobserved heterogeneity ε_i is added to the Poisson model

$$ln\, \lambda_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{, where } exp(\varepsilon) \sim \Gamma$$

Poisson assumes

$$E(u_i \mid x_i) = \lambda_i$$

$$V(u_i \mid x_i) = \lambda_i$$

Negative binomial assumes

$$E(u_i \mid x_i) = \lambda_i$$

$$V(u_i \mid x_i) = \lambda_i (1 + \lambda_i\, \alpha)$$

NB with $\alpha = 0$ gives Poisson. When the dispersion parameter $\alpha > 0$, the NB model gives substantially higher probability for low counts and somewhat higher probability for high counts than Poisson.

Further variations are zero-inflated NB and zero-truncated NB (hurdle model or two-part model).

36

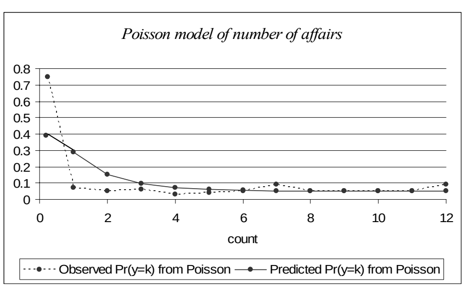
Mixture ZIP Regression

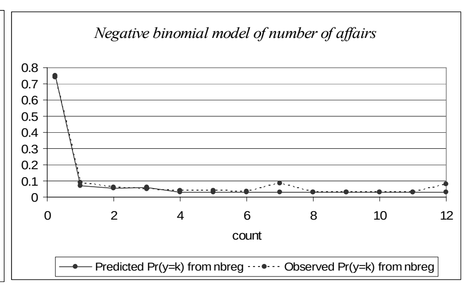
$$\text{logit}\,(\pi_i) = \gamma_0 + \gamma_1\, x_i\,,$$

$$\ln \lambda_{i\,|\,C=c_i} = \beta_{0c} + \beta_1 x_i$$

Equivalent generalization of zero-inflated negative binomial possible

Counts Of Marital Affairs





Dependent variable: Number of affairs reported in the last year
Covariates: Having kids, marital happiness, religiosity, years married
Sample size: 601

Source: Hilbe (2007). *Negative Binomial Regression.* Cambridge.

Model Alternatives For Counts Of Marital Affairs

Model	Log Likelihood	# of Parameters	BIC
Poisson	-1,399.913	13	2883
Negative Binomial	-724.240	14	1538
Zero-inflated Poisson	-783.002	14	1656
Zero-inflated negative binomial	-718.064	15	1532
2-class Poisson mixture	-728.001	15	1552
2-class negative binomial mixture	-718.064	16	1539
2-class zero-inflated Poisson	**-700.718**	**16**	**1504**
2-class zero-inflated negative binomial	-700.718	17	1510

Model Alternatives For Counts Of Marital Affairs (Continued)

Model	Log Likelihood	# of Parameters	BIC
2-class negative binomial hurdle	-726.039	15	1548
Poisson with normal residual	-735.953	14	1561

Input For Two-Class ZIP Regression

```
TITLE:      Hilbe page 112 example

DATA:       FILE = affairs1.dat;

VARIABLE:   NAMES = ID
            male age yrsmarr kids relig educ occup ratemarr
            naffairs affair vryhap hapavg avgmarr unhap vryrel
            smerel slghtrel notrel;
            USEVAR  = naffairs kids vryhap hapavg avgmarr vryrel
            smerel slghtrel notrel yrsmarr3 yrsmarr4 yrsmarr5
            yrsmarr6;
            COUNT = naffairs(pi);
            CLASSES = c(2);

DEFINE:     IF (yrsmarr==4) THEN yrsmarr3=1 ELSE yrsmarr3=0;
            IF (yrsmarr==7) THEN yrsmarr4=1 ELSE yrsmarr4=0;
            IF (yrsmarr==10) THEN yrsmarr5=1 ELSE yrsmarr5=0;
            IF (yrsmarr==15) THEN yrsmarr6=1 ELSE yrsmarr6=0;

ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR = ML;
```


Input For Two-Class ZIP Regression (Continued)

```
MODEL:             %OVERALL%
                   naffairs ON  kids-yrsmarr6 (p1-p12);

! compute incidence rate ratios (Hilbe, p.109)

MODEL CONSTRAINT:  new(e1-e12);
                   e1=exp(p1);
                   e2=exp(p2);
                   e3=exp(p3);
                   e4=exp(p4);
                   e5=exp(p5);
                   e6=exp(p6);
                   e7=exp(p7);
                   e8=exp(p8);
                   e9=exp(p9);
                   e10=exp(p10);
                   e11=exp(p11);
                   e12=exp(p12);

OUTPUT:            TECH1;
```

Randomized Trials With Non-Compliance

Randomized Trials With NonCompliance

- Tx group (compliance status observed)
  - Compliers
  - Noncompliers
- Control group (compliance status unobserved)
  - Compliers
  - NonCompliers

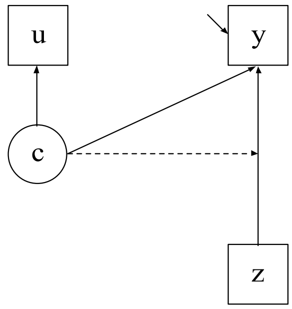
Compliers and Noncompliers are typically not randomly equivalent subgroups.

Four approaches to estimating treatment effects:

1. Tx versus Control (Intent-To-Treat; ITT)
2. Tx Compliers versus Control (Per Protocol)
3. Tx Compliers versus Tx NonCompliers + Control (As-Treated)
4. Mixture analysis (Complier Average Causal Effect; CACE):
   - Tx Compliers versus Control Compliers
   - Tx NonCompliers versus Control NonCompliers

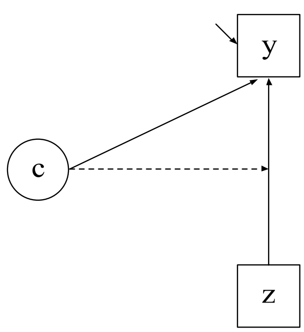
CACE: Little & Yau (1998) in Psychological Methods

CACE Estimation Via Mixture Modeling And ML Estimation In Mplus



z is a 0/1 dummy variable indicating treatment assignment

Randomized Trials with NonCompliance: Complier Average Causal Effect (CACE) Estimation



- UG Ex 7.24
  Including u, no training data
- UG Ex 7.23
  Excluding u, training data

TRAINING DATA

Training data can be used when latent class membership is known for certain individuals in the sample.

Training data must include one variable for each latent class. Each individual receives a value of 0 or 1 for each class variable. A zero indicates that the individual is not allowed to be in the class. A one indicates that the individual is allowed to be in the class.

CACE Application

With CACE models, there are two classes, compliers and noncompliers. The treatment group has known class membership. The control group does not. Therefore, the training data is as follows:

	Class 1 Compliers	Class 2 Non-Compliers
Control Group	1	1
Treatment Group Compliers	1	0
Treatment Group NonCompliers	0	1

JOBS Data

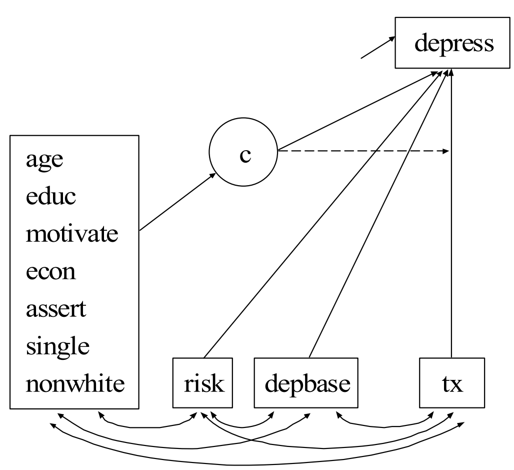
The JOBS data are from a Michigan University Prevention Research Center study of interventions aimed at preventing poor mental health of unemployed workers and promoting high quality of reemployment. The intervention consisted of five half-day training seminars that focused on problem solving, decision making group processes, and learning and practicing job search skills. The control group received a booklet briefly describing job search methods and tips. Respondents were recruited from the Michigan Employment Security Commission. After a series of screening procedures, 1801 were randomly assigned to treatment and control conditions. Of the 1249 in the treatment group, only 54% participated in the treatment.

The variables collected in the study include depression scores and outcome measures related to reemployment. Background variables include demographic and psychosocial variables.

JOBS Data (Continued)

Data for the analysis include the outcome variable of depression and the background variables of treatment status, age, education, marital status, SES, ethnicity, a risk score for depression, a pre-intervention depression score, a measure of motivation to participate, and a measure of assertiveness. A subset of 502 individuals classified as having high-risk of depression were analyzed.

The analysis replicates that of Little and Yau (1998).



Input For Complier Average Causal Effect (CACE) Model

```
TITLE:       Complier Average Causal Effect (CACE) estimation in a
             randomized trial.
DATA:        FILE IS wjobs.dat;
VARIABLE:    NAMES ARE depress risk Tx depbase age motivate educ assert
             single econ nonwhite x10 c1 c2;

             USEV ARE depress risk Tx depbase age motivate educ assert
             single econ nonwhite c1-c2;

             CLASSES = c(2);
             TRAINING = c1-c2;
ANALYSIS:    TYPE = MIXTURE;
MODEL:       %OVERALL%

             depress ON Tx risk depbase;
             c#1 ON age educ motivate econ assert single nonwhite;

             %C#2%           !c#2 is the noncomplier class (noshows)

             [depress];

             depress ON Tx@0;
OUTPUT:      TECH8;
```


Output Excerpts Complier Average Causal Effect (CACE) Model

Tests Of Model Fit

```
Loglikelihood

          H0 Value                      -729.414

Information Criteria

          Number of Free Parameters          14
          Akaike (AIC)                   1486.828
          Bayesian (BIC)                 1545.888
          Sample-Size Adjusted BIC       1501.451
             (n* = (n + 2) / 24)
          Entropy                           0.727
```

Output Excerpts Complier Average Causal Effect (CACE) Model (Continued)

Model Results

```
FINAL CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE

   Class 1       271.93488       0.54170
   Class 2       230.06512       0.45830

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY CLASS
MEMBERSHIP

Class Counts and Proportions

   Class 1            278         0.55378
   Class 2            224         0.44622

Average Latent Class Probabilities for Most Likely Latent Class
Membership (Row) by Latent Class (Column)
                      1             2

   Class 1          0.900         0.100
   Class 2          0.097         0.903
```


Output Excerpts Complier Average Causal Effect (CACE) Model (Continued)

Model Results (Continued)

```
                        Estimates        S.E.     Est./S.E.
Class 1

 Depress   ON
    TX                       -.310        .130        -2.378
    RISK                      .912        .247         3.685
    DEPBASE                 -1.463        .181        -8.077

Residual Variances
    DEPRESS                   .506        .037        13.742

Intercepts
    DEPRESS                  1.812        .299         6.068
```

Output Excerpts Complier Average Causal Effect (CACE) Model (Continued)

Model Results (Continued)

	Estimates	S.E.	Est./S.E.
Class 2			
Depress ON			
TX	.000	.000	.000
RISK	.912	.247	3.685
DEPBASE	-1.463	.181	-8.077
Residual Variances			
DEPRESS	.506	.037	13.742
Intercepts			
DEPRESS	1.633	.273	5.977

Output Excerpts Complier Average Causal Effect (CACE) Model (Continued)

Model Results (Continued)

LATENT CLASS REGRESSION MODEL PART

C#1 ON			
AGE	.079	.015	5.184
EDUC	.300	.068	4.390
MOTIVATE	.667	.157	4.243
ECON	-.159	.152	-1.045
ASSERT	-.376	.143	-2.631
SINGLE	.540	.283	1.908
NONWHITE	-.499	.317	-1.571
Intercepts			
C#1	-8.740	1.590	-5.498

Further Readings On CACE

Angrist, J.D., Imbens, G.W., Rubin, D.B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91, 444-445.

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. Journal of Educational and Behavioral Statistics, 27, 385-409.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. Psychological Methods, 7, 178-193.

Jo, B., Asparouhov, T., Muthén, B., Ialongo, N. & Brown, H. (2007). Cluster randomized trials with treatment noncompliance. Accepted for publication in Psychological Methods.

Little, R.J. & Yau, L.H.Y. (1998). Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. Psychological Methods, 3, 147-159.

