Latent Class Analysis

- Jouko Miettunen, PhD
- Department of Psychiatry
- University of Oulu
- phone: +358-8-3156923
- e-mail: jouko.miettunen@oulu.fi

1

Topics of this presentation

- Introduction
- Related methods
- Applications
- Statistical issues
- Software
- Examples
- References

2

Introduction (1)

Latent Class Analysis

- Specific statistical modelling method developed to group subjects according to selected characteristics
  - ◆ Classifies subjects to groups
  - ◆ Identifies characteristics that indicate groups

3

Related methods (1)

Cluster analysis

- Based on distances between observations not on modelling

Fuzzy clustering

- ”Grades of membership”
- Do not give characteristics for groups

Structural equation modelling and factor analyses

- Investigate variables not subjects
  - ◆ used to classify questionnaire items into subsets
  - ◆ models can include also confounders

4

Related methods (2)

Several names are used to label LCA as a cluster analysis method

- mixture likelihood approach to clustering
- model-based clustering
- mixture-model clustering
- Bayesian classification
- unsupervised learning
- latent class cluster analysis

5

Applications (1)

- PubMed found 159 and PsycINFO 211 articles (December 2004)
- Several in psychiatry and psychology
  - ◆ already 1982 in schizophrenia
- Categories of people
  - ◆ who are smokers or alcoholists
  - ◆ with eating disorders
  - ◆ with breast cancer

6

Applications (2)

- Diagnostic validity / syndromes
  - ADHD
  - Hypochondriasis
  - psychosis
  - etc.
- Genetic studies
  - Alzheimer disease, etc.

7

Statistical issues (1)

- Model based clustering method
  - E.g. maximum likelihood methods
  - Most packages use EM algorithm or some modification
- All observed variables (e.g., symptoms) should be statistically independent (roughly, uncorrelated) within each latent class

8

Statistical issues (2)

- Problem of "local maxima," where the computer program, trying to find best-fitting values for quantities such as the population base rates of the latent classes, instead converges on values that are not best-fitting
- Many programs use multiple sets of starting values

9

Statistical issues (3)

Model comparisons

- Information Criteria
  - Akaike's IC
  - Bayesian IC
  - CAIC
- Likelihood-ratio tests
- Entropy based measures
  - indicate how well the indicators predict class membership

10

Software (1)

Mplus

- Integrates random effect, factor, and latent class analysis in both cross-sectional and longitudinal settings and for both single-level and multi-level data.
- Latent class analysis
  - unique features for covariates and complex sample data
- www.statmodel.com

11

Software (2)

Stata package: GLLAMM

- Generalized Linear Latent And Mixed Models
- Can be used for
  - generalised linear mixed models
  - multilevel factor models
  - latent trait models
  - multilevel structural equation models
  - latent class models
- http://www.gllamm.org/

12

Software (3)

R package: LCA 1.1

- A library of 15 programs and utility functions for performing exploratory latent class analysis on binary data.
- Reference:
  - Waller N. An R package for exploratory latent class analysis. Applied Psychological Measurement. Vol 28(2), Mar 2004, pp. 141-142

13

Software (4)

Latent GOLD

- http://www.latentGOLD.com/
- http://www.statisticalinnovations.com
- Reference:
  - Vermunt JK & Magidson J. Applications of Latent Class Analysis: An introduction to the technique and the Latent GOLD software. (http://www.gla.ac.uk/External/RSS/RSScomp/vermunt.pdf)
- LEM (older free LCA software)

14

Example: Anti-Social Behavior

- National Longitudinal Survey of Youth (NLSY)
- Respondent ages between 16 and 23
- Background information: age, gender and ethnicity
- N=7,326

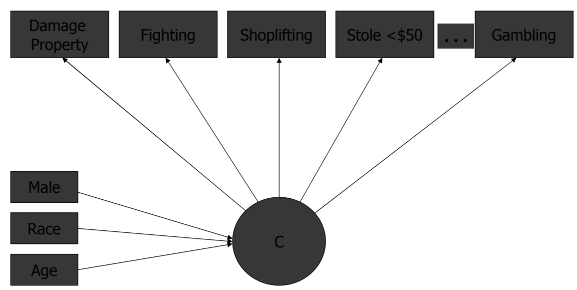
17 antisocial dichotomously scored behavior items:

- Damaged property
- Fighting
- Shoplifting
- Stole <$50
- Stole >$50
- Use of force
- Seriously threaten
- Intent to injure
- Use Marijuana
- Use other drug
- Sold Marijuana
- Sold hard drugs
- 'Con' somebody
- Stole an Automobile
- Broken into a building
- Held stolen goods
- Gambling Operation

Reference:
http://www.ats.ucla.edu/stat/mplus/seminars/lca/default.htm

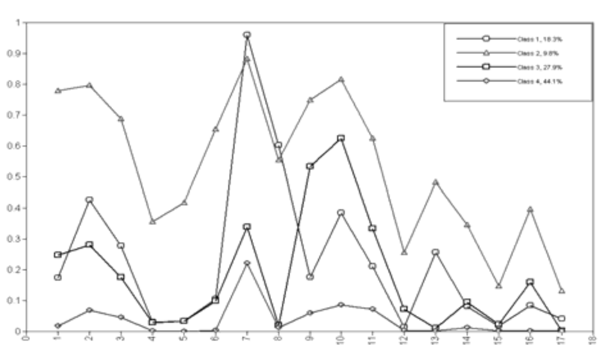
15

Model for Anti-Social Behavior



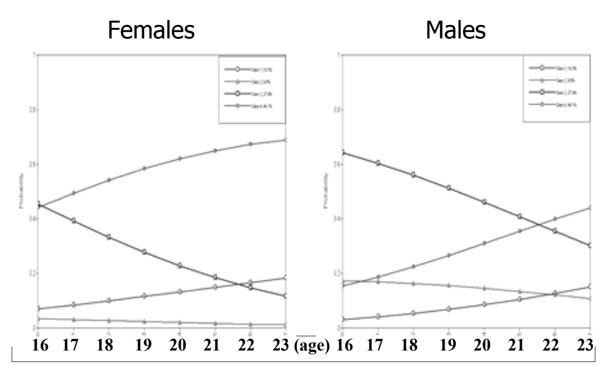
16

Anti-Social Behavior probabilities



17

Relationship between class probabilities and age by gender



18

Results: Anti-Social Behavior

- Summary of four classes:
  - Property Offense Class (9.8%)
  - Substance Involvement Class (18.3%)
  - Person Offenses Class (27.9%)
  - Normative Class (44.1%)
- Classification Table:

Columns: Latent class

Rows: Average latent class probability for most likely latent class membership

	1	2	3	4
1	**0.854**	0.031	0.070	0.040
2	0.041	**0.917**	0.040	0
3	0.058	0.021	**0.820**	0.100
4	0.038	0	0.080	**0.880**

19

Example: Temperament

Northern Finland 1966 Birth Cohort

- Women who were living in the provinces of Oulu and Lapland and were due to deliver during 1966 (12,058 live births)
- Data from the 31 year follow-up
  - Temperament and Character Inventory (TCI)

20

Variables for analyses

- Temperament and Character Inventory
  - True/False items (temperament part)
    - Novelty Seeking (40 items)
    - Harm Avoidance (35 items)
    - Reward Depence (24 items)
    - Persistence (8 items)
  - covariates
    - educational level 1997
      - basic, secondary, tertiary
    - gender
- N = 4850

21

Temperament Example I

- Mplus 3.01 program
- LCA with 4 categories
- No covariates
- No intercorrelations between temperament variables

22

Mplus input

```
MPLUS INPUT INSTRUCTIONS

title: Mplus example on TCI variables - four factors
data: file is c:\tci\tcimplus.dat;
variable: names IDKT TCIP TCINS TCIHA TCIRD SUKUP TUT97LK;
   missing = all (99);
   usev are TCIP TCINS TCIHA TCIRD;
   classes = c(4);
analysis: type = mixture; starts = 20 2;
model:
   %overall%
   %c#1%
   %c#2%
   %c#3%
   %c#4%
output: tech1 tech8 tech11;
savedata: file is tciout.sav; save = cprobabilities;
plot: type=plot3;
   series tciha tcins tcird tcip(*);
```