Causal Inference

Causal Inference Concepts

- Potential outcomes
- Principal Stratification
- Finite mixtures

Potential Outcomes Framework

- Treatment variable X (e.g., X dichotomous with X=1 or X=0)
- Observed outcome Y, potential outcome variables Y(1), Y(0)
- Observed outcome under selected trmt x equals potential outcome under trmt assignment X=x : $y_i = y_i(x)$ if $x_i = x$

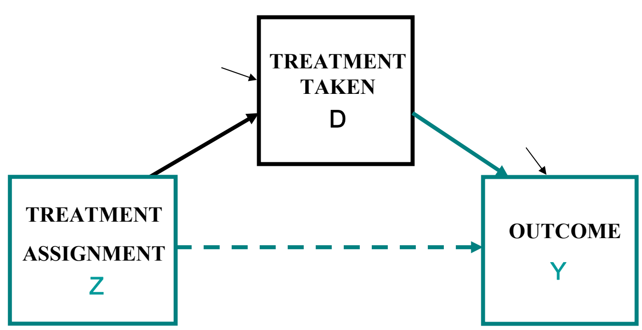
Subject #	**X**	**Y(1)**	**Y(0)**	**Causal Effect**	**Y**
1	$x_1 = 1$	$y_1(1)$	$y_1(0)$	$y_1(1)- y_1(0)$	$y_1 = y_1(1)$
2	$x_2 = 0$	$y_2(1)$	$y_2(0)$	$y_2(1)- y_2(0)$	$y_2 = y_2(0)$
…	…	…	…	…	…
N	$x_N = 1$	$y_N(1)$	$y_N(0)$	$y_N(1)- y_N(0)$	$y_N = y_N(1)$
Means				**ACE**	

Causal Inference And Non-Compliance

Causal Effects: The AIR (1996) Vietnam Draft Example

Angrist, Imbens & Rubin (1996) in JASA

- Conscription into the military randomly allocated via draft lottery



Causal Effects: The AIR (1996) Vietnam Draft Example (Continued)

- Z: treatment assignment (draft status)
  - Z = 1: assigned to serve in the military (for low lottery numbers)
  - Z = 0: not assigned to serve (for high lottery numbers)
- D: treatment taken (veteran status)
  - D = 1: served in the military
  - D = 0: did not serve in the military
- Y: health outcome (mortality after discharge)
- Note that D is not always = Z
  - avoid the draft (or deferred for medical reasons);

    non-compliance: Z = 1, D=0
  - volunteer for military service: Z = 0, D = 1

Causal Effect Of Z On Y, $Y_i(1, D_i(1)) - Y_i(0, D_i(0))$, For The Population Of Units Classified By $D_i(0)$ And $D_i(1)$

		$D_i(0)$: 0	$D_i(0)$: 1
$D_i(1)$	0	$Y_i(1, 0) - Y_i(0, 0) = 0$ Never-taker (π_n ; μ_{1n}, μ_{0n})	$Y_i(1, 0) - Y_i(0, 1) = -(Y_i(1) - Y_i(0))$ Defier ($\pi = 0$)
	1	$Y_i(1, 1) - Y_i(0, 0) = Y_i(1) - Y_i(0)$ Complier (π_c ; μ_{1c}, μ_{0c})	$Y_i(1, 1) - Y_i(0, 1) = 0$ Always-taker (π_a ; μ_{1a}, μ_{0a})

Average causal effect for compliers

$$E[(Y_i(1) - Y_i(0)) \mid D_i(1) - D_i(0) = 1] = \frac{E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]}{E[D_i(1) - D_i(0)]} \quad (12)$$

Or, $\mu_{1c} - \mu_{0c} = (\mu_1 - \mu_0)/\pi_c$

Z → Y is attributed to D → Y under the exclusion restriction

Causal Effect of D → Y Continued

- Mixture of 3 latent classes. Identification of parameters. Mixture means:

$$\mu_1 = \pi_c \mu_{1c} + \pi_n \mu_{1n} + \pi_a \mu_{1a}.$$
$$\mu_0 = \pi_c \mu_{0c} + \pi_n \mu_{0n} + \pi_a \mu_{0a}.$$
$$\mu_1 - \mu_0 = \pi_c (\mu_{1c} - \mu_{0c}) + \pi_n \times 0 + \pi_a \times 0$$

- Average causal effect

$$\mu_{1c} - \mu_{0c} = (\mu_1 - \mu_0)/\pi_c$$

- Estimated average causal effect

$$(\bar{y}_1 - \bar{y}_0)/(p_{c+a} - p_a),$$

where

p_{c+a} is the proportion in the treatment group who take the treatment

p_a is the proportion in the control group who take the treatment

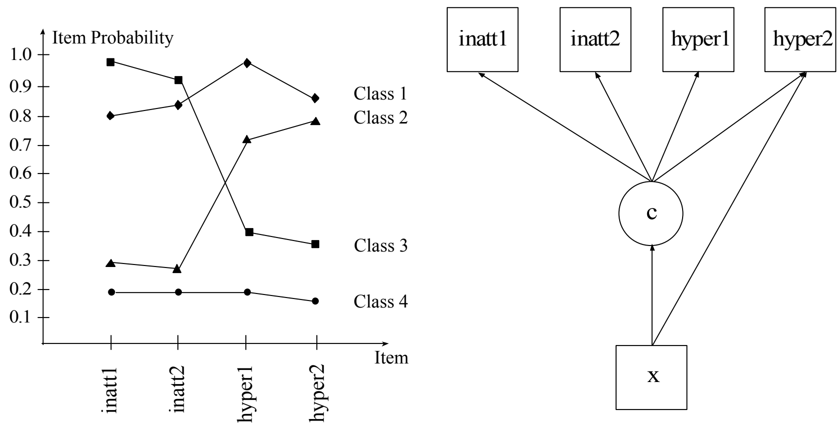
In JOBS (Little & Yau, 1998), there are no always-takers (could not get into the seminars if not assigned), so

$$p_a = 0$$
$$(\bar{y}_1 - \bar{y}_0)/p_c,$$

which is the Bloom (1984) IV estimate (the less compliance, the more attenuated the treatment and the more you upweight the mean difference).

Latent Class Analysis

Latent Class Analysis

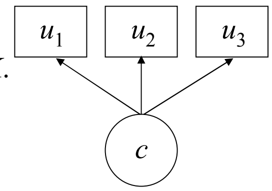


Latent Class Analysis (Continued)

Introduced by Lazarsfeld & Henry, Goodman, Clogg, Dayton & Mcready

- Setting
  - Cross-sectional data
  - Multiple items measuring a construct
  - Hypothesized construct represented as latent class variable (categorical latent variable
- Aim
  - Identify items that indicate classes well
  - Estimate class probabilities
  - Relate class probabilities to covariates
  - Classify individuals into classes (posterior probabilities)
- Applications
  - Diagnostic criteria for alcohol dependence. National sample, n = 8313
  - Antisocial behavior items measured in the NLSY. National sample, n = 7326

Latent Class Analysis Model



Dichotomous (0/1) indicators u: $u_1, u_2, \ldots, u_r$

Categorical latent variable c: $c = k$; $k = 1, 2, \ldots, K$.

Marginal probability for item $u_j = 1$,

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

Joint probability of all u's, assuming conditional independence

$$P(u_1, u_2, \ldots, u_r) = \sum_{k=1}^{K} P(c = k)\, P(u_1 \mid c = k)\, P(u_2 \mid c = k) \ldots P(u_r \mid c = k)$$

Note analogies with the case of continuous outcomes and continuous factors

LCA Estimation

Posterior Probabilities:

$$P(c = k \mid u_1, u_2, \ldots, u_r) = \frac{P(c = k)\, P(u_1 \mid c = k)\, P(u_2 \mid c = k) \ldots P(u_r \mid c = k)}{P(u_1, u_2, \ldots, u_r)}$$

Maximum-likelihood estimation via the EM algorithm:
c seen as missing data. EM: maximize E(complete-data log likelihood $|u_{i1}, u_{i2}, \ldots, u_{ir}$) wrt parameters.

- E (Expectation) step: compute $E(c_i \mid u_{i1}, u_{i2}, \ldots, u_{ir})$ = posterior probability for each class and $E(c_i\, u_{ij} \mid u_{i1}, u_{i2}, \ldots, u_{ir})$ for each class and u_j
- M (Maximization) step: estimate $P(u_j \mid c_k)$ and $P(c_k)$ parameters by regression and summation over posterior probabilities

LCA Parameters

Number of H_0 parameters in the (exploratory) LCA model with K classes and r binary u's: $K - 1 + K \times r$ (H_1 has $2^r - 1$ parameters).

$$\quad\quad\quad\quad\quad\quad\quad\quad\quad H_1 \quad\quad H_0$$

- 2 classes, 3 u: df = 0 computed as (8 – 1) – (1 + 6)
- 2 classes, 4 u: df = 6 computed as (16 – 1) – (1 + 8)
- 3 classes, 4 u: df = 1, but not identified because of 1 indeterminacy
- 3 classes, 5 u: df = 14 computed as (32 – 1) – (2 + 15)

Confirmatory LCA modeling applies restrictions to the parameters.

Logit versus Probability Scale. The u-c relation is a logit regression (binary u),

$$P(u = 1 \mid c) = \frac{1}{1 + exp\,(-Logit)}, \quad (81)$$

$$Logit = log\,[P/(1 - P)]. \quad (82)$$

For example:

Logit = 0: P = 0.5
Logit = -1: P = 0.27
Logit = 1: P = 0.73
Logit = -3: P = 0.05
Logit = -10: P = 0.00005

LCA Testing Against Data

- Model fit to frequency tables. Overall test against data
  – When the model contains only **u**, summing over the cells,

$$\chi^2_P = \sum_i \frac{(o_i - e_i)^2}{e_i}, \quad (82)$$

$$\chi^2_{LR} = 2\sum_i o_i \, log\, o_i / e_i. \quad (83)$$

A cell that has non-zero observed frequency and expected frequency less than .01 is not included in the χ^2 computation as the default. With missing data on **u**, the EM algorithm described in Little and Rubin (1987; chapter 9.3, pp. 181-185) is used to compute the estimated frequencies in the unrestricted multinomial model. In this case, a test of MCAR for the unrestricted model is also provided (Little & Rubin, 1987, pp. 192-193).

- Model fit to univariate and bivariate frequency tables. Mplus TECH10

Latent Class Analysis
Alcohol Dependence Criteria, NLSY 1989 (n = 8313)

Source: Muthén & Muthén (1995)

	Latent Classes				
	Two-class solution[1]		Three-class solution[2]		
	I	II	I	II	III
Prevalence	0.78	0.22	0.75	0.21	0.03
DSM-III-R Criterion	Conditional Probability of Fulfilling a Criterion				
Withdrawal	0.00	0.14	0.00	0.07	0.49
Tolerance	0.01	0.45	0.01	0.35	0.81
Larger	0.15	0.96	0.12	0.94	0.99
Cut down	0.00	0.14	0.01	0.05	0.60
Time spent	0.00	0.19	0.00	0.09	0.65
Major role-Hazard	0.03	0.83	0.02	0.73	0.96
Give up	0.00	0.10	0.00	0.03	0.43
Relief	0.00	0.08	0.00	0.02	0.40
Continue	0.00	0.24	0.02	0.11	0.83

[1]Likelihood ratio chi-square fit = 1779 with 492 degrees of freedom
[2]Likelihood ratio chi-square fit = 448 with 482 degrees of freedom

Latent Class Membership By Number Of DSM-III-R Alcohol Dependence Criteria Met (n=8313)

Source: Muthén & Muthén (1995)

		Latent Classes				
Number of Criteria Met		Two-class solution		Three-class solution		
	%	I	II	I	II	III
0	64.2	5335	0	5335	0	0
1	14.0	1161	1	1161	1	0
2	10.2	0	845	0	845	0
3	5.6	0	469	0	469	0
4	2.6	0	213	0	211	2
5	1.4	0	116	0	19	97
6	0.8	0	68	0	0	68
7	0.5	0	42	0	0	42
8	0.5	0	39	0	0	39
9	0.3	0	24	0	0	24
%	100.0	78.1	21.9	78.1	18.6	3.3

LCA Testing Of K – 1 Versus K Classes

Model testing by χ^2, BIC, and LRT

- Overall test against data: likelihood-ratio χ^2 with H_1 as the unrestricted multinomial (problem: sparse cells)
- Comparing models with different number of classes:
  - Likelihood-ratio χ^2 cannot be used
  - Bayesian information criterion (Schwartz, 1978)

$$BIC = -2 logL + h \times ln\ n, \qquad (81)$$

where h is the number of parameters and n is the sample size. Choose model with smallest BIC value.

  - Vuong-Lo-Mendell-Rubin likelihood-ratio test (Biometrika, 2001). Mplus TECH11
  - Bootstrapped likelihood ratio test. Mplus TECH14 (Version 4)

More On LCA Testing Of K – 1 Versus K Classes Bootstrap Likelihood Ratio Test (LRT): TECH14

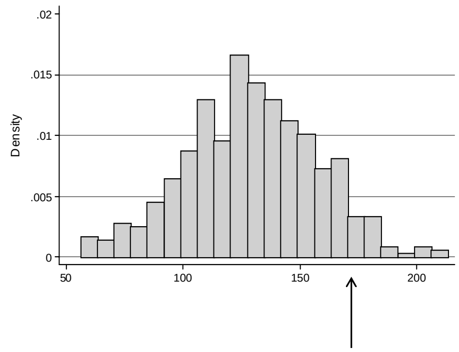
- LRT = 2*[logL(model 1) – logL(model2)], where model 2 is nested within model 1
- When testing a k-1-class model against a k-class model, the LRT does not have a chi-square distribution due to boundary conditions, but its distribution can be determined empirically by bootstrapping

Bootstrap steps:

1. In the k-class run, estimate both the k-class and the k-1-class model to get the LRT value for the data
2. Generate (at most) 100 samples using the parameter estimates from the k-1-class model and for each generated sample get the log likelihood value for both the k-1 and the k-class model to compute the LRT values for all generated samples
3. Get the p value for the data LRT by comparing its value to the distribution in 2.

More On LCA Testing
Bootstrap Likelihood Ratio Test (LRT): TECH14 (Continued)

2- vs. 3-class LRT Distribution



Observed-data LRT =169.41

Bootstrap LRT (Continued)

Technical 14 Output

```
    Number of initial stage random starts for k-1 class model       0
    Number of final stage optimizations for the k-1 class model     0
    Number of initial stage random starts for k class model        20
    Number of final stage optimizations for the k class model       5
    Number of bootstrap draws requested                        Varies

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR
6 (H0) VERSUS 7 CLASSES

    H0 Loglikelihood Value                          -3431.567
    2 Times the Loglikelihood Difference               37.009
    Difference in the Number of Parameters                 19
    Approximate P-Value                                0.2667
    Successful Bootstrap Draws                             15

WARNING: THE BEST LOGLIKELIHOOD VALUE WAS NOT REPLICATED IN 6 OUT OF
11 BOOTSTRAP DRAWS.  THE P-VALUE MAY NOT BE TRUSTWORTHY DUE TO LOCAL
MAXIMA. INCREASE THE NUMBER OF BOOTSTRAP LRT RANDOM STARTS.
```

Specifications Related To Tech14

- LRTSTARTS is used to give more random starts for the default sequential approach (2 to 100 bootstrap draws)
  - Default is LRTSTARTS = 0 0 20 5;
  - More thorough is LRTSTARTS = 2 1 50 15;

- LRTBOOTSTRAP is used to request number of bootstrap draws for non-sequential approach
  - Default is LRTBOOTSTRAP = 100;

Lo-Mendell-Rubin LRT: TECH11

The TECH11 option is used in conjunction with TYPE=MIXTURE to produce the Lo-Mendell-Rubin (Biometrika, 2001) likelihood ratio test of model fit that compares the estimated model with a model with one less class than the estimated model. The p-value obtained represents the probability that H0 is true, that the data have been generated by the model with one less class. A low p-value indicates that the estimated model is preferable. An adjustment to the test according to the Lo-Mendell-Rubin is also given. The model with one class less is obtained by deleting the first class in the estimated model. Because of this, it is recommended that the last class be the largest class. TECH11 is available only for ESTIMATOR=MLR. TECH11 is not available with training data.

Lo-Mendell-Rubin LRT: TECH11 (Continued)

Technical 11 Output

```
VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 5 (H0) VERSUS
6 CLASSES

            H0 Loglikelihood Value                  -40808.314
            2 Times the Loglikelihood Difference       408.167
            Difference in the Number of Parameters          18
            Mean                                        52.168
            Standard Deviation                          70.681
            P-Value                                     0.0019

       LO-MENDELL-RUBIN ADJUSTED LRT TEST

            Value                                      405.634
            P-Value                                     0.0020
```


Deciding On The Number Of Classes: Bootstrapped LRT (BLRT)

- Nylund, Muthen and Asparouhov (2006) simulation study
- BLRT has better Type I error than NCS and LMR
- BLRT finds the right number of classes better than BIC, NCS and LMR

BLRT: Bootstrap likelihood ratio test (TECH14)
NCS: Naïve Chi-square (2 × LL difference)
LMR: Lo-Mendell-Rubin (TECH11)

Monte Carlo Simulation Excerpt From Nylund, Asparouhov And Muthen (2006)

Latent class analysis with categorical outcomes

Which percent of the time does a certain number of classes get picked?

Model		BIC			NCS			LMR			BLRT		
		Classes			Classes			Classes			Classes		
	n	3	4	5	3	4	5	3	4	5	3	4	5
10-Item (Complex Structure) with 4 Latent Classes	200	92	**8**	0	2	**48**	41	34	**43**	9	16	**78**	6
	500	24	**76**	0	0	**34**	45	9	**72**	14	0	**94**	6
	1000	0	**100**	0	0	**26**	41	2	**80**	17	0	**94**	6

Other Considerations In Determining The Number Of Classes

Interpretability and usefulness:

- Substantive theory
- Auxiliary (external) variables
- Predictive validity

Quality Of Classification

- Classification table based on posterior class probabilities p_{ij}
  - Rows are individuals who have their highest probability in this class; entries are averaged p_{ij} over individuals
- Entropy

$$E_{\kappa} = 1 - \frac{\Sigma_i \, \Sigma_k \, (-\hat{p}_{ik} \, ln\hat{p}_{ik})}{n \; lnK} \, . \qquad (84)$$

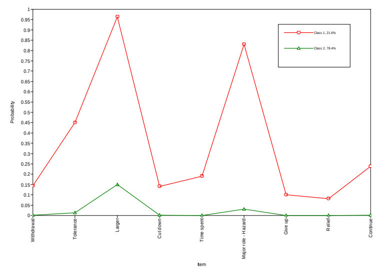
A value close to 1 indicates good classification in that many individuals have p_{ij} values close to either 0 or 1.

LCA Model Results For NLSY Alcohol Dependence Criteria

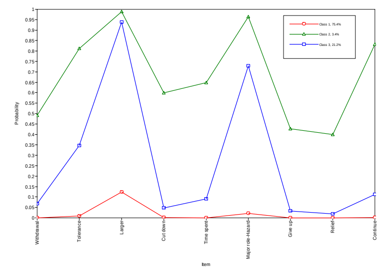
	Number of Classes			
	2	3	4	5
Pearson χ^2	128,906	773	664	585
LR χ^2	1,779	448	326	263
χ^2 df	492	482	472	462
# significant bivariate residuals (TECH10)	84	4	1	0
Loglikelihood	-14,804	-14,139	-14,078	-14,046
# of parameters	19	29	39	49
BIC	29,780	28,539	28,508	28,535
LMR (TECH11) p	0.000	0.000	0.008	0.082
BLRT (TECH14) p	0.000	0.000	0.000	0.000
Entropy	0.901	0.892	0.844	0.852

LCA Item Profiles For NLSY Alcohol Criteria

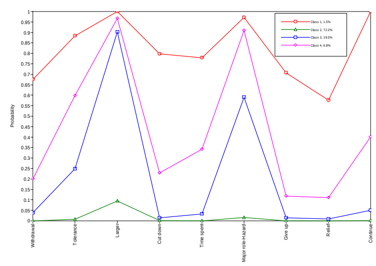
2-class LCA Item Profiles



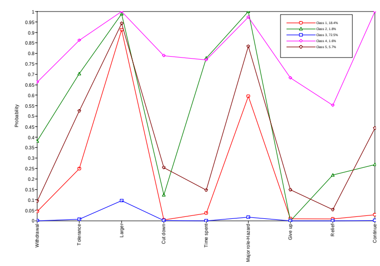
3-class LCA Item Profiles



4-class LCA Item Profiles



5-class LCA Item Profiles



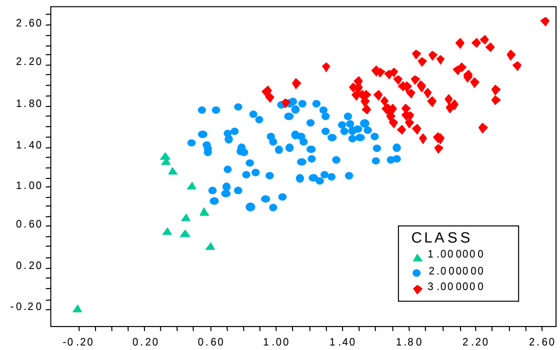
Input For NLSY Alcohol LCA

```
TITLE:          Alcohol LCA M & M (1995)

DATA:           FILE = bengt05_spread.dat;

VARIABLE:       NAMES = u1-u9;

                CATEGORICAL = u1-u9;

                CLASSES = c(3);

ANALYSIS:       TYPE = MIXTURE;

PLOT:           TYPE = PLOT3;
                SERIES = u1-u9(*);
```

Testing By Fitting Neighboring Models: Alcohol Dependence Criteria In The NLSY

- LCA, 3 classes: logL = -14,139, 29 parameters, BIC = 28,539
- FA, 2 factors: logL = -14,083, 26 parameters, BIC = 28,401

Factor 1: Alcohol Dependence



Factor 2: Alcohol Abuse

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.

Antisocial Behavior (ASB) Data (Continued)

Following is a list of the 17 items:

Property offense:
- Damaged property
- Shoplifting
- Stole < $50
- Stole > $50
- "Con" someone
- Take auto
- Broken into building
- Held stolen goods

Person offense:
- Fighting
- Use of force
- Seriously threaten
- Intent to injure
- Gambling operation

Drug offense:
- Use marijuana
- Use other drugs
- Sold marijuana
- Sold hard drugs

The items were dichotomized 0/1 with 0 representing never in the last year

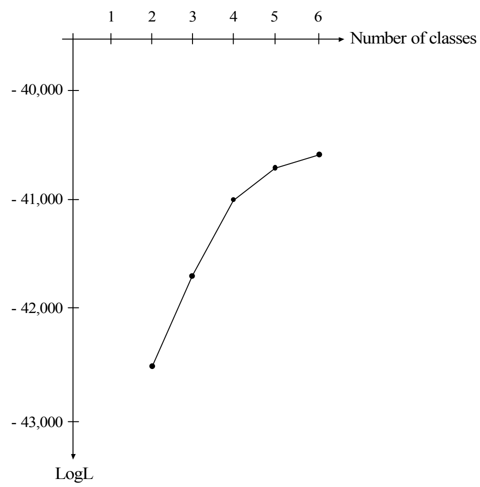
Are there different groups of people with different ASB profiles?

Deciding On The Number Of Classes For The ASB Items

Number of Classes	1	2	3	4	5	6
Loglikelihood	-48,168.475	-42,625.653	-41,713.142	-41,007.498	-40,808.314	-40,604.231
# par.	17	35	53	71	89	107
BIC	96,488	85,563	83,898	82,647	82,409	82,161
ABIC		85,452	83,730	82,421	82,126	81,821
AIC	96,370	85,321	83,532	82,157	81,795	81,422
2*LogL k – 1 vs. k #par. diff. = 18		11,085.644	1,825.022	1,411.288	398.368	408.166
TECH14 LRT p-value for k-1		.0000	.0000	.0000	.0000	.0000
TECH11 LRT p-value for k-1	NA	.0000	.0000	.0000	.0000	.0019
Entropy	NA	.838	.743	.742	.741	.723

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

Log Likelihood Plot



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)
```

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

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

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;
```

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

```
ANALYSIS:  TYPE = MIXTURE;
MODEL:                                  ! Not needed
           %OVERALL%                    ! Not needed
           %c#1%                        ! Not needed
           [property$1-gambling$1*0];   ! Not needed
           %c#2%                        ! Not needed
           [property$1-gambling$1*1];   ! Not needed
           %c#3%                        ! Not needed
           [property$1-gambling$1*2];   ! Not needed
           %c#4%                        ! Not needed
           [property$1-gambling$1*3];   ! Not needed
OUTPUT:    TECH8 TECH10;
SAVEDATA:  FILE IS asb.sav;
           SAVE IS CPROB;
```


Output Excerpts 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
```

Output Excerpts 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.
```


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

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column)

	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

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

TECHNICAL 10

UNIVARIATE MODEL FIT OF 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

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

BIVARIATE MODEL FIT INFORMATION

Estimated Probabilities

Variable	Variable	H1	H0	Standard Residual
PROPERTY	FIGHT			
Category 1	Category 1	0.635	0.631	0.668
Category 1	Category 2	0.180	0.184	-0.833
Category 2	Category 1	0.084	0.088	-1.141
Category 2	Category 2	0.101	0.097	1.088
PROPERTY	SHOPLIFT			
Category 1	Category 1	0.656	0.646	1.779
Category 1	Category 2	0.159	0.169	-2.269
Category 2	Category 1	0.080	0.090	-2.971
Category 2	Category 2	0.105	0.095	2.904

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

Class 1

Thresholds	Estimates	S. E.	Est./S.E.
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
POT$1	-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
CON$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	0.122	-5.699
GAMBLING$1	1.722	0.125	13.774

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

Class 2

Thresholds	Estimates	S. E.	Est./S.E.
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
POT$1	-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
CON$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	Estimate	S.E.	Est./S.E.
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
POT$1	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
CON$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	Estimate	S.E.	Est./S.E.
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
POT$1	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
CON$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