23

Mplus output (1)

Output: TESTS OF MODEL FIT

INFORMATION CRITERIA

Akaike (AIC)	107150.024
Bayesian (BIC)	107299.218
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	107226.133
Entropy	0.579

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

H0 Loglikelihood Value	-53675.803
2 Times the Loglikelihood Difference	247.582
Difference in the Number of Parameters	5
Mean	-10.709
Standard Deviation	35.960
P-Value	0.0000

24

Mplus output (2)

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Latent Classes	Class Counts and Proportions	
1	714	0.14722
2	1358	0.28000
3	482	0.09938
4	2296	0.47340

Mplus output (3)

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

	1	2	3	4
1	**0.730**	0.002	0.059	0.209
2	0.003	**0.807**	0.063	0.128
3	0.070	0.134	**0.696**	0.100
4	0.116	0.090	0.035	**0.759**

Mplus output (4)

MEANS FOR LATENT CLASSES

	1	2	3	4
NOVELTY SEEKING (NS)	17.300	21.554	15.414	22.059
HARM AVOIDANCE (HA)	20.309	9.109	18.766	13.583
REWARD DEPENDENCE (RD)	13.358	14.599	14.066	15.543
PERSISTENCE (P)	2.750	5.949	5.690	3.458

Class characteristics (subjective):

Class 1: LOW P, HIGH HA
Class 2: HIGH P, LOW HA, HIGH NS
Class 3: LOW NS, HIGH HA, HIGH P
Class 4: LOW P, HIGH NS

Temperament Example II

- Mplus 3.01 program
- LCA with 4 categories
- Intercorrelations
- Covariates

Mplus input

MPLUS INPUT INSTRUCTIONS, MODEL PART

```
model:
  %overall%
    tcip with tciha;
    tciha with tcins;
    tcird with tcins;
    c#1 c#2 c#3 on SUKUP TUT97LK;
  %c#1%
  %c#2%
  %c#3%
  %c#4%
```

with = ”correlated with”
on = ”regression on”
- last class is reference class

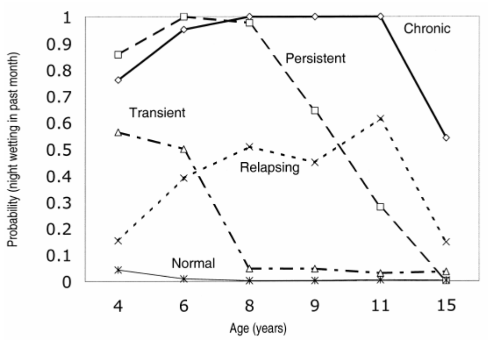
Mplus output

MODEL FIT

Akaike (AIC)	105992.455
Bayesian (BIC)	106258.412
Sample-Size Adjusted BIC	106128.128
Entropy	0.618

Latent Classes	Class Counts	Proportions
1	460	0.09485
2	761	0.15691
3	1763	0.36351
4	1866	0.38474

Example: Nightime wetting



Croudace et al. Developmental typology of trajectories to nighttime bladder control: Epidemiologic application of longitudinal latent class analysis. (2003)

31

References (1)

- Muthén B & Muthén M. Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. Alcohol Clin Exp Res, 24, 882-91, 2000.
- Croudace TJ, Järvelin MR, Wadsworth ME & Jones PB. Developmental typology of trajectories to nighttime bladder control: Epidemiologic application of longitudinal latent class analysis. American Journal of Epidemiology 157(9), 834-42, 2003.
- Rindskopf D & Rindskopf W. The value of latent class analysis in medical diagnosis. Statistics in Medicine, 5, 21-7, 1986.

32

References (2)

- Skrondal A & Rabe-Hesketh S. Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. Chapman & Hall/CRC, 2004.
- Hagenaars JA & McCutcheon AL. Applied latent class analysis. Cambridge, UK: Cambridge University Press, 2002.
- More references and examples, e.g. in
  - www.statmodel.com

33