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

LATENT CLASS INDICATOR MODEL PART IN PROBABILITY SCALE

Class 3

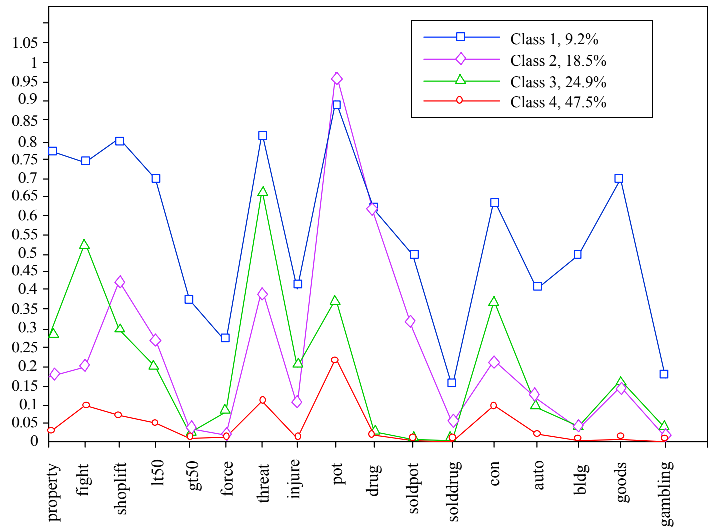
PROPERTY			
Category 2	0.277	0.021	13.321
FIGHT			
Category 2	0.533	0.022	24.193
SHOPLIFT			
Category 2	0.314	0.021	15.120
LT50			
Category 2	0.206	0.018	11.635
GT50			
Category 2	0.034	0.006	5.256
FORCE			
Category 2	0.079	0.008	9.379
THREAT			
Category 2	0.678	0.023	29.703
INJURE			
Category 2	0.188	0.015	12.118
POT			
Category 2	0.362	0.020	17.887

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

LATENT CLASS INDICATOR MODEL PART IN PROBABILITY SCALE

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

4-Class LCA Item Profiles
Antisocial Behavior



LCA Probabilities For Antisocial Behavior (n=7326)

	C#1: Combined class	C#2: Drug class	C#3: Person offense class	C#4: Normative class
Property	**0.78**	0.18	0.28	0.02
Fighting	0.74	0.20	**0.53**	0.09
Shoplifting	**0.82**	0.42	0.31	0.07
Stole < $50	**0.70**	0.27	0.21	0.05
Stole > $50	**0.37**	0.03	0.03	0.00
Use of force	0.26	0.02	0.08	0.01
Seriously threaten	0.82	0.38	**0.68**	0.10
Intent to injure	0.40	0.08	0.19	0.00
Use marijuana	0.90	**0.96**	0.36	**0.22**
Use other drugs	0.60	**0.58**	0.03	0.01
Sold marijuana	0.51	0.26	0.01	0.00
Sold hard drugs	0.14	0.04	0.00	0.00
“Con” someone	0.64	0.20	**0.38**	0.07
Take auto	0.36	0.08	0.11	0.01
Broken into bldg.	**0.43**	0.03	0.04	0.00
Held stolen goods	**0.67**	0.10	0.12	0.00
Gambling operation	0.15	0.02	0.03	0.00
Class Prob.	0.09	0.18	0.25	0.47

Posterior 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
```


Technical Issues Related To LCA

Convergence Behavior – check TECH8 output

- Loglikelihood should increase smoothly and reach a stable maximum

Checking Local Maxima

- Run with more than one set of starting values to see if convergence can be obtained at another set of parameter estimates – done by default using random starts
- Compare loglikelihood values – select solution with the largest value
- The best loglikelihood should be replicated in several solutions

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

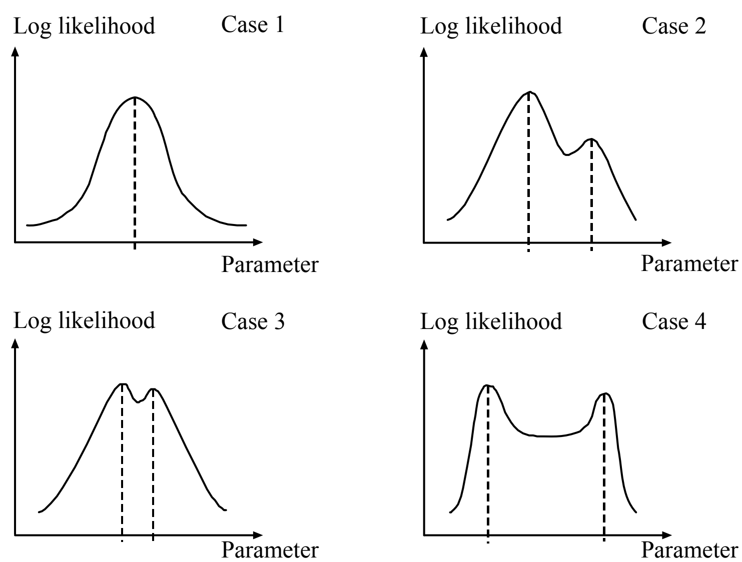
Technical 8 Output

E STEP ITER	LOGLIKELIHOOD	ABS CHANGE	REL CHANGE	CLASS COUNTS	
1	-0.50814249D+05	0.0000000	0.0000000	888.234 2208.576	1659.562 2569.628
2	-0.41810482D+05	9003.7666995	0.1771898	831.174 2165.174	1722.366 2606.555
3	-0.41706620D+05	103.8616123	0.0024841	767.588 2146.235	1807.168 2605.009
4	-0.41657122D+05	49.4986699	0.0011868	714.379 2146.792	1867.660 2597.170
5	-0.41623995D+05	33.1269450	0.0007952	671.621 2162.382	1905.257 2586.740
96	-0.41007499D+05	0.0002095	0.0000000	673.025 1824.820	1354.398 3473.758
97	-0.41007499D+05	0.0001814	0.0000000	672.982 1824.606	1354.419 3473.999
98	-0.41007499D+05	0.0001572	0.0000000	672.906 1824.408	1354.439 3474.211
99	-0.41007499D+05	0.0001362	0.0000000	672.906 1824.222	1354.457 3474.414
100	-0.41007499D+05	0.0001180	0.0000000	672.872 1824.050	1354.475 3474.604

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

153	-0.41007498D+05	0.0000001	0.0000000	672.424 1821.771	1354.725 3477.081
154	-0.41007498D+05	0.0000001	0.0000000	672.423 1821.767	1354.726 3477.085
155	-0.41007498D+05	0.0000000	0.0000000	672.422 1821.764	1354.726 3477.088
171	-0.41007498D+05	0.0000000	0.0000000	672.416 1821.733	1354.730 3477.121
172	-0.41007498D+05	0.0000000	0.0000000	672.416 1821.732	1354.730 3477.122
173	-0.41007498D+05	0.0000000	0.0000000	672.416 1821.731	1354.730 3477.123

Global And Local Solutions



Random Starts

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 2 final stage optimizations which are 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 when there are more than two classes:

STARTS = 50 5;

or with many classes

STARTS = 500 10; STITERATIONS = 20;

Loglikelihood Values At Local Maxima

Results from 10 final stage solutions for ASB example

Good Loglikelihood Behavior: 4-Class LCA			Poor Loglikelihood Behavior: 5-Class LCA		
Loglikelihood	*Seed*	*Initial stage start numbers*	*Loglikelihood*	*Seed*	*Initial stage start numbers*
-41007.498	462953	7	-40808.314	195353	225
-41007.498	608496	4	-40808.406	783165	170
-41007.498	415931	10	-40808.406	863691	481
-41007.498	285380	1	-40815.960	939709	112
-41007.498	93468	3	-40815.960	303634	169
-41007.498	195873	6	-40815.960	85734	411
-41007.498	127215	9	-40815.960	316165	299
-41007.498	253358	2	-40815.960	458181	189
-41010.867	939021	8	-40815.960	502532	445
-41023.043	903420	5	-40816.006	605161	409

- OPTSEED option
- Default STARTS = 10 2 is sufficient for 1-4 classes and 6 classes, but not for 5 classes. 5 classes needs STARTS = 300 10.

Further Readings On Latent Class Analysis

Clogg, C.C. (1995). Latent class models. In G. Arminger, C.C. Clogg & M.E. Sobel (eds.), Handbook of statistical modeling for the social and behavioral sciences (pp. 311-359). New York: Plenum Press.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, 61, 215-231.

Hagenaars, J.A & McCutcheon, A. (2002). Applied latent class analysis. Cambridge: Cambridge University Press.

Nestadt, G., Hanfelt, J., Liang, K.Y., Lamacz, M., Wolyniec, P., & Pulver, A.E. (1994). An evaluation of the structure of schizophrenia spectrum personality disorders. Journal of Personality Disorders, 8, 288-298.

Rindskopf, D., & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. Statistics in Medicine, 5, 21-27.

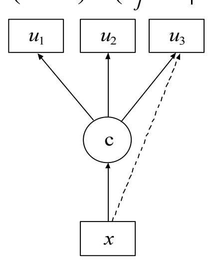
Uebersax, J.S., & Grove, W.M. (1990). Latent class analysis of diagnostic agreement. Statistics in Medicine, 9, 559-572.

Latent Class Analysis With Covariates

LCA With Covariates

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

$$P(u_j = 1) = \sum_{k=1}^{K} P(c = k)\, P(u_j = 1|c = k). \qquad (5)$$



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

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

$$logit[P(c = k|x)] = \alpha_k + \gamma_k\, x. \qquad (7)$$

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*,

$$P(c_i = k \mid x_i) = \frac{e^{\alpha_k + \gamma_k x_i}}{\sum_{s=1}^{K} e^{\alpha_s + \gamma_s x_i}}, \qquad (90)$$

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

$$log[P(c_i = k \mid x_i)/P(c_i = K \mid x_i)] = \alpha_k + \gamma_k x_i . \qquad (91)$$

LCA With Covariates; Multiple-Group LCA

Example: Stouffer & Toby (1951) study of universalistic and particularistic values when confronted with situations involving role conflict. Four dichotomous items.

Respondents divided into thirds (each having $n = 216$) based on who faced the role conflict: Ego, Smith, Close Friend.

Invariance across groups?

Models:

- Invariant measurement characteristics
- Partial measurement invariance
- Measurement invariance and structural invariance

All analyses can be done with groups as covariates (also via KNOWNCLASS)

Source: Clogg, C. & Goodman (1985). Simultaneous latent structure analysis in several groups. In N.B. Tuma (Ed.), Sociological Methodology.

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

```
MODEL:

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

           %c#1%                             !Not needed
           [property$1-goods$1*0];           !Not needed

           %c#2%                             !Not needed
           [property$1-goods$1*1];           !Not needed

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

           %c#4%                             !Not needed
           [property$1-goods$1*3];           !Not needed
OUTPUT:    TECH1 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
```


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

Model Results

```
LATENT CLASS REGRESSION MODEL PART
C#1         ON               Estimates       S.E.    Est./S.E.
   AGE94                         -2.85       .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
```

Calculating Latent Class Probabilities For Different Covariate Values

Consider the multinomial logistic regression logit for a latent 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 = 0 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

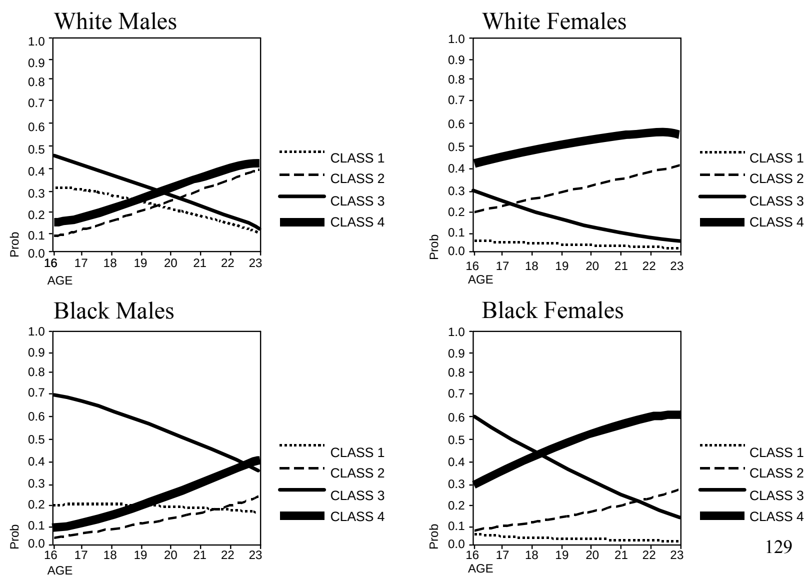
Calculating Latent Class Probabilities For Different Covariate Values (Continued)

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

ASB Classes Regressed On Age, Male, Black In The NLSY (n=7326)



Further Readings On Latent Class Regression Analysis

Bandeen-Roche, K., Miglioretti, D.L., Zeger, S.L. & Rathouz, P.J. (1997). Latent variable regression for multiple discrete outcomes. Journal of the American Statistical Association, 92, 1375-1386.

Clogg, C.C. & Goodman, L.A. (1985). Simultaneous latent structural analysis in several groups. In Tuma, N.B. (ed.), Sociological Methodology, 1985 (pp. 81-110). San Francisco: Jossey-Bass Publishers.

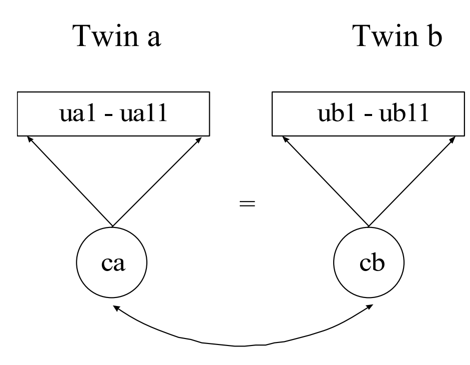
Dayton, C.M. & Macready, G.B. (1988). Concomitant variable latent class models. Journal of the American Statistical Association, 83, 173-178.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. Journal of the American Statistical Association, 87, 476-486.

Confirmatory Latent Class Analysis With Several Latent Class Variables

131

Twin Latent Class Analysis



132

Input For Twin LCA

```
TITLE:      Heath twins; 2,510 complete pairs (5,020 individuals)

DATA:       FILE = twins.dat;
            FORMAT = f9.0 12f3.0 f7.0 2f3.0 f4.0 4f3.0/
                     f9.0 12f3.0 f7.0 2f3.0 f4.0 4f3.0;
VARIABLE:   NAMES = id1 ua1-ua11 male1 xfam1 xid1 zyg891 age1
            majrdep1 conduct1 nalcdep1 alcbuse1
                    id2 ub1-ub11 male2 xfam2 xid2 zyg892 age2
            majrdep2 conduct2 nalcdep2 alcbuse2;
            MISSING = .;
            USEVAR = ua1-ua11 ub1-ub11;
            CATEGORICAL = ua1-ub11;
            CLASSES = ca(5) cb(5);

ANALYSIS:   TYPE = MIXTURE;
            STARTS = 50 5;
            MCONV = .00001;
            STSCALE = 1;
            PARAMETERIZATION = LOGLINEAR;
```


Input For Twin LCA (Continued)

```
MODEL:      %OVERALL%
            ca#1-ca#4 WITH cb#1-cb#4;
            ca#1 WITH cb#2 (991);
            ca#2 WITH cb#1 (991);
            ca#1 WITH cb#3 (992);
            ca#3 WITH cb#1 (992);
            ca#2 WITH cb#3 (993);
            ca#3 WITH cb#2 (993);
            ca#1 WITH cb#4 (994);
            ca#4 WITH cb#1 (994);
            ca#2 WITH cb#4 (995);
            ca#4 WITH cb#2 (995);
            ca#3 WITH cb#4 (996);
            ca#4 WITH cb#3 (996);

            [ca#1] (901);
            [cb#1] (901);
            [ca#2] (902);
            [cb#2] (902);
            [ca#3] (903);
            [cb#3] (903);
            [ca#4] (904);
            [cb#4] (904);
```

Input For Twin LCA (Continued)

```
MODEL ca:
%ca#1%
[ua1$1-ua11$1] (101-111);
.
.
.
%ca#5%
[ua1$1-ua11$1] (501-511);

MODEL cb:
%cb#1%
[ub1$1-ub11$1] (101-111);
.
.
.
%cb#5%
[ub1$1-ub11$1] (501-511);

OUTPUT: TECH1 TECH8 TECH10 STANDARDIZED;

PLOT: TYPE = PLOT3;
SERIES = ua1-ua11(*) | ub1-ub11(*);
```


Output Excerpts: Twin LCA

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS PATTERNS BASED ON THE ESTIMATED MODEL

Latent Class Pattern			
1	1	22.47245	0.00895
1	2	26.35396	0.01050
1	3	30.53171	0.01216
1	4	31.34914	0.01249
1	5	31.82931	0.01268
2	1	26.35396	0.01050
2	2	86.70107	0.03454
2	3	1.67049	0.00067
2	4	35.08729	0.01398
2	5	53.29870	0.02123
3	1	30.53171	0.01216
3	2	1.67049	0.00067
3	3	73.37778	0.02923
3	4	105.55970	0.04206
3	5	31.73730	0.01264

Output Excerpts: Twin LCA (Continued)

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS PATTERNS BASED ON THE ESTIMATED MODEL

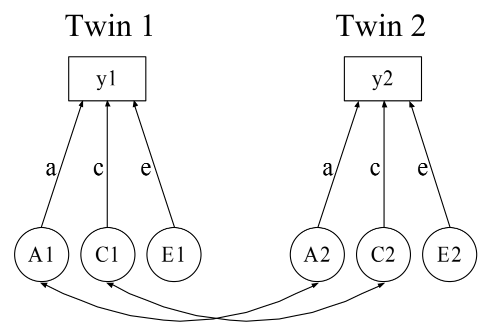
Latent Class Pattern			
4	1	31.34914	0.01249
4	2	35.08729	0.01398
4	3	105.55970	0.04206
4	4	425.40291	0.16948
4	5	296.48193	0.11812
5	1	31.82931	0.01268
5	2	53.29870	0.02123
5	3	31.73730	0.01264
5	4	296.48193	0.11812
5	5	614.24676	0.24472

Output Excerpts: Twin LCA (Continued)

FINAL CLASS COUNTS AND PROPORTIONS FOR EACH LATENT CLASS VARIABLE BASED ON THE ESTIMATED MODEL

Latent Class Variable	Class		
CA	1	142.53658	0.05679
	2	203.11148	0.08092
	3	242.87698	0.09676
	4	893.88098	0.35613
	5	1027.59399	0.40940
CA	1	142.53658	0.05679
	2	203.11148	0.08092
	3	242.87698	0.09676
	4	893.88098	0.35613
	5	1027.59399	0.40940

Classic Twin Modeling



1.0 for MZ, 0.5 for DZ 1.0

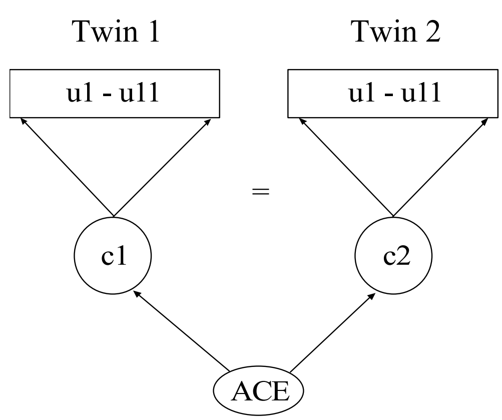
ACE Model

- Continuous or categorical outcome
- MZ, DZ twins jointly in 2-group analysis

$$\Sigma_{DZ} = \begin{bmatrix} a^2 + c^2 + e^2 & \text{symm.} \\ 0.5 \text{ x } a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix} \quad \Sigma_{MZ} = \begin{bmatrix} a^2 + c^2 + e^2 & \text{symm.} \\ a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}$$

For Mplus inputs, see User's Guide ex5.18, ex5.21

ACE Latent Class Analysis



Second-order LCA (c = class)

- Conventional approach: 3 steps – LCA, classification, ACE (or twin concordance)
- New approach: 1 step (latent class variables regressed on continuous latent variables)

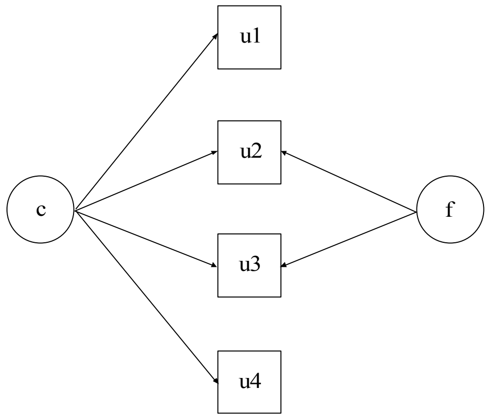
Further Readings On Twin LCA

Muthen, Asparouhov & Rebollo (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. Twin Research and Human Genetics, 9, 313-324.

Prescott, C.A. (2004). Using the Mplus computer program to estimate models for continuous and categorical data from twins. Behavior Genetics, 34, 17- 40.

Latent Class Analysis With A Random Effect

Latent Class Analysis With A Random Effect



Input For A Latent Class Analysis With A Random Effect

```
TITLE:      Example of a latent class analysis from Qu-Tan-
            Kutner (1996), Biometrics, 52, 797-810 Example 1.
            2LCD model: random effect for u2 and u3 in class 1

DATA:       FILE IS alvordhiv1.dat;

VARIABLE:   NAMES ARE u1-u4 id;
            USEV ARE u1-u4;
            CATEGORICAL ARE u1-u4;
            CLASSES = c (2);

ANALYSIS:   TYPE = MIXTURE;
            ALGORITHM = INTEGRATION;
```

Input For A Latent Class Analysis With A Random Effect (Continued)

```
MODEL:      %OVERALL%
            f BY u2-u3@0;
            f@1; [f@0];
            %c#1%
            [u1$1@-15 u4$1@-15];
            f BY u2-u3*1 (1);

OUTPUT:     SAMPSTAT TECH1 TECH8 TECH10;
```


Output Excerpts A Latent Class Analysis With A Random Effect

Tests Of Model Fit

```
Loglikelihood

             H0 Value                          -623.299
Information Criteria

             Number of Free Parameters                7
             Akaike (AIC)                      1260.599
             Bayesian (BIC)                    1289.013
             Sample-Size Adjusted BIC          1266.799
                (n* = (n + 2) / 24)
             Entropy                              0.993
```

Output Excerpts A Latent Class Analysis With A Random Effect (Continued)

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

             Value                              4.487
             Degrees of Freedom                     8
             P-Value                           0.8107

             Likelihood Ratio Chi-Square

             Value                              3.057
             Degrees of Freedom                     8
             P-Value                           0.9307
```


Output Excerpts A Latent Class Analysis With A Random Effect (Continued)

Final Class Counts

```
FINAL  CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE BASED
ON ESTIMATED POSTERIOR PROBABILITIES

 Class 1         231.57388          0.54106
 Class 2         196.42612          0.45894
```

Output Excerpts A Latent Class Analysis With A Random Effect (Continued)

Technical 10 Output

MODEL FIT INFORMATION FOR THE LATENT CLASS INDICATOR MODEL PART

RESPONSE PATTERNS

No.	Pattern	No.	Pattern	No.	Pattern	No.	Pattern
1	0000	2	1000	3	0100	4	1100
5	0001	6	1001	7	1101	8	1011
9	1111						

Output Excerpts A Latent Class Analysis With A Random Effect (Continued)

RESPONSE PATTERN FREQUENCIES AND CHI-SQUARE CONTRIBUTIONS

Response Pattern	Frequency Observed	Frequency Estimated	Standard Residual	Chi-square Pearson	Contribution Loglikelihood	Deleted
1	170.00	169.71	0.03	0.00	0.58	
2	4.00	4.82	0.38	0.14	-1.50	
3	6.00	6.29	0.11	0.01	-0.56	
4	1.00	0.18	1.94	3.78	3.44	
5	15.00	14.47	0.14	0.02	1.09	
6	17.00	16.94	0.02	0.00	0.13	
7	4.00	4.04	0.02	0.00	-0.09	
8	83.00	83.05	0.01	0.00	-0.10	
9	128.00	127.97	0.00	0.00	0.06	

THE TOTAL PEARSON CHI-SQUARE CONTRIBUTION FROM EMPTY CELLS IS 0.54

Output Excerpts A Latent Class Analysis With A Random Effect (Continued)

Bivariate Model Fit Information

Variable	Variable	Estimated Probabilities H1	H0	Standard Residual
U1	U2			
Category 1	Category 1	0.432	0.430	0.080
Category 1	Category 2	0.014	0.016	-0.317
Category 2	Category 1	0.243	0.245	-0.091
Category 2	Category 2	0.311	0.309	0.084
U1	U3			
Category 1	Category 1	0.446	0.446	0.000
Category 1	Category 2	0.000	0.000	-0.011
Category 2	Category 1	0.061	0.061	0.004
Category 2	Category 2	0.493	0.493	-0.002
U1	U4			
Category 1	Category 1	0.411	0.411	0.000
Category 1	Category 2	0.035	0.035	0.000
Category 2	Category 1	0.012	0.012	-0.001
Category 2	Category 2	0.542	0.542	0.000

Output Excerpts A Latent Class Analysis With A Random Effect (Continued)

U2	U3			
Category 1	Category 1	0.481	0.481	0.006
Category 1	Category 2	0.194	0.194	-0.006
Category 2	Category 1	0.026	0.026	-0.013
Category 2	Category 2	0.299	0.299	0.003
U2	U4			
Category 1	Category 1	0.407	0.408	-0.053
Category 1	Category 2	0.269	0.267	0.060
Category 2	Category 1	0.016	0.015	0.212
Category 2	Category 2	0.308	0.310	-0.058
U3	U4			
Category 1	Category 1	0.423	0.423	0.000
Category 1	Category 2	0.084	0.084	0.003
Category 2	Category 1	0.000	0.000	-0.011
Category 2	Category 2	0.493	0.493	-0.002

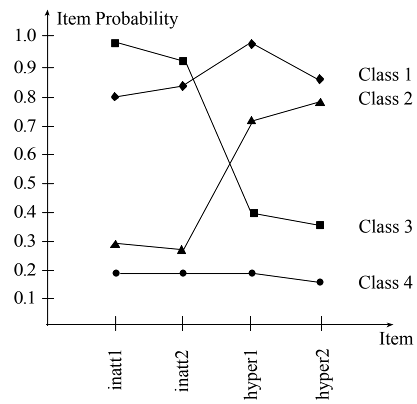
Modeling With A Combination Of Continuous And Categorical Latent Variables

Modeling With A Combination Of Continuous And Categorical Latent Variables

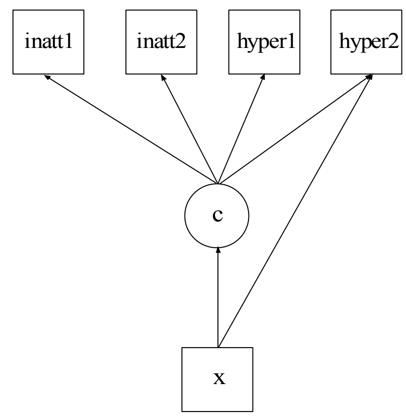
- Factor mixture analysis
  - Generalized factor analysis
  - Generalized latent class analysis
- Structural equation mixture modeling
- Growth mixture modeling

Latent Class Analysis

a. Item Profiles

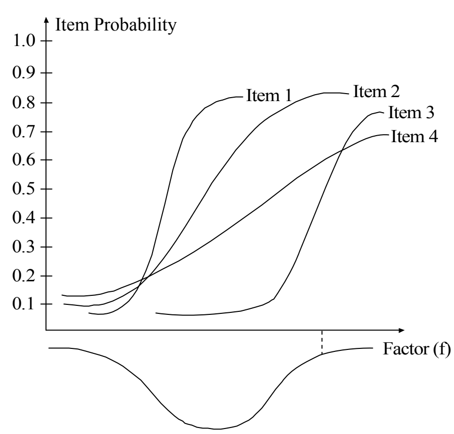


b. Model Diagram

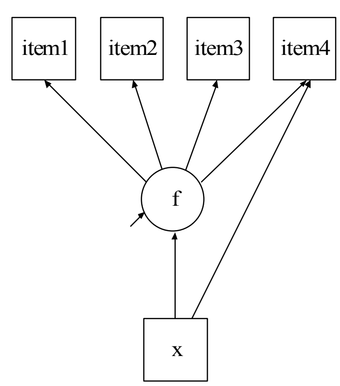


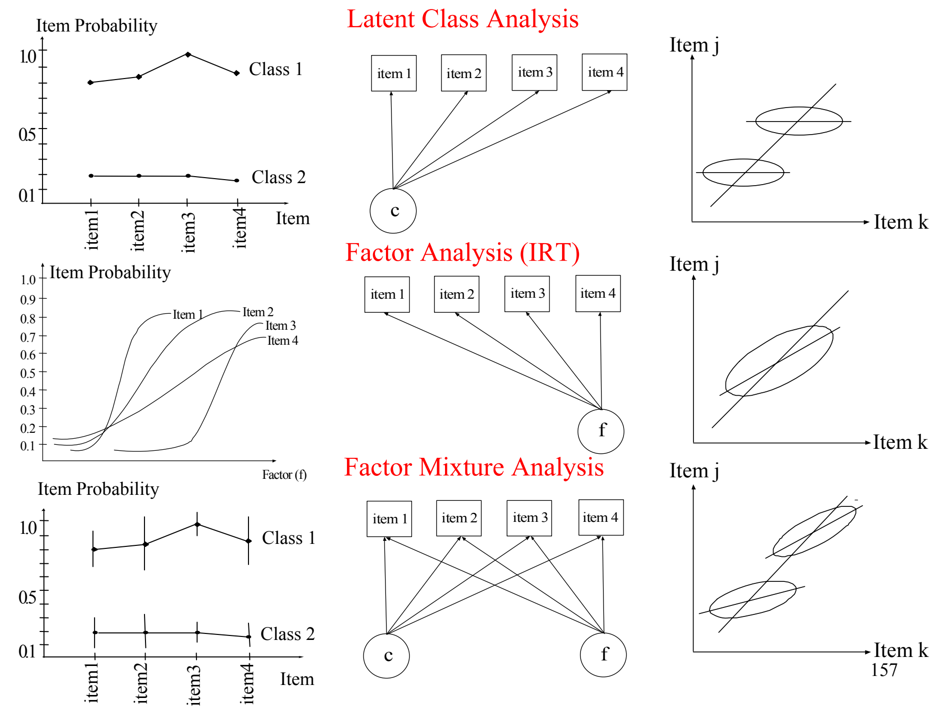
Factor Analysis (IRT, Latent Trait)

a. Item Response Curves



b. Model Diagram





Latent Class, Factor, And Factor Mixture Analysis Alcohol Dependence Criteria, NLSY 1989 (n = 8313)

Source: Muthén & Muthén (1995)

	Latent Classes				
	Two-class solution[1]		Three-class solution[2]		
	I	II	I	II	III
Prevalence	0.78	0.22	0.75	0.21	0.03
DSM-III-R Criterion	Conditional Probability of Fulfilling a Criterion				
Withdrawal	0.00	0.14	0.00	0.07	0.49
Tolerance	0.01	0.45	0.01	0.35	0.81
Larger	0.15	0.96	0.12	0.94	0.99
Cut down	0.00	0.14	0.01	0.05	0.60
Time spent	0.00	0.19	0.00	0.09	0.65
Major role-Hazard	0.03	0.83	0.02	0.73	0.96
Give up	0.00	0.10	0.00	0.03	0.43
Relief	0.00	0.08	0.00	0.02	0.40
Continue	0.00	0.24	0.02	0.11	0.83

[1]Likelihood ratio chi-square fit = 1779 with 492 degrees of freedom
[2]Likelihood ratio chi-square fit = 448 with 482 degrees of freedom

LCA, FA, And FMA For NLSY 1989

- LCA, 3 classes: logL = -14,139, 29 parameters, BIC = 28,539
- FA, 2 factors: logL = -14,083, 26 parameters, BIC = 28,401
- FMA 2 classes, 1 factor, loadings invariant:
  logL = -14,054, 29 parameters, BIC = 28,370

Models can be compared with respect to fit to the data

- Standardized bivariate residuals
- Standardized residuals for most frequent response patterns

Estimated Frequencies And Standardized Residuals

Obs Freq.	LCA 3c		FA 2f		FMA 1f, 2c	
	Est. Freq.	Res.	Est. Freq.	Res.	Est. Freq.	Res.
5335	5332	-0.07	5307	-0.64	5331	-0.08
941	945	0.12	985	1.48	946	0.18
601	551	**-2.22**	596	-0.22	606	0.21
217	284	**4.04**	211	-0.42	228	0.75
155	111	**-4.16**	118	**-3.48**	134	1.87
149	151	0.15	168	1.45	147	0.17
65	68	0.41	46	**-2.79**	53	1.60
49	52	0.42	84	**3.80**	59	1.27
48	54	0.81	44	-0.61	46	0.32
47	40	-1.09	45	-0.37	45	0.33

Bolded entries are significant at the 5% level.

Input For FMA Of 9 Alcohol Items In The NLSY 1989

```
TITLE:          Alcohol LCA M & M (1995)

DATA:           FILE = bengt05_spread.dat;

VARIABLE:       NAMES = u1-u9;

                CATEGORICAL = u1-u9;

                CLASSES = c(2);

ANALYSIS:       TYPE = MIXTURE;

                ALGORITHM = INTEGRATION;

                STARTS =  200 10; STITER = 20;

                ADAPTIVE = OFF;

                PROCESS = 4;
```


Input For FMA Of 9 Alcohol Items In The NLSY 1989 (Continued)

```
MODEL:          OVERALL%

                f BY u1-u9;
                f*1; [f@0];

                %c#1%
                [u1$1-u9$1];
                f*1;

                %c#2%
                [u1$1-u9$1];
                f*1;

OUTPUT:         TECH1 TECH8 TECH10;

PLOT:           TYPE = plot3;
                SERIES = u1-u9(*);
```

Output Excerpts FMA Of 9 Alcohol Items In The NLSY 1989

Latent Class 1

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
f BY				
u1	1.000	0.000	999.000	999.000
u2	0.769	0.067	11.418	0.000
u3	0.867	0.115	7.540	0.000
u4	1.157	0.131	8.839	0.000
u5	1.473	0.191	7.727	0.000
u6	0.998	0.105	9.533	0.000
u7	1.037	0.102	10.176	0.000
u8	1.229	0.152	8.097	0.000
u9	1.597	0.264	6.058	0.000

Output Excerpts FMA Of 9 Alcohol Items In The NLSY 1989 (Continued)

Latent Class 1

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Means				
f	0.000	0.000	999.000	999.000
Thresholds				
u1$1	3.332	0.357	9.334	0.000
u2$1	1.269	0.243	5.223	0.000
u3$1	-2.367	0.399	-5.929	0.000
u4$1	4.586	0.453	10.128	0.000
u5$1	3.482	0.500	6.967	0.000
u6$1	-0.398	0.391	-1.017	0.309
u7$1	4.665	0.363	12.851	0.000

Output Excerpts FMA Of 9 Alcohol Items In The NLSY 1989 (Continued)

Latent Class 1

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
u6$1	4.767	0.440	10.825	0.000
u7$1	4.782	0.910	5.256	0.000
Variances				
f	2.636	0.610	4.323	0.000

Output Excerpts FMA Of 9 Alcohol Items In The NLSY 1989 (Continued)

Latent Class 2

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
f BY				
u1	1.000	0.000	999.000	999.000
u2	0.769	0.067	11.418	0.000
u3	0.867	0.115	7.540	0.000
u4	1.157	0.131	8.839	0.000
u5	1.473	0.191	7.727	0.000
u6	0.998	0.105	9.533	0.000
u7	1.037	0.102	10.176	0.000
u8	1.229	0.152	8.097	0.000
u9	1.597	0.264	6.058	0.000

Output Excerpts FMA Of 9 Alcohol Items In The NLSY 1989 (Continued)

Latent Class 2

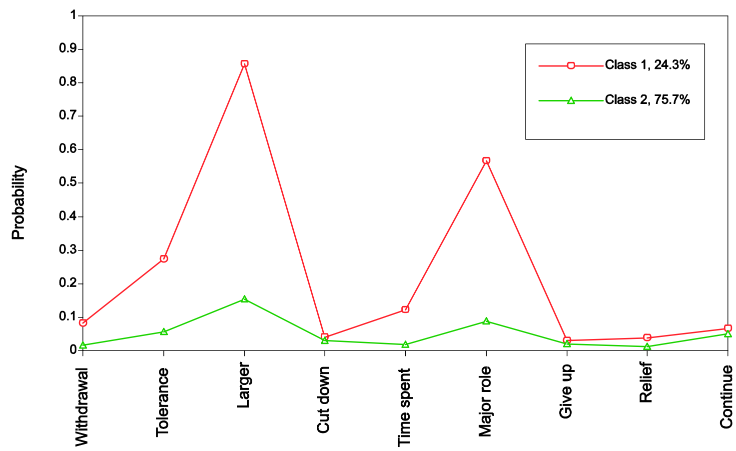
Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Means				
f	0.000	0.000	999.000	999.000
Thresholds				
u1$1	10.486	1.074	9.764	0.000
u2$1	6.286	0.632	9.942	0.000
u3$1	4.392	1.060	4.142	0.000
u4$1	10.516	0.923	11.393	0.000
u5$1	14.730	1.995	7.382	0.000
u6$1	6.635	1.054	6.296	0.309
u7$1	10.365	0.922	11.243	0.000

Output Excerpts FMA Of 9 Alcohol Items In The NLSY 1989 (Continued)

Latent Class 2

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
u6$1	13.241	1.842	7.188	0.000
u7$1	12.278	2.055	5.974	0.000
Variances				
f	20.569	4.559	4.511	0.000
Categorical Latent Variables				
Means				
c#1	-1.135	0.265	-4.290	0.000

FMA Profile Plot



Factor (IRT) Mixture Example: The Latent Structure Of ADHD

- UCLA clinical sample of 425 males ages 5-18, all with ADHD diagnosis
- Subjects assessed by clinicians:
  1) direct interview with child (> 7 years),
  2) interview with mother about child
- KSADS: Nine inattentiveness items, nine hyperactivity items; dichotomously scored
- Families with at least 1 ADHD affected child
- Parent data, candidate gene data on sib pairs
- What types of ADHD does a treatment population show?

The Latent Structure Of ADHD (Continued)

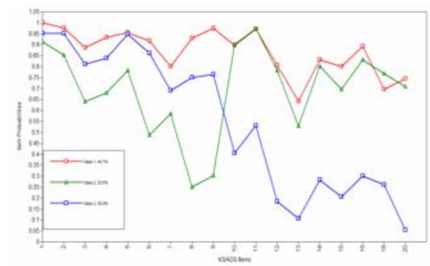
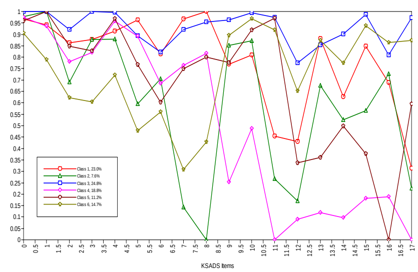
Inattentiveness Items:	Hyperactivity Items:
'Difficulty sustaining attn on tasks/play'	'Difficulty remaining seated'
'Easily distracted'	'Fidgets'
'Makes a lot of careless mistakes'	'Runs or climbs excessively'
'Doesn't listen'	'Difficulty playing quietly'
'Difficulty following instructions'	'Blurts out answers'
'Difficulty organizing tasks'	'Difficulty waiting turn'
'Dislikes/avoids tasks'	'Interrupts or intrudes'
'Loses things'	'Talks excessively'
'Forgetful in daily activities'	'Driven by motor'

The Latent Structure Of ADHD: Model Results

Model	Likelihood	# Parameters	BIC	BLRT p value for k-1 classes
LCA – 2c	-3650	37	7523	0.
LCA – 3c	**-3545**	**56**	**7430**	**0.**
LCA – 4c	-3499	75	7452	0.
LCA – 5c	-3464	94	7496	0.
LCA – 6c	**-3431**	**113**	**7547**	**0.**
LCA – 7c	-3413	132	7625	0.27

LCA-3c is best by BIC and LCA-6c is best by BLRT

Three-Class And Six-Class LCA Item Profiles

LCA – 3c	LCA – 6c
	

The Latent Structure Of ADHD: Model Results

Model	Likelihood	# Parameters	BIC	BLRT p value for k-1 classes
LCA – 2c	-3650	37	7523	0.
LCA – 3c	**-3545**	**56**	**7430**	**0.**
LCA – 4c	-3499	75	7452	0.
LCA – 5c	-3464	94	7496	0.
LCA – 6c	**-3431**	**113**	**7547**	**0.**
LCA – 7c	-3413	132	7625	0.27
EFA – 2f	**-3505**	**53**	**7331**	

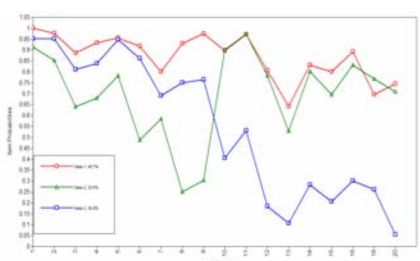
The EFA model is better than LCA - 3c, but no classification of individuals is obtained

The Latent Structure Of ADHD: Model Results

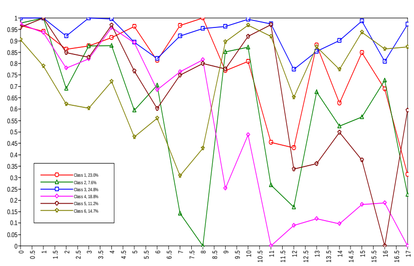
Model	Likelihood	# Parameters	BIC	BLRT p value for k-1
LCA – 2c	-3650	37	7523	0.
LCA – 3c	**-3545**	**56**	**7430**	**0.**
LCA – 4c	-3499	75	7452	0.
LCA – 5c	-3464	94	7496	0.
LCA – 6c	**-3431**	**113**	**7547**	**0.**
LCA – 7c	-3413	132	7625	0.27
EFA – 2f	**-3505**	**53**	**7331**	
FMA – 2c, 2f	-3461	59	7280	
FMA – 2c, 2f Class-varying Factor loadings	**-3432**	**75**	**7318**	**χ^2-diff (16) = 58 $p < 0.01$**

Item Profiles For Three-Class LCA, Six-Class LCA And Two-Class, Two-Factor FMA

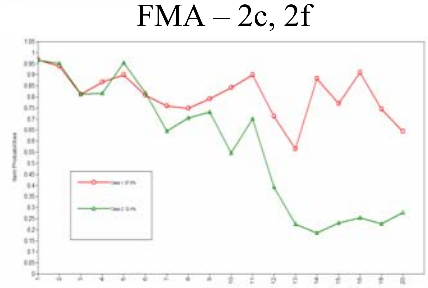
LCA – 3c



LCA – 6c



FMA – 2c, 2f



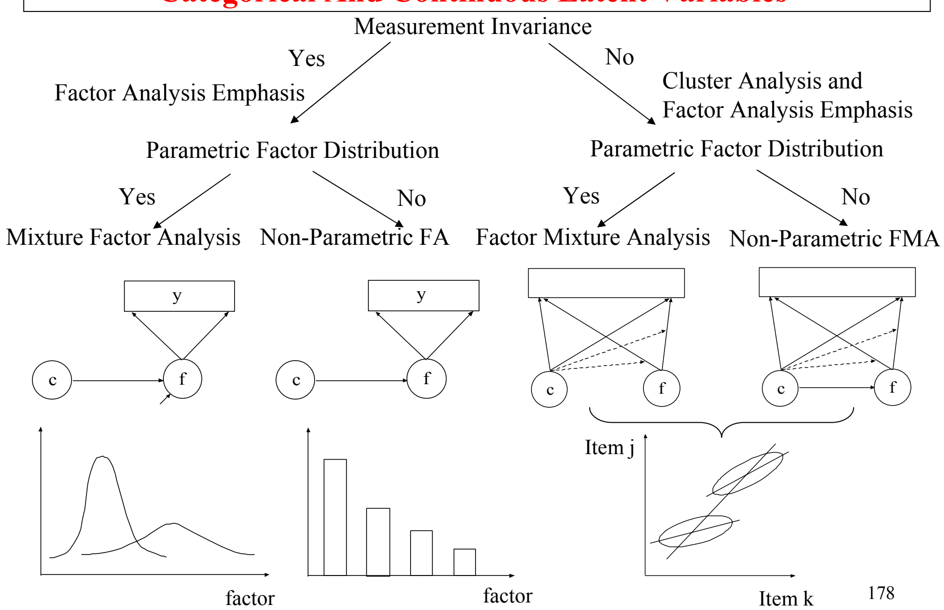
Factor Mixture Modeling Issues

Categorical outcomes plus continuous-normal latent variables have the computational and statistical disadvantage of

- heavy computations due to numerical integration
- normality assumption

Non-parametric latent variable distribution avoids the normality assumption and at the same time the computational disadvantage!

Overview Of Cross-Sectional Hybrids: Modeling With Categorical And Continuous Latent Variables



Further Readings On Factor Mixture Analysis

Lubke, G. & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. Structural Equation Modeling, 14(1), 26-47.

McLachlan, Do & Ambroise (2004). Analyzing microarray gene expression data. Wiley.

Muthen (2006). Should substance use disorders be considered as categorical or dimensional? Addiction, 101, 6-16.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 1-24. Charlotte, NC: Information Age Publishing, Inc.

Muthen & Asparouhov (2006). Item response mixture modeling: Application to tobacco dependence criteria. Addictive Behaviors, 31, 1050-1066.

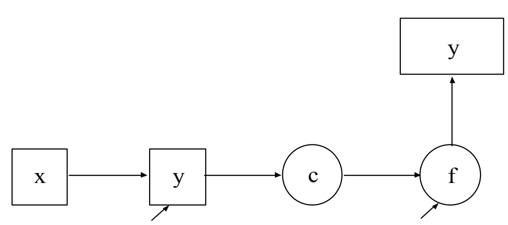
Muthen, Asparouhov & Rebollo (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. Twin Research and Human Genetics, 9, 313-324.

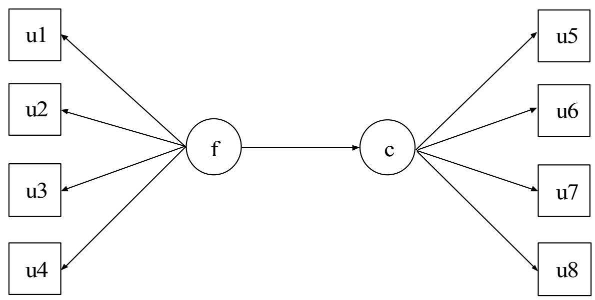
EFA Mixture Analysis

- Class-varying intercepts\thresholds, loading matrices, residual variances, and factor correlation matrices.
- User's Guide example 4.4:

```
TITLE:      This is an example of an exploratory factor mixture
            analysis with continuous latent class indicators
DATA:       FILE = ex4.4.dat;
VARIABLE:   NAMES = y1-y8;
            CLASSES = c(2);
ANALYSIS:   TYPE = MIXTURE EFA 1 2;
```

Structural Equation Mixture Modeling





Input For A Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators

```
TITLE:      this is an example of a latent class model with two
            classes influenced by a continuous latent variable
            with categorical indicators
DATA:       FILE = firstcetaMC.dat;
VARIABLE:   NAMES ARE u1-u8 c;
            USEV = u1-u8;
            CATEGORICAL = u1-u8;
            CLASSES = c(2);
ANALYSIS:   TYPE = MIXTURE;
            ALGORITHM = INTEGRATION’
            STARTS = 100 5;
MODEL:
            %OVERALL&
            f BY u1-u4;
            c#1 ON f;
OUTPUT:     TECH1 TECH8;
```

Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators

Random Starts

```
Preliminary loglikelihood values and seeds:

                  -5362.831   68985
                  -5363.809   392418
                  -5364.300   851945
                  -5364.532   939021
                  -5364.594   848890
                  -5364.747   696773
                  -5364.750   963053
                  -5364.874   754100
                  -5365.008   415931
```


Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators (Continued)

```
                                .
                                .
                                .
                  -5405.339   311214
                  -5406.840   761633
                  -5407.270   575700
                  -5408.512   391179
                  -5411.759   284109
                  -5417.058   462953
                  -5417.195   347515
                  -5425.535   285380
```

Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators (Continued)

```
Loglikelihood values at local maxima and seeds:

                  -5346.581   851945
                  -5346.583   848890
                  -5346.583   392418
                  -5346.591   939021
                  -5346.603   68985

Unperturbed starting value run did not converge.
```


Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators (Continued)

Tests Of Model Fit

```
Loglikelihood

             H0 Value                          -5346.581
Information Criteria

             Number of Free Parameters                18
             Akaike (AIC)                      10729.161
             Bayesian (BIC)                    10817.501
             Sample-Size Adjusted BIC          10760.332
                (n* = (n + 2) / 24)
             Entropy                               0.511
```

Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators (Continued)

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

          Pearson Chi-Square

          Value                           281.968
          Degrees of Freedom                  238
          P-Value                          0.0266

          Likelihood Ratio Chi-Square

          Value                           296.114
          Degrees of Freedom                  238
          P-Value                          0.0062
```


Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators (Continued)

Final Class Counts

```
FINAL  CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE BASED
ON ESTIMATED POSTERIOR PROBABILITIES

 Class 1        450.85360          0.45085
 Class 2        549.14640          0.54915
```

Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators (Continued)

	Estimates	S.E.	Est./S.E.
Class 1			
F BY			
U1	1.000	0.000	0.000
U2	0.786	0.185	4.258
U3	0.760	0.220	3.461
U4	1.019	0.233	4.381
Variances			
F	1.137	0.379	3.001
Class 2			
F BY			
U1	1.000	0.000	0.000
U2	0.786	0.185	4.258
U3	0.760	0.220	3.461
U4	1.019	0.233	4.381
Variances			
F	1.137	0.379	3.001

Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators (Continued)

	Estimates	S.E.	Est./S.E.
Class 1			
Thresholds			
U1$1	0.089	0.078	1.136
U2$1	-0.028	0.073	-0.379
U3$1	0.027	0.072	0.379
U4$1	0.020	0.079	0.253
U5$1	-1.126	0.177	-6.356
U6$1	-1.066	0.212	-5.040
U7$1	-0.919	0.174	-5.289
U8$1	-1.156	0.175	-6.622

Output Excerpts Latent Class Model With Two Classes Influenced By A Continuous Latent Variable With Categorical Indicators (Continued)

```
                    Estimates       S.E.  Est./S.E.
 Class 2
  Thresholds
    U1$1              0.089      0.078     1.136
    U2$1             -0.028      0.073    -0.379
    U3$1              0.027      0.072     0.379
    U4$1              0.020      0.079     0.253
    U5$1              1.058      0.179     5.909
    U6$1              1.020      0.141     7.247
    U7$1              0.839      0.133     6.316
    U8$1              0.979      0.171     5.727

 LATENT CLASS REGRESSION MODEL PART
  C#1      ON
    F                 0.684      0.242     2.829
 Intercepts
    C#1              -0.220      0.202    -1.089
```


Further Readings On Structural Equation Mixture Modeling

Guo, J., Wall, M. & Amemiya, Y. (2006). Latent class regression on latent factors. Biostatistics, 7, 145 - 163.

Jedidi, K., Jagpal. H.S. & DeSarbo, W.S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. Marketing Science, 16, 39-59.

Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. Psychometrika, 62, 297-330.

References

(To request a Muthén paper, please email bmuthen@ucla.edu and refer to the number in parenthesis.)

Analysis With Categorical Outcomes

General

Agresti, A. (2002). Categorical data analysis. Second edition. New York: John Wiley & Sons.

Agresti, A. (1996). An introduction to categorical data analysis. New York: Wiley.

Hosmer, D.W. & Lemeshow, S. (2000). Applied logistic regression. Second edition. New York: John Wiley & Sons.

McKelvey, R.D. & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. Journal of Mathematical Sociology, 4, 103-120.

Censored and Poisson Regression

Hilbe (2007). Negative Binomial Regression. Cambridge.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34, 1-13.

References (Continued)

Long, S. (1997). Regression models for categorical and limited dependent variables. Thousand Oaks: Sage.

Maddala, G.S. (1983). Limited-dependent and qualitative variables in econometrics. Cambridge: Cambridge University Press.

Tobin, J (1958). Estimation of relationships for limited dependent variables. Econometrica, 26, 24-36.

IRT

Baker, F.B. & Kim, S.H. (2004). Item response theory. Parameter estimation techniques. Second edition. New York: Marcel Dekker.

Bock, R.D. (1997). A brief history of item response theory. Educational Measurement: Issues and Practice, 16, 21-33.

du Toit, M. (2003). IRT from SSI. Lincolnwood, IL: Scientific Software International, Inc. (BILOG, MULTILOG, PARSCALE, TESTFACT)

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory. Boston: Kluwer-Nijhoff.

MacIntosh, R. & Hashim, S. (2003). Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. Applied Psychological Measurement, 27, 372-379.

References (Continued)

Muthén, B. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. Journal of Educational Statistics, 10, 121-132. (#13)

Muthén, B. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer & H. Braun (Eds.), Test Validity (pp. 213-238). Hillsdale, NJ: Erlbaum Associates. (#18)

Muthén, B. (1989). Using item-specific instructional information in achievement modeling. Psychometrika, 54, 385-396. (#30)

Muthén, B. (1994). Instructionally sensitive psychometrics: Applications to the Second International Mathematics Study. In I. Westbury, C. Ethington, L. Sosniak & D. Baker (Eds.), In search of more effective mathematics education: Examining data from the IEA second international mathematics study (pp. 293-324). Norwood, NJ: Ablex. (#54)

Muthén, B. & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus Web Note #4 (www.statmodel.com).

Muthén, B., Kao, Chih-Fen & Burstein, L. (1991). Instructional sensitivity in mathematics achievement test items: Applications of a new IRT-based detection technique. Journal of Educational Measurement, 28, 1-22. (#35)

References (Continued)

Muthén, B. & Lehman, J. (1985). Multiple-group IRT modeling: Applications to item bias analysis. Journal of Educational Statistics, 10, 133-142. (#15)

Takane, Y. & DeLeeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. Psychometrika, 52, 393-408.

Factor Analysis

Bartholomew, D.J. (1987). Latent variable models and factor analysis. New York: Oxford University Press.

Bock, R.D., Gibbons, R., & Muraki, E.J. (1988). Full information item factor analysis. Applied Psychological Measurement, 12, 261-280.

Blafield, E. (1980). Clustering of observations from finite mixtures with structural information. Unpublished doctoral dissertation, Jyvaskyla studies in computer science, economics, and statistics, Jyvaskyla, Finland.

Flora, D.B. & Curran P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods, 9, 466-491.

References (Continued)

Lord, F.M. & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley Publishing Co.

Millsap, R.E. & Yun-Tien, J. (2004). Assessing factorial invariance in ordered-categorical measures. Multivariate Behavioral Research, 39, 479-515.

Mislevy, R. (1986). Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-31.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. Psychometrika, 43, 551-560. (#3)

Muthén, B. (1989). Dichotomous factor analysis of symptom data. In Eaton & Bohrnstedt (Eds.), Latent variable models for dichotomous outcomes: Analysis of data from the Epidemiological Catchment Area program (pp. 19-65), a special issue of Sociological Methods & Research, 18, 19-65. (#21)

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585. (#24)

Muthén, B. (1996). Psychometric evaluation of diagnostic criteria: Application to a two-dimensional model of alcohol abuse and dependence. Drug and Alcohol Dependence, 41, 101-112. (#66)

References (Continued)

Muthén, B. & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus Web Note #4 (www.statmodel.com).

Muthén, B. & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. Psychometrika, 46, 407-419. (#6)

Muthén, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.

Muthén, B. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. British Journal of Mathematical and Statistical Psychology, 45, 19-30.

Muthén, B. & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. Psychometrika, 60, 489-503.

Takane, Y. & DeLeeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. Psychometrika, 52, 393-408.

Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. Psychometrika, 62, 297-330.

References (Continued)

MIMIC

Gallo, J.J., Anthony, J. & Muthén, B. (1994). Age differences in the symptoms of depression: a latent trait analysis. Journals of Gerontology: Psychological Sciences, 49, 251-264. (#52)

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585. (#24)

Muthén, B., Tam, T., Muthén, L., Stolzenberg, R.M. & Hollis, M. (1993). Latent variable modeling in the LISCOMP framework: Measurement of attitudes toward career choice. In D. Krebs & P. Schmidt (Eds.), New directions in attitude measurement, Festschrift for Karl Schuessler (pp. 277-290). Berlin: Walter de Gruyter. (#46)

SEM

Browne, M.W. & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C.C. Clogg & M.E. Sobel (Eds.), Handbook of statistical modeling for the social and behavioral sciences (pp. 311-359). New York: Plenum Press.

Muthén, B. (1979). A structural probit model with latent variables. Journal of the American Statistical Association, 74, 807-811. (#4)

References (Continued)

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. Journal of Econometrics, 22, 48-65. (#9)

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132. (#11)

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585. (#24)

Muthén, B. (1993). Goodness of fit with categorical and other non-normal variables. In K.A. Bollen, & J.S. Long (Eds.), Testing structural equation models (pp. 205-243). Newbury Park, CA: Sage. (#45).

Muthén, B. & Speckart, G. (1983). Categorizing skewed, limited dependent variables: Using multivariate probit regression to evaluate the California Civil Addict Program. Evaluation Review, 7, 257-269. (#3)

Muthén, B. du Toit, S.H.C. & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Accepted for publication in Psychometrika. (#75)

Prescott, C.A. (2004). Using the Mplus computer program to estimate models for continuous and categorical data from twins. Behavior Genetics, 34, 17-40.

References (Continued)

Xie, Y. (1989). Structural equation models for ordinal variables. Sociological Methods & Research, 17, 325-352.

Yu, C.-Y. & Muthén, B. (2002). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Technical report.

Analysis With Categorical Latent Variables (Mixture Modeling)

General

Albert, P.A., McShane, L. M., Shih, J.H. & The US NCI Bladder Tumor Marker Network (2001). Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p53 immunohistochemical assays in bladder tumors. Biometrics, 57, 610-619.

Clark, S. & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. Submitted for publication.

Clark, S. & Muthén, B. (2009). Models and strategies for factor mixture analysis: Three examples concerning the structure underlying psychological disorders.

Everitt, B.S. & Hand, D.J. (1981). Finite mixture distributions. London: Chapman and Hall.

References (Continued)

Follmann, D.A. & Lambert, D. (1989). Generalizing logistic regression by nonparametric mixing. Journal of American Statistical Association, 84, 295-300.

Hox, J. & Lensvelt-Mulders, G. (2004). Randomized response analysis in Mplus. Structural Equation Modeling, 11, 615-620.

Lo, Y., Mendell, N.R. & Rubin, D.B. (2001). Testing the number of components in a normal mixture. Biometrika, 88, 767-778.

Lubke, G. & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. Structural Equation Modeling, 14(1), 26-47.

McLachlan, G.J. & Peel, D. (2000). Finite mixture models. New York: Wiley & Sons.

McLachlan, Do & Ambroise (2004). Analyzing microarray gene expression data. Wiley.

Hagenaars, J.A & McCutcheon, A. (2002). Applied latent class analysis. Cambridge: Cambridge University Press.

McLachlan, G.J. & Peel, D. (2000). Finite mixture models. New York: Wiley & Sons.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. Behaviormetrika, 29, 81-117. (#96)

References (Continued)

Muthén (2006). Should substance use disorders be considered as categorical or dimensional? Addiction, 101, 6-16.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 1-24. Charlotte, NC: Information Age Publishing, Inc.

Muthen & Asparouhov (2006). Item response mixture modeling: Application to tobacco dependence criteria. Addictive Behaviors, 31, 1050-1066.

Muthen, Asparouhov & Rebollo (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. Twin Research and Human Genetics, 9, 313-324.

Muthén, L.K. & Muthén, B. (1998-2006). Mplus User's Guide. Los Angeles, CA: Muthén & Muthén.

Schwartz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6, 461-464.

Titterington, D.M., Smith, A.F.M., & Makov, U.E. (1985). Statistical analysis of finite mixture distributions. Chichester, U.K.: John Wiley & Sons.

References (Continued)

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, 57, 307-333.

Noncompliance (CACE)

Angrist, J.D., Imbens, G.W., Rubin. D.B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91, 444-445.

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. Journal of Educational and Behavioral Statistics, 27, 385-409.

Jo, B. (2002). Model misspecification sensitivity analysis in estimating causal effects of interventions with noncompliance. Statistics in Medicine, 21, 3161-3181.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. Psychological Methods, 7, 178-193.

Jo, B., Asparouhov, T., Muthén, B., Ialongo, N. & Brown, H. (2007). Cluster randomized trials with treatment noncompliance. Accepted for publication in Psychological Methods.

References (Continued)

Jo, B. & Muthén, B. (2003). Longitudinal studies with intervention and noncompliance: Estimation of causal effects in growth mixture modeling. In N. Duan and S. Reise (Eds.), Multilevel modeling: methodological advances, issues, and applications, Multivariate Applications Book Series (pp. 112-139). Lawrence Erlbaum Associates.

Jo, B. & Muthén, B. (2001). Modeling of intervention effects with noncompliance: A latent variable approach for randomized trials. In G. A. Marcoulides & R. E. Schumacker (eds.), New developments and techniques in structural equation modeling (pp.57-87). Lawrence Erlbaum Associates.

Little, R.J. & Yau, L.H.Y. (1998). Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. Psychological Methods, 3, 147-159.

Latent Class Analysis

Bandeen-Roche, K., Miglioretti, D.L., Zeger, S.L. & Rathouz, P.J. (1997). Latent variable regression for multiple discrete outcomes. Journal of the American Statistical Association, 92, 1375-1386.

Bartholomew, D.J. (1987). Latent variable models and factor analysis. New York: Oxford University Press.

References (Continued)

Bucholz, K.K., Heath, A.C., Reich, T., Hesselbrock, V.M., Kramer, J.R., Nurnberger, J.I., & Schuckit, M.A. (1996). Can we subtype alcoholism? A latent class analysis of data from relatives of alcoholics in a multi-center family study of alcoholism. Alcohol Clinical Experimental Research, 20, 1462-1471.

Clogg, C.C. (1995). Latent class models. In G. Arminger, C.C. Clogg & M.E. Sobel (eds.), Handbook of statistical modeling for the social and behavioral sciences (pp 331-359). New York: Plenum Press.

Clogg, C.C. & Goodman, L.A. (1985). Simultaneous latent structural analysis in several groups. In Tuma, N.B. (ed.), Sociological Methodology, 1985 (pp. 18-110). San Francisco: Jossey-Bass Publishers.

Dayton, C.M. & Macready, G.B. (1988). Concomitant variable latent class models. Journal of the American Statistical Association, 83, 173-178.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. Journal of the American Statistical Association, 87, 476-486.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, 61, 215-231.

Guo, J., Wall, M. & Amemiya, Y. (2006). Latent class regression on latent factors. Biostatistics, 7, 145 - 163.

References (Continued)

Hagenaars, J.A. & McCutcheon, A.L. (2002). Applied latent class analysis. Cambridge, UK: Cambridge University Press.

Heijden, P.G.M., Dressens, J. & Bockenholt, U. (1996). Estimating the concomitant-variable latent-class model with the EM algorithm. Journal of Educational and Behavioral Statistics, 21, 215-229.

Heinen, D. (1996). Latent class and discrete latent trait models: similarities and differences. Thousand Oakes, CA: Sage Publications.

Jedidi, K., Jagpal. H.S. & DeSarbo, W.S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. Marketing Science, 16, 39-59.

Lazarsfeld, P.F. & Henry, N.W. (1968). Latent structure analysis. New York: Houghton Mifflin.

Muthén, B. (2001). Latent variable mixture modeling. In G.A. Marcoulides & R.E. Schumacker (eds.), New developments and techniques in structural equation modeling (pp. 1-33). Lawrence Erlbaum Associates.

Muthén, B. & Muthén, L. (2000). Integrating person-centered and variable-centered analysis: growth mixture modeling with latent trajectory classes. Alcoholism: Clinical and Experimental Research, 24, 882-891.

References (Continued)

Nestadt, G., Hanfelt, J., Liang, K.Y., Lamacz, M., Wolyniec, P., Pulver, A.E. (1994). An evaluation of the structure of schizophrenia spectrum personality disorders. Journal of Personality Disorders, 8, 288-298.

Nylund, K.L., Asparouhov, T., & Muthen, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. Structural Equation Modeling, 14, 535-569.

Qu, Y., Tan, M., & Kutner. M.H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics, 52, 797-810.

Rindskopf, D. (1990). Testing developmental models using latent class analysis. In A. von Eye (ed.), Statistical methods in longitudinal research: Time series and categorical longitudinal data (Vol 2, pp. 443-469). Boston: Academic Press.

Rindskopf, D., & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. Statistics in Medicine, 5, 21-27.

Stoolmiller, M. (2001). Synergistic interaction of child manageability problems and parent-discipline tactics in predicting future growth in externalizing behavior for boys. Developmental Psychology, 37, 814-825.

Uebersax, J.S., & Grove, W.M. (1990). Latent class analysis of diagnostic agreement. Statistics in Medicine, 9, 559-572.

References (Continued)

Yamamoto, K. & Gitomer, D. (1993). Application of a HYBRID model to a test of cognitive skill representation. In N. Frederiksen, R. Mislevy, and I Beijar (eds.), Test theory for new generation of tests. Hillsdale, NJ: LEA.

Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. Psychometrika, 62, 297-330